Exploring the evolutionary dynamics of sound symbolism

Chundra A. Cathcart (chundra.cathcart@uzh.ch)
Department of Comparative Language Science, University of Zurich
Center for the Interdisciplinary Study of Language Evolution, University of Zurich
DFG Center "Words, Bones, Genes, Tools", Universität Tübingen

Gerhard Jäger (gerhard.jaeger@uni-tuebingen.de)
Seminar für Sprachwissenschaft, Universität Tübingen
DFG Center "Words, Bones, Genes, Tools", Universität Tübingen

Abstract

This paper uses phylogenetic modeling to investigate the evolutionary mechanisms responsible for the maintenance of sound symbolism in the world's languages. Applying our model to sound-meaning correspondences reported in the literature, we find that many previously established associations are weaker than expected when analyzed using our framework. This is possibly because certain sound-meaning associations are artifacts of slow-changing vocabulary items rather than specific preferences for certain sounds in words with certain meanings. For sound-meaning associations for which we find evidence, the maintenance of sound symbolism appears to be due to a tendency to preserve words in certain meanings if certain sounds are present.

Keywords: Sound symbolism; iconicity; sound patterns; linguistic evolution; phylogenetic modeling

Introduction

A key characteristic of spoken human languages is the largely arbitrary nature of sound patterns (Hockett, 1959): the first consonant of *cat* has no iconic relationship to the entity to which it refers, nor does it provide any cues to word class or any other information that might assist a learner who has not encountered the form before in deducing its meaning. Arbitrariness is a major driver of flexibility in language use, as the set of possible forms that can be assigned to a particular concept is not necessarily constrained by the need for a relationship between form and meaning that is iconic (i.e., grounded in perceptuomotor analogies).

At the same time, exceptions to arbitrariness are well documented. Certain sounds are over-represented in words denoting smallness (Ohala et al., 1994; Ekström, 2022). Sound symbolism, a manifestation of iconicity, is shown to aid in word learning (Imai & Kita, 2014). Research shows also that semantically similar forms cluster together phonologically (Monaghan, Shillcock, Christiansen, & Kirby, 2014), though it is not clear whether this represents general cross-linguistic sound symbolic patterns or phenomena that tend to be more language specific, like phonaesthemes or systematicity (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; Kwon & Round, 2015). Recent work identifies robust sound-meaning correspondences in representative samples of the world's languages, while controlling for nuisance factors such as genetic relatedness and geographic proximity (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Johansson, Anikin, Carling, & Holmer, 2020; Winter, Sóskuthy, Perlman, & Dingemanse, 2022), indicating that sound symbolism is deeply entrenched in linguistic systems. Semantic information has been shown to reduce uncertainty in character-level language models of word form prediction (Pimentel, McCarthy, Blasi, Roark, & Cotterell, 2019), further supporting the notion that sound and meaning are intertwined in languages' lexicons. However, our understanding of the pathways of development of the sound symbolic component in the world's languages is currently underdeveloped. Though a handful of studies document the diachrony of motivated sound patterns in individual languages (e.g., Carling & Johansson, 2014), little is known regarding the overarching processes that introduce and maintain sound symbolism in languages' lexicons.

In this paper, we investigate the evolutionary dynamics of sound symbolic patterns in the basic vocabulary of the world's languages. We use a phylogenetic model capable of shedding light on the conditions under which vocabulary items arise and are replaced. For instance, we can infer whether a word form is more likely to be used for a particular concept (e.g., *Nase* for NOSE) if it contains a particular sound (e.g., /n/). Additionally, we can infer whether processes which mutate word forms (e.g., sound change, analogical change, compounding, etc.) are more likely to introduce or remove certain sounds in words with certain meanings. Finally, we can assess whether a vocabulary item is more likely to be replaced if a particular sound is absent (e.g., if it lacks a beneficial sound symbolic cue). In addition to inferring the dynamics of individual sounds within concepts, we can compare these quantities across concepts for a given sound, asking, e.g., if loss of the sound /n/ is more likely to lead to vocabulary replacement in the concept NOSE than in the concept EYE.

We conduct two studies using phylogenetic comparative methods. First, we model the evolution of individual sound patterns independently within lexical items corresponding to different concepts. This allows us to investigate previously reported associations between concepts and sounds as well as explain how these associations take root and are maintained over time. Additionally, we analyze the diachronic preferences for certain sounds in lexical items with particular meanings in comparison to lexical items with other meanings. This allows us to assess whether certain concepts exhibit stronger relative long-term preferences for certain sounds in comparison to other concepts.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

We find that within lexical items corresponding to individual concepts, there is little support for the idea that evolutionary forces work to maintain sound symbolism; in many cases they work *against* sound symbolism. A comparison of evolutionary trends across pairs of meanings fails to find strong support for certain previously reported associations; we discuss reasons for this negative finding. In cases where we find evidence for relative sound symbolic preferences within concepts, we find that these trends are due largely to selection-like pressure toward preserving forms with certain sound patterns in specific meaning slots.

Materials and methods

Data

Lexical ROOT-MEANING TRAITS are an important feature in phylogenetic linguistics. Related languages share a root-meaning trait if they express the same concept with forms from the same cognate class (Dyen, Kruskal, & Black, 1992; Ringe, Warnow, & Taylor, 2002; Bouckaert et al., 2012; Chang, Cathcart, Hall, & Garrett, 2015). As an example, Spanish *comer* 'eat' (< Latin *com-edere* 'chew up') agrees with ancestral Latin *edere* in employing the Proto-Indo-European root $*h_1ed$- in the meaning slot EAT, whereas French has *manger* (< Latin *manducare* 'chew'), reflecting a different cognate class. The evolution of root-meaning traits was analyzed using data from a subset of the Lexibank repository (List et al., 2021) that has been processed to normalize orthographic forms and link records in different languages to the Concepticon semantic taxonomy (List, Tjuka, Rzymski, Greenhill, & Forkel, 2022). Among the datasets in the Lexibank repository, a number contain data corresponding to the same languages. This overlap can potentially duplicate data from the same families, biasing results. For this reason, we work with a subset of datasets coded for cognacy.

We investigate sound-concept relationships reported in two recent data-driven studies (Blasi et al., 2016; Johansson et al., 2020). For comparability with previous results, we convert forms to the orthography used by the Automated Similarity Judgment Program (Wichmann, Holman, & Brown, 2018) using the Python package `asjp` (Sofroniev, 2018). For phylogenetic comparative analyses, we code root-meaning traits from Lexibank according to whether or not a sound reported to be symbolic for a given concept is present or absent (e.g., whether /n/ is present in the root-meaning trait for NOSE in a particular language).

The phylogenetic comparative methods we employ require phylogenies of the languages under study. We use MrBayes v. 3.2.7 (Huelsenbeck & Ronquist, 2001) to infer phylogenies for the language families in the data sample we use on the basis of the binary (i.e., not coded for the presence of sound symbolism) root-meaning traits found in Lexibank. We assume gamma-distributed rate variation across classes of root-meaning traits, along with a uniform tree prior and a relaxed clock model with independent gamma-distributed rate variation across branches. Phylogenies were inferred over 1e8 iterations of Markov chain Monte Carlo over two chains, with a thinning interval of 1000 and the first quarter of samples discarded as burn-in. This resulted in a posterior sample of 75000 trees for each family under study.

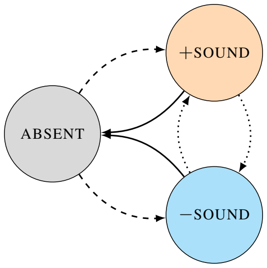


Figure 1: Continuous-time Markov process modeling the evolution of sound patterns within root-meaning traits. The system transitions between three states, in which the root meaning trait is ABSENT, or present with a sound pattern present (+SOUND) or absent (−SOUND).

Model

In a given language, root-meaning trait is either absent or present, and if present, either contains a symbolic sound or does not. We assume that root-meaning traits evolve between these three states according to a continuous-time Markov (CTM) process parameterized by transition rates representing the frequencies at which transition types occur. These rates are unobserved and must be estimated on the basis of observed linguistic data and language phylogenies, usually via Bayesian inference. This model is analogous to biological models of morphologically dependent traits (Maddison, 1993): just as the evolutionary dynamics of trait values specific to an anatomical character (e.g., tail color) are relevant only when the character is present, the evolutionary dynamics of sounds within root-meaning traits depend on the root-meaning trait's presence. The CTM process involved is schematized in Figure 1.

Of particular interest to the study of the evolutionary persistence of sound symbolism are asymmetries in pairs of transition rates governing the evolution of root-meaning traits. Fitted independently for a root-meaning trait corresponding to a single concept, this model can tell us whether the trait is more likely to arise (i.e., a reflex of a cognate class is used to represent a concept) if a certain sound is present versus absent (by comparing the frequencies with which the transitions marked by dashed arrows, representing BIRTH RATES, occur). It can also shed light on whether gains of a sound (e.g., via sound change, analogical change, and related processes; cf. Nuckolls, 1999; Carling & Johansson, 2014; Flaksman, 2017) are more frequent than losses (by comparing transitions marked with dotted lines representing MUTATION RATES). Rate asymmetries can tell us whether the trait is more likely to be lost if one of these conditions is not met (by comparing transitions marked by solid lines representing

DEATH RATES). Finally, the STATIONARY PROBABILITY or equilibrium distribution of the CTM process governing root-meaning trait evolution, which represents the proportion of time in which the system is expected to be in a particular state as time approaches infinity, can tell us whether a concept exhibits a greater long-term propensity towards a sound's presence than its absence. If used to analyze the behavior of a single sound across root-meaning traits corresponding to different concepts, the same model can tell us whether the pressures represented by each of these rate asymmetries (mediating the presence of a particular sound) as well as asymmetries in stationary probability are stronger in root-meaning traits with certain meanings than in others (e.g., we can compare propensity toward the sound /n/ in the concept NOSE vs. TONGUE).

We model the evolution of these traits according to the hierarchical model outlined below. A root-meaning trait expressing a concept with index $c \in \{1,...,C\}$ evolves in a family with index $f \in \{1,...,F\}$ according to the following transition rates:

- Birth rates:
  - $\lambda^{-}_{c,f}$ (the rate at which transitions of the type ABSENT $\rightarrow$ $-$SOUND occur)
  - $\lambda^{+}_{c,f}$ (the rate at which transitions of the type ABSENT $\rightarrow$ $+$SOUND occur)
- Mutation rates:
  - $\rho^{+-}_{c,f}$ (the rate at which transitions of the type $+$SOUND $\rightarrow$ $-$SOUND occur)
  - $\rho^{-+}_{c,f}$ (the rate at which transitions of the type $-$SOUND $\rightarrow$ $+$SOUND occur)
- Death rates:
  - $\mu^{-}_{c,f}$ (the rate at which transitions of the type $-$SOUND $\rightarrow$ ABSENT occur)
  - $\mu^{+}_{c,f}$ (the rate at which transitions of the type $+$SOUND $\rightarrow$ ABSENT occur)

Transition rates are distributed as follows. Below, we write $r_{c,f}$ to denote a given transition rate (e.g., $\lambda^{-}_{c,f}$).

For analyses of individual sound-mean associations and pairs of sound-mean associations, the generative process is as follows:

- $\mu_r \sim \text{Normal}(0,1)$
- $\sigma_r \sim \text{HalfNormal}(0,1)$
- $r_c \sim \text{Normal}(\mu_r, \sigma_r) : c \in \{1,...,C\}$
- $\sigma_{r_c} \sim \text{HalfNormal}(0,1) : c \in \{1,...,C\}$
- $r_{c,f} \sim \text{LogNormal}(r_c, \sigma_{r_c}) : f \in \{1,...,F\}$

Our hierarchical model structure allows rates to vary across concepts and families in a restricted manner. This potentially accounts for the fact that families may differ in terms of coverage. For a given concept, an expected evolutionary rate is given by $\exp(r_c)$, since family-level rates are log-normally distributed with mean r_c.

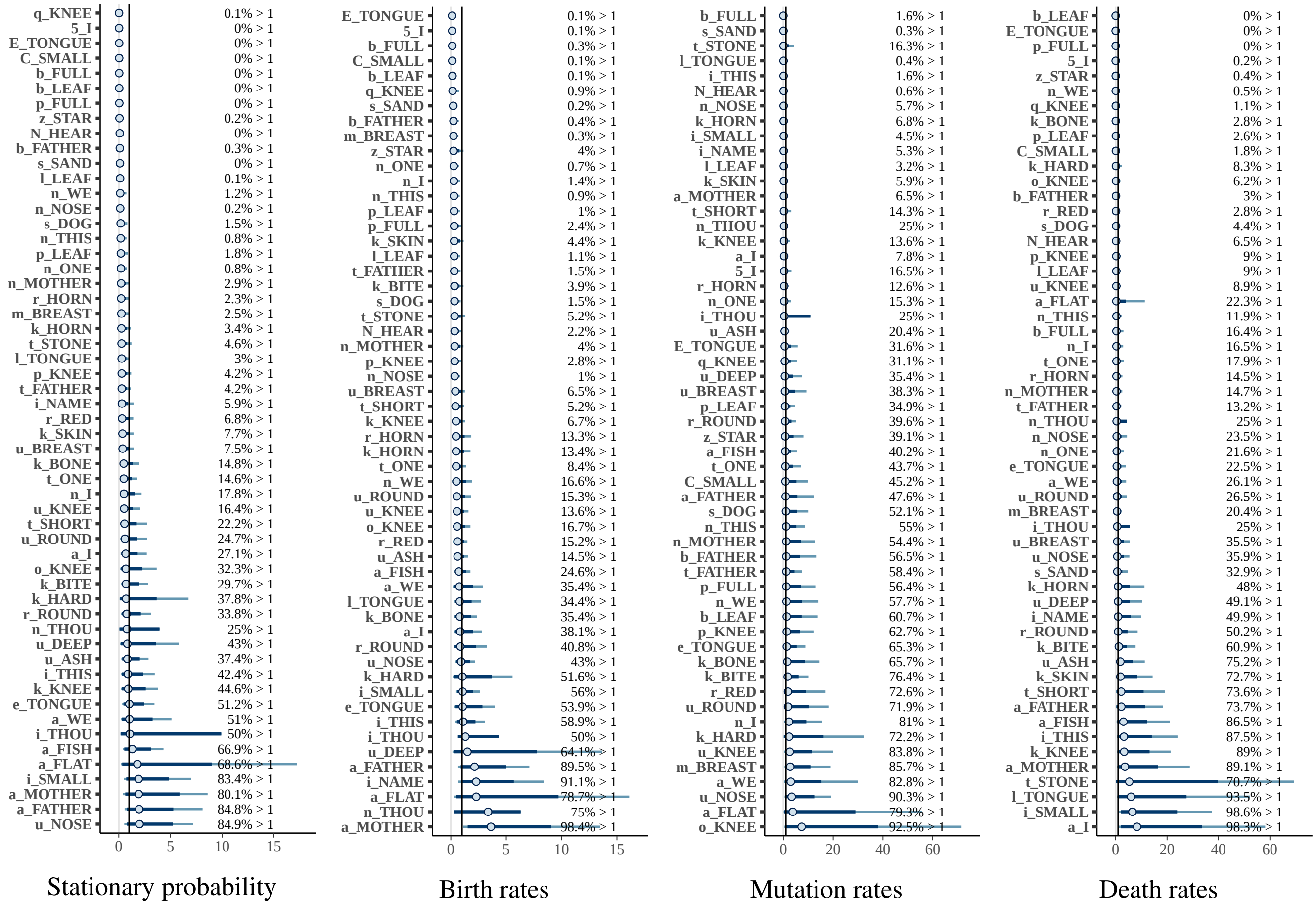
Posterior probabilities of rates are computed via the pruning algorithm (Felsenstein, 1981, 2004). Rates were inferred using RStan version 2.26.13 (Carpenter et al., 2017), running the No U-Turn Sampler (NUTS) over 4 chains for 2000 iterations, with the first half discarded as burn-in. Model convergence was assessed via the potential scale reduction factor (Gelman & Rubin, 1992), with values under 1.1 taken to indicate convergence. Rates were inferred on the basis of the maximum clade credibility (MCC) tree for each family.

We carry out two sets of analyses using this model. In the first, we infer transition rates for individual sound-meaning associations. In the second, we infer transition rates for certain sounds across selected pairs of concepts. For analyses of individual sound-concept associations, we run models for the union of the two sets of sound-meaning associations reported by Blasi et al. (2016) and Johansson et al. (2020). For analyses of pairs of sound-concept associations, we take the intersection between the two sets of associations, and restrict our pairwise comparisons to contrasts between (1) a concept reported as exhibiting a particular type of sound symbolism (e.g., /n/ in NOSE) and (2) concepts within this list of concepts exhibiting sound symbolism of any type (e.g., /n/ in TONGUE). We use this somewhat restricted data set due to the otherwise very large number of contrasts that we would need to take into account if we conducted this study in a more exploratory manner. Of the models we fitted to investigate these comparisons, some did not run to completion within a week on the computing cluster that we used, and were excluded from further analyses.

Results

Individual sound-meaning associations We inspect the evolutionary behavior of individual sound-meaning associations with respect to four quantities of interest.

Stationary probability For root-meaning traits corresponding to each individual sound-meaning association, we compute the stationary probability of the CTM process parameterized by the expected rates ($\exp(r_c)$) for the concept in question using the R function `svd()` for each sample from the posterior distribution. This results in a sample of simplices made up of the probabilities $P(\text{ABSENT})$, $P(-\text{SOUND})$, and $P(+\text{SOUND})$, representing the proportion of time expected to be spent in each state. We quantify the LONG-TERM PREFERENCE for sound symbolism within a concept via the ratio $\frac{P(+\text{SOUND})}{P(-\text{SOUND})}$. A ratio greater than 1 indicates that over the course of evolution of root-meaning traits corresponding to a given concept, a symbolic sound is expected to be present for longer than it is absent. Following standard conventions for one-tailed Bayesian hypothesis testing, we take there to be decisive evidence for a long-term preference for sound symbolism within a concept if the ratio $\frac{P(+\text{SOUND})}{P(-\text{SOUND})}$ is greater than 1 in 95% or more of samples. Figure 2 shows 95% and 85% credible intervals (CIs) for these ratios annotated with the proportion of samples for which the ratio is greater than 1. As is clear, there is no decisive evidence for long-term preferences for sound symbolism in any of the concept-sound



Stationary probability   Birth rates   Mutation rates   Death rates

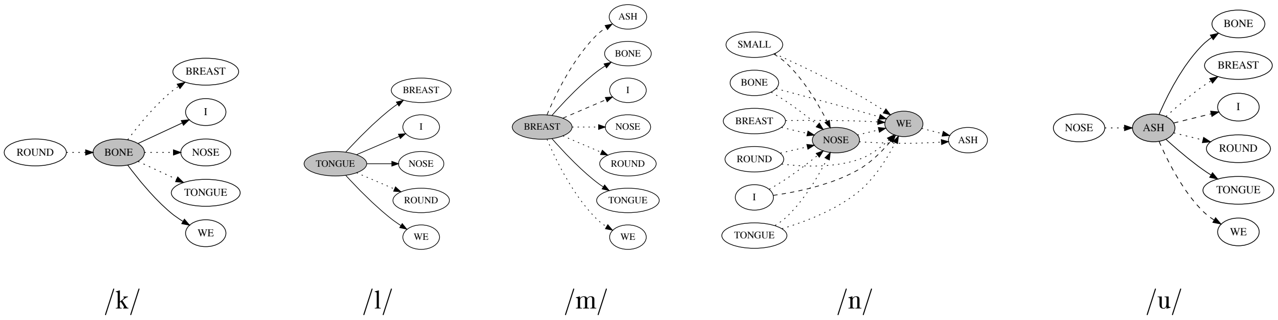
Figure 2: 95% and 85% credible intervals (CIs) for ratios representing asymmetries in quantities of interest (stationary probability, birth rates, mutation rates, and death rates). Ratios greater than one point to forces that foster sound symbolism during the evolution of root-meaning traits.

pairs analyzed. In fact, there appears to be decisive evidence *against* long-term preferences for sound symbolism in many of the concept-sound pairs analyzed.

Birth rates For each individual sound-meaning association, we compare the frequency with which corresponding root-meaning traits are born with symbolic sounds PRESENT versus ABSENT by comparing the quantities $\exp(\lambda_c^+)$ and $\exp(\lambda_c^-)$. The probability of a symbolic sound being present in the initial state of the system can be computed via the formula $\frac{\exp(\lambda_c^+)}{\exp(\lambda_c^+)+\exp(\lambda_c^-)}$ (Tarasov, 2019). Here, we assess the frequency with which root-meaning traits are born with versus without symbolic sounds via the ratio of the quantities $\exp(\lambda_c^+)$ and $\exp(\lambda_c^-)$. Values greater than one indicate that word forms are more likely to enter into particular meaning functions if a symbolic sound is present. Hypothesis assessment is as above; CIs and annotation are found in Figure 2. There is decisive evidence for an asymmetry in birth rates that favors sound symbolism in only one sound-meaning association, /a/ in MOTHER. This asymmetry may be trivial, due to the high overall frequency of sounds like /a/ in word lists (though similar effects are not found for /a/ in other concepts, e.g., FLAT), and in any event, it does not lead to a higher overall stationary probability of /a/ in words meaning MOTHER.

Mutation rates For each individual sound-meaning association, we compare the frequency with which changes altering word forms used for certain concepts INTRODUCE versus REMOVE a symbolic sound by comparing the quantities $\exp(\rho_c^+)$ and $\exp(\rho_c^-)$. We assess asymmetries in the frequencies of these different types of change via the ratio of these two quantities. Hypothesis assessment is as above; CIs and annotation are found in Figure 2. There is no decisive evidence that mutational changes altering word forms are more likely to introduce versus remove a symbolic sound for any of the sound-meaning pairs analyzed.

Death rates For each individual sound-meaning association, we compare the frequency with which forms CONTAINING versus LACKING a symbolic sound fall out of use in a particular meaning function by comparing the quantities $\exp(\mu_c^+)$ and $\exp(\mu_c^-)$. We assess asymmetries in the frequencies of



/k/ /l/ /m/ /n/ /u/

Figure 3: Pairwise differences in long-term preferences towards sounds (stationary probabilities) across concepts. Sounds in question are given below each graphviz (Ellson et al., 2002) visualization. Concepts are shaded if an association is reported for them with the sound in question in the literature cited.

these different change types via the ratio of these two quantities. Hypothesis assessment is as above; CIs and annotation are found in Figure 2. We find decisive evidence that root-meaning traits corresponding to the concept SMALL are more likely to die out (i.e., the words used will fall out of use in this particular meaning slot) if the sound /i/ is absent, and find the same effect for the concept I with respect to the sound /a/. Again here, this result may be trivial, due to the high overall frequency of sounds like /a/ in word lists. We find near-decisive evidence for this effect for the concept TONGUE with respect to /l/.

It is worth noting that although we find some decisive evidence for rate asymmetries both in birth and death rates, these do not ultimately impact the long-term preference for sound symbolism in the sound-meaning associations in question.

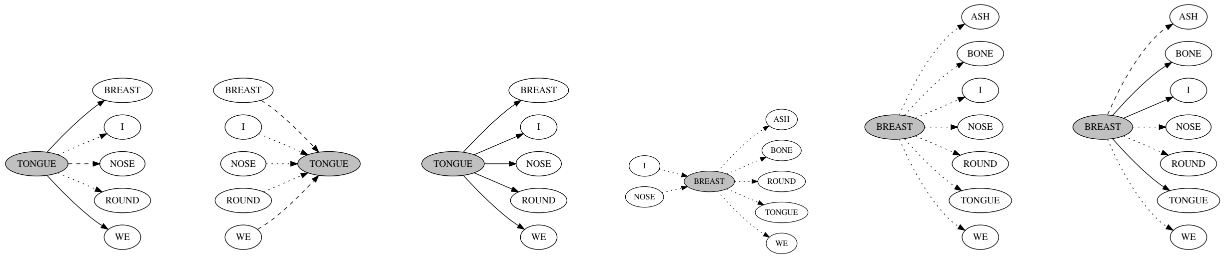
Pairs of features

Above, analyses of asymmetries in stationary probabilities and transition rates showed that evolutionary forces do not overwhelmingly favor sound symbolism in the histories of root-meaning traits — overall, /n/ is not expected during the majority of time that a words meaning NOSE are present. Still, these asymmetries, however weak, may foster the presence of symbolic sounds in certain concepts more than in others. We use the hierarchical model described above to investigate the evolution of symbolic sounds across selected pairs of concepts. We assess whether one concept has a greater sound-symbolic propensity than another concept (in terms of stationary probability, birth rates, mutation rates, or death rates) by comparing the ratios of the quantities discussed in the previous section. This could tell us, for instance, that one concept (e.g., TONGUE) has a greater long-term propensity for a sound (e.g., /l/) than another concept (e.g., NOSE), even if the stationary probability of the sound's presence in the first concept is lower than the stationary probability of its absence (as demonstrated to be the case in results for individual features).

Figure 3 shows pairwise differences in concept-level ratios of the stationary probability of presence versus absence of several sounds. Solid lines directed from one concept to another indicate that the former concept has a higher long-term preference for the sound in question than the latter in more than 95% of samples (dashed and dotted lines indicate that this relationship holds in more than 85% and 50%, respectively, of samples). Concepts are shaded if an association is reported for them with the sound in question in the literature cited.

These pairwise comparisons indicate relatively weaker support for certain sound-meaning associations than is reported in the literature that serves as the basis for our investigations. Surprisingly, the concept NOSE does not show a decisively higher preference for the sound /n/ than other concepts, even though this is a frequently reported association. Additionally, while the concept BREAST exhibits decisive evidence for a higher propensity toward /m/ than several other concepts, there is weaker evidence with respect to certain concepts not known for an association with this sound (e.g., ROUND).

We inspect pairwise concept-level differences in asymmetries in transition rates for two sounds within which there is stronger evidence for sound symbolism, /l/ (vis à vis TONGUE) and /m/ (vis à vis BREAST). Figure 4 gives graphviz visualizations of differences in concept-level rate asymmetries involving these sounds. We see that root-meaning traits for TONGUE are more likely to be born if /l/ is present than those meaning BREAST and WE, but root-meaning traits for BREAST are not more likely to be born with /m/ than the other concepts under study here. Mutation rates are no more likely to introduce the sounds /l/ and /m/, respectively, in the concepts TONGUE and BREAST than they are in other concepts. Finally, we see that root-meaning traits for TONGUE are decisively more likely to die out if /l/ is absent than in all other concepts investigated; this behavior also holds for /m/ to some extent. These results show us that where symbolism persists, it is largely mediated by quasi-selectional forces which conserve forms in certain meaning roles if they exhibit particular symbolic sounds.



/l/, birth /l/, mutation /l/, death /m/, birth /m/, mutation /m/, death

Figure 4: Pairwise differences in rate asymmetries involving the sounds /l/ and /m/. Concepts are shaded if an association is reported for them with the sound in question in the literature cited.

Discussion and Outlook

In this paper, we used a phylogenetic comparative model to investigate the development of sound symbolism within lexical items. We applied this model to individual sound-meaning associations in order to better understand aspects of the propensity toward sound symbolism within individual concepts, and also analyzed relative propensities towards particular sounds across multiple concepts. Our chief goal was to understand the specific diachronic mechanisms involved in the maintenance of sound symbolism, namely whether the presence of symbolic sounds within root-meaning traits corresponding to particular concepts is mediated by a tendency of forms with certain sounds to enter into use in certain meaning functions, mutational forces giving rise to certain sounds in forms with certain meanings, or the tendency of a form to be preserved in particular meaning function if a symbolic sound is present in it. A better understanding of these mechanisms sheds light on our understanding of the selectional forces underlying this phenomenon.

Restricting analyses to individual sound-meaning associations, we find virtually no evidence that symbolic sounds are overwhelmingly favored during the history of root-meaning traits corresponding to particular concepts; on average, symbolic sounds are expected to be present a minority of the time, and evolutionary mechanisms are often more likely to remove symbolic sounds than they are to create them. At first blush, this comes as a surprise, but this result makes greater sense in light of the fact that the large-scale analyses of sound symbolism cited in this paper demonstrate evidence for relative propensities for certain sounds among concepts that are higher than baseline propensities across all concepts. The sound /l/ may be present in only a small fraction of words meaning TONGUE, but this proportion will be larger than that of words for concepts exhibiting the same sound.

Our pairwise analyses find some support for sound-meaning associations reported in the literature. For instance, we find that TONGUE has a higher overall propensity for the sound /l/ than the majority of other concepts taken into consideration, and that BREAST exhibits a higher propensity for /m/ than a number of concepts. However, we find that some frequently reported associations are not supported by the modeling framework that we use. For instance, there is no evidence that words for NOSE exhibit a higher propensity for /n/ than in other concepts. There are a number of possible reasons for this finding. We use a different sample from the two studies cited which is restricted to relatively large families, a prerequisite for phylogenetic methods (though see Jäger & Wahle, 2021 for extensions to small families, orphans, and isolates, or alternatively, the global phylogenetic approach of Shcherbakova et al., 2023). However, the families we use exhibit broad geographic coverage and should be capable of capturing global evolutionary trends. We might also expect different results when probing associations between meanings and features (e.g., NASAL) as opposed to specific sounds; we investigated associations of the latter type because they are reported in the literature. The negative results we report may also be due to the fact that some previously reported associations are an artifact of slow-changing vocabulary items and historically deep areal patterns, confounds that were not captured by the phylogenetic and areal controls employed in previous studies. In cases where we find robust relative evidence for a sound-meaning correspondence, associations persist largely because forms with a particular sound (e.g., /l/ in TONGUE) are less likely to be replaced in particular meanings than words for other concepts with the same sound, but not necessarily due to processes which alter word forms or because forms with particular sounds are more likely to enter into use in particular meaning functions. Our results serve as an initial step in clarifying the diachronic mechanisms underlying sound symbolism in the world's languages.

Acknowledgements

This work was supported by the NCCR Evolving Language (SNSF Agreement 51NF40_180888), the ERC under the EU's Horizon 2020 research and innovation programme (Grant Agreement 834050), and the DFG CAS "Words, Bones, Genes, Tools." SNSF grant No. 207573 is gratefully acknowledged as well. Code used in analyses is found at `github.com/chundrac/sound-symbolism-cogsci2024`.

References

Birchall, J., Dunn, M., & Greenhill, S. J. (2021, July). *CLDF dataset derived from Birchall et al.'s "A Combined Comparative and Phylogenetic Analysis of the Chapacuran Language Family" from 2016.* Zenodo. doi: 10.5281/zenodo.5119306

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences*, *113*(39), 10818–10823.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., … Atkinson, Q. D. (2012). Mapping the origins and expansion of the indo-european language family. *Science*, *337*(6097), 957-960.

Bowern, C. (2021, July). *CLDF dataset derived from Bowern and Atkinson's "Internal Structure of Pama-Nyungan" from 2012.* Zenodo. doi: 10.5281/zenodo.5119341

Carling, G., & Johansson, N. (2014). Motivated language change: processes involved in the growth and conventionalization of onomatopoeia and sound symbolism. *Acta linguistica hafniensia*, *46*(2), 199–217.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76(1)*(1), 1-32.

Chacon, T. C. (2021a, July). *CLDF dataset derived from Chacon's "Annotated Swadesh Lists for Arawakan Languages" from 2017.* Zenodo. doi: 10.5281/zenodo.5118556

Chacon, T. C. (2021b, July). *CLDF dataset derived from Chacon's "A revised proposal of Proto-Tukanoan consonants and Tukanoan family classification" from 2014.* Zenodo. doi: 10.5281/zenodo.5118723

Chang, W., Cathcart, C., Hall, D., & Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European Steppe Hypothesis. *Language*, *91*(1), 194-244.

Constenla Umaña, A. (2021, July). *CLDF dataset derived from Constenla Umaña's "¿Existe relación genealógica entre las lenguas misumalpas y las chibchenses?" from 2005.* Zenodo. doi: 10.5281/zenodo.5121347

Deepadung, S., Buakaw, S., & Rattanapitak, A. (2021, July). *CLDF dataset derived from Deepadung et al.'s "Lexical Comparison of Palaung Dialects" from 2015.* Zenodo. doi: 10.5281/zenodo.5121402

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, *19*(10), 603–615.

Dunn, M., & Tresoldi, T. (2021, October). *evotext/ielex-data-and-tree: IELex data and tree (2021/11/08).* Zenodo. doi: 10.5281/zenodo.5556801

Dyen, I., Kruskal, J., & Black, P. (1992). An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society*, *82*(5).

Ekström, A. G. (2022). What's next for size-sound symbolism? *Frontiers in Language Sciences*, *1*. doi: 10.3389/flang.2022.1046637

Ellson, J., Gansner, E., Koutsofios, L., North, S. C., & Woodhull, G. (2002). Graphviz—open source graph drawing tools. In *Graph drawing: 9th international symposium, gd 2001 vienna, austria, september 23–26, 2001 revised papers 9* (pp. 483–484).

Feleke, T. L. (2021, July). *CLDF dataset derived from Feleke's "Ethiosemitic Languages" from 2021.* Zenodo. doi: 10.5281/zenodo.5126691

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, *17*(6), 367-76.

Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, Mass.: Sinauer Associates.

Ferraz Gerardi, F., & Reichert, S. (2021, July). *CLDF dataset derived from Gerardi and Reichert's "The Tupí-Guaraní Language Family: A Phylogenetic Classification" from 2020.* Zenodo. doi: 10.5281/zenodo.5127906

Flaksman, M. (2017). Iconic treadmill hypothesis. In A. Zirker, M. Bauer, O. Fischer, & C. Ljungberg (Eds.), *Dimensions of iconicity* (pp. 15–38). Amsterdam: John Benjamins.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457-511.

Greenhill, S. J., Haynie, H. J., Ross, R. M., Chira, A. M., List, J.-M., Campbell, L., … Gray, R. D. (2023). A recent northern origin for the uto-aztecan family. *Language*.

Hockett, C. F. (1959). Animal" languages" and human language. *Human Biology*, *31*(1), 32–39.

Huelsenbeck, J. P., & Ronquist, F. (2001). Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical transactions of the Royal Society B: Biological sciences*, *369*(1651), 20130298.

Johansson, N. E., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*, *24*(2), 253–310.

Jäger, G., & Wahle, J. (2021). Phylogenetic typology. *Frontiers in Psychology*, *12*. doi: 10.3389/fpsyg.2021.682132

Kolipakam, V. (2021, July). *CLDF dataset derived from Kolipakam et al.'s "DravLex:" from 2018.* Zenodo. doi: 10.5281/zenodo.5121580

Kwon, N., & Round, E. R. (2015). Phonaesthemes in morphological theory. *Morphology*, *25*, 1–27.

Lee, S. (2021, July). *CLDF dataset derived from Lee's "Sketch of Language History in the Korean Peninsula" from 2015.* Zenodo. doi: 10.5281/zenodo.5126904

Lee, S., & Hasegawa, T. (2021a, July). *CLDF dataset derived* *from Lee and Hasegawa's "Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages" from 2011.* Zenodo. doi: 10.5281/zenodo.5126801

Lee, S., & Hasegawa, T. (2021b, July). *CLDF dataset derived from Lee and Hasegawa's "Evolution of the Ainu Language in Space and Time" from 2013.* Zenodo. doi: 10.5281/zenodo.5126890

List, J.-M., Forkel, R., Greenhill, S. J., Rzymski, C., Englisch, J., & Gray, R. D. (2021, August). *Lexibank: A publicly available repository of standardized lexical datasets with automatically computed phonological and lexical features for more than 2000 language varieties.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.5227818` doi: 10.5281/zenodo.5227818

List, J. M., Tjuka, A., Rzymski, C., Greenhill, S., & Forkel, R. (Eds.). (2022). *Clld concepticon 3.0.0.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from `https://concepticon.clld.org/`

Maddison, W. P. (1993). Missing data versus missing characters in phylogenetic analysis. *Systematic Biology*, *42*(4), 576–581. Retrieved from `http://www.jstor.org/stable/2992490`

McElhanon, K. A. (2021, July). *CLDF dataset derived from McElhanon's "Preliminary Observations on Huon Peninsula Languages" from 1967.* Zenodo. doi: 10.5281/zenodo.5127348

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130299.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual review of anthropology*, *28*(1), 225–252.

Ohala, J. J., et al. (1994). The frequency code underlies the sound-symbolic use of voice pitch. *Sound symbolism*, *2*, 325–347.

Peiros, I. (2021, July). *CLDF dataset derived from Peiros' "Genetic classification of Austro-Asiatic languages" from 2004.* Zenodo. doi: 10.5281/zenodo.5127536

Pimentel, T., McCarthy, A. D., Blasi, D., Roark, B., & Cotterell, R. (2019). Meaning to form: Measuring systematicity as information. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 1751–1764).

Ringe, D., Warnow, T., & Taylor, A. (2002). Indo-European and computational cladistics. *Transactions of the Philological Society*, *100*(1), 59-129.

Robinson, L. C., & Holton, G. (2021, July). *CLDF dataset derived from Robinson and Holton's "Internal Classification of the Alor-Pantar Language Family" from 2012.* Zenodo. doi: 10.5281/zenodo.5121340

Sagart, L., Jacques, G., Lai, Y., & List, J.-M. (2021, July). *CLDF dataset derived from Sagart et al.'s "Sino- Tibetan Database of Lexical Cognates" from 2019.* Zenodo. doi: 10.5281/zenodo.5121409

Savelyev, A. (2021, July). *CLDF dataset derived from Savelyev and Robbeet's "Internal Structure of the Turkic Language Family" from 2020.* Zenodo. doi: 10.5281/zenodo.5137274

Shcherbakova, O., Michaelis, S. M., Haynie, H. J., Passmore, S., Gast, V., Gray, R. D., … Skirgård, H. (2023). Societies of strangers do not speak less complex languages. *Science Advances*, *9*(33), eadf7704.

Sofroniev, P. (2018). *asjp 0.0.2 (python library).* Retrieved from `https://pypi.org/project/asjp/`

Syrjänen, K., Honkola, T., Korhonen, K., Lehtinen, J., Vesakoski, O., & Wahlber, N. (2021, July). *CLDF Dataset derived from Syrjänen et al.'s "Shedding more light on language classification" from 2013.* Zenodo. doi: 10.5281/zenodo.5137236

Tarasov, S. (2019). Integration of anatomy ontologies and evo-devo using structured markov models suggests a new framework for modeling discrete phenotypic traits. *Systematic biology*, *68*(5), 698–716.

Walworth, M. (2021, July). *CLDF dataset derived from Walworth's "Polynesian Segmented Data" from 2019.* Zenodo. doi: 10.5281/zenodo.5126932

Wichmann, S. (2021, July). *CLDF dataset derived from Wichmann's "Lexicostatistical Dataset of Mixe-Zoquean" from 2006.* Zenodo. doi: 10.5281/zenodo.5126948

Wichmann, S., Holman, E. W., & Brown, C. H. (2018). *The ASJP database (version 18).* http://asjp.clld.org/.

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), 1035.