The building blocks of sound symbolism

Erben Johansson, Niklas

2020

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Erben Johansson, N. (2020). *The building blocks of sound symbolism*. [Doctoral Thesis (compilation), Division of Linguistics and Cognitive Semiotics]. Lund University (Media-Tryck).

Total number of authors:
1

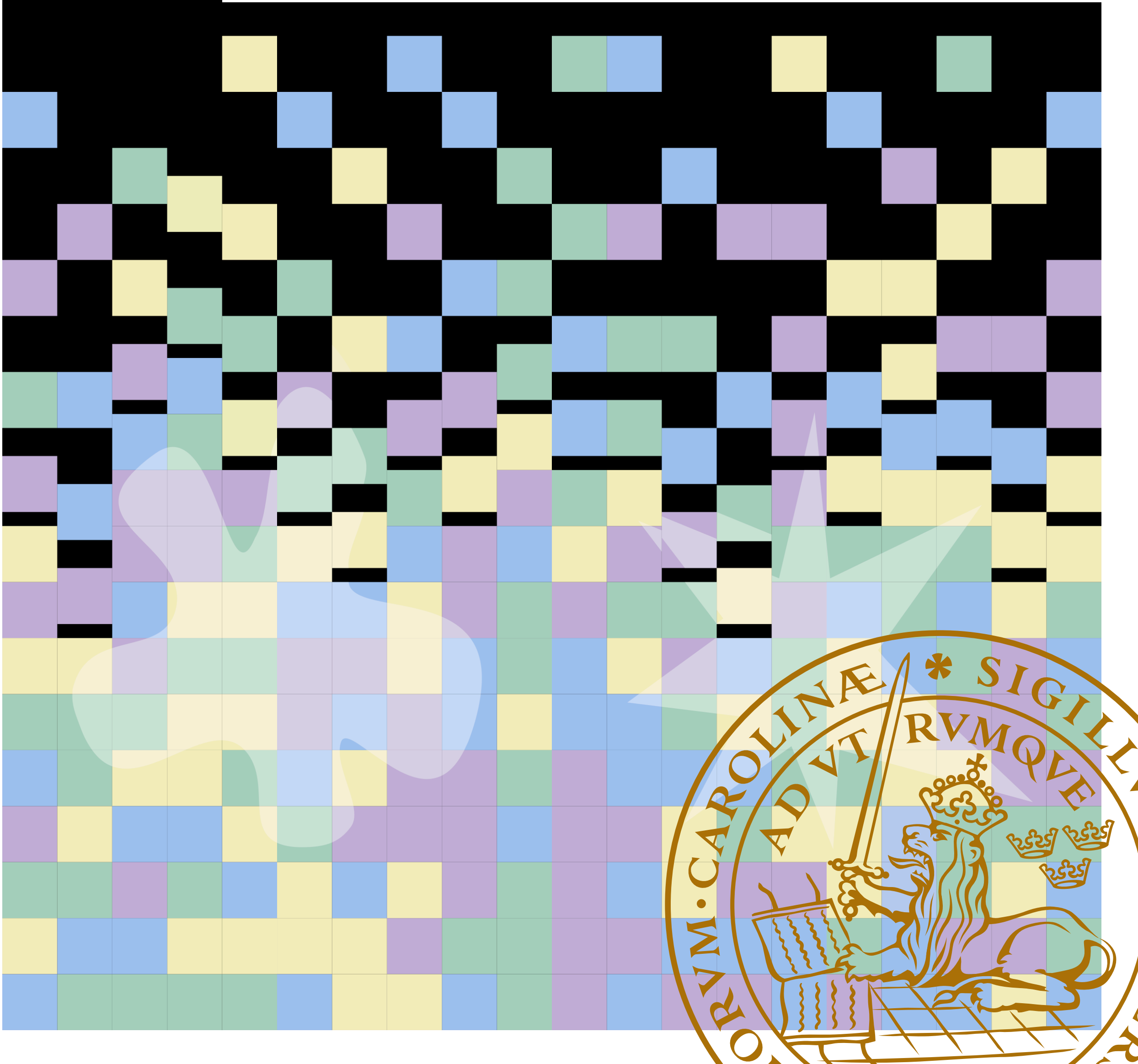
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

The building blocks of sound symbolism

NIKLAS ERBEN JOHANSSON

CENTRE FOR LANGUAGES AND LITERATURE | LUND UNIVERSITY



The building blocks of sound symbolism

The building blocks of sound symbolism

Niklas Erben Johansson

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculties of Humanities and Theology,
Lund University, Sweden.
To be defended at room H104, Centre for Languages and Literature,
Saturday June 6, 2020, 10:00

Faculty opponent
Associate professor Mark Dingemanse
Centre for Language Studies, Radboud University

Organization LUND UNIVERSITY Centre for Languages and Literature Author(s) Niklas Erben Johansson	**Document name** **Dcotoral disseration** **Date of issue** Saturday June 6 Sponsoring organization

Title and subtitle: The building blocks of sound symbolism

Abstract

Languages contain thousands of words each and are made up by a seemingly endless collection of sound combinations. Yet a subsection of these show clear signs of corresponding word shapes for the same meanings which is generally known as vocal iconicity and sound symbolism. This dissertation explores the boundaries of sound symbolism in the lexicon from typological, functional and evolutionary perspectives in an attempt to provide a deeper understanding of the role sound symbolism plays in human language. In order to achieve this, the subject in question was triangulated by investigating different methodologies which included lexical data from a large number of language families, experiment participants and robust statistical tests.

Study I investigates basic vocabulary items in a large number of language families in order to establish the extent of sound symbolic items in the core of the lexicon, as well as how the sound-meaning associations are mapped and interconnected. This study shows that by expanding the lexical dataset compared to previous studies and completely controlling for genetic bias, a larger number of sound-meaning associations can be established. In addition, by placing focus on the phonetic and semantic features of sounds and meanings, two new types of sounds symbolism could be established, along with 20 semantically and phonetically superordinate concepts which could be linked to the semantic development of the lexicon.

Study II explores how sound symbolic associations emerge in arbitrary words through sequential transmission over language users. This study demonstrates that transmission of signals is sufficient for iconic effects to emerge and does not require interactional communication. Furthermore, it also shows that more semantically marked meanings produce stronger effects and that iconicity in the size and shape domains seems to be dictated by similarities between the internal semantic relationships of each oppositional word pair and its respective associated sounds.

Studies III and IV use color words to investigate differences and similarities between low-level cross-modal associations and sound symbolism in lexemes. Study III explores the driving factors of cross-modal associations between colors and sounds by experimentally testing implicit preferences between several different acoustic and visual parameters. The most crucial finding was that neither specific hues nor specific vowels produced any notable effects and it is therefore possible that previously reported associations between vowels and colors are actually dependent on underlying visual and acoustic parameters.

Study IV investigates sound symbolic associations in words for colors in a large number of language families by correlating acoustically described segments with luminance and saturation values obtained from cross-linguistic color-naming data. In accordance with Study III, this study showed that luminance produced the strongest results and was primarily associated with vowels, while saturation was primarily associated with consonants. This could then be linked to cross-linguistic lexicalization order of color words.

To summarize, this dissertation shows the importance of studying the underlying parameters of sound symbolism semantically and phonetically in both language users and cross-linguistic language data. In addition, it also shows the applicability of non-arbitrary sound-meaning associations for gaining a deeper understanding of how linguistic categories have developed evolutionarily and historically.

Key words: sound symbolism, iconicity, basic vocabulary, lexical semantics, language evolution, typology

Classification system and/or index terms (if any)		
Supplementary bibliographical information		**Language: English**
ISSN and key title		**ISBN** 978-91-89213-02-9 (print) 978-91-89213-03-6 (digital)
Recipient's notes	**Number of pages** 58	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature Date 2020-04-22

The building blocks of sound symbolism

Niklas Erben Johansson

LUND
UNIVERSITY

Cover photo by Niklas Erben Johansson

Faculties of Humanities and Theology
General Linguistics
The Centre for Languages and Literature

ISBN 978-91-89213-02-9 (print)
978-91-89213-03-6 (digital)

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel
certified provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

MADE IN SWEDEN

Knowledge rests not upon truth alone, but upon error also

Table of Contents

Acknowledgements

This dissertation has grown out of a childish fascination for the human condition and specifically for how animals such as humans communicate in order to ensure progress. Correspondingly, this dissertation would never have been written without the help of all the people that have supported me.

My most meaningful thanks go to my love, Šárka Erben Johansson. Even if you would not admit it, you have been completely instrumental in writing this dissertation because you have helped me progress academically and emotionally over the eight wonderful years I have had the privilege to spend my life with you. I am deeply grateful for the countless hours you have spent discussing ideas with me, and reading, proofreading and editing various versions of my articles throughout this process. You have an amazing ability to help me break down my thoughts sprung from pure excitement to logical, concrete and comprehensive hypotheses. I am incredibly grateful that you keep challenging me intellectually when I oversimplify or when I am too impatient while you simultaneously support and encourage me when I feel overwhelmed. Likewise, I would like to thank my parents, Bengt-Göran Johansson and Piia Kullman, for their never-ending support throughout this process. I am eternally grateful for you taking a genuine and inquisitive interest in all aspects of this dissertation and for the mindboggling endurance you have shown when listening to me going on for hours about how exciting my new results were, how annoying writing can be and how hard the choices I had to make were. I love you all.

A very special thanks goes to my supervisors, Gerd Carling and Arthur Holmer, who guided me excellently through the various academic and administrative hurdles during my years as a PhD student. You have always made yourself available when I have been in need and you always listened to my questions and thoughts attentively and with great enthusiasm. Your broad expertise in linguistic research in conjunction with the academic freedom you have allowed me to have encouraged me to constantly progress. This has not only made me a far better researcher than I would ever have thought I could become during these few years, but it has also helped me build confidence to develop hunches into ideas into scientific studies.

One of the most valuable things I have learnt as a PhD student has been that conducting research and writing articles cooperatively yields unquestionable qualitative and quantitative benefits. I would therefore like to thank my coauthors, Andrey Anikin, Nikolay Aseyev, Jon W. Carr and Simon Kirby, for their efforts, insights and willingness to delve into this strange subfield of linguistics despite various academic backgrounds. It has been a real pleasure working with all of you. I specifically want to thank Andrey Anikin who let me kidnap him to aid me with the statistical components in all four articles in this dissertation, which has entailed not only working over data to make sense out of it, but also running statistical models on multiple devices together which has sometimes taken weeks. Our collaboration has led to countless stimulating discussions and it has helped me to develop my methodological thinking considerably. My stay at The University of Edinburgh also had a major impact on me and I would therefore like to thank the brilliant people at the Centre for Language Evolution who warmly welcomed into their community.

Lastly, I would like to thank the people at the linguistics department at Lund University, of which many are my dear friends, for contributing to the inviting, open and invigorating atmosphere I have been allowed to be a part of.

Abstract

Languages contain thousands of words each and are made up by a seemingly endless collection of sound combinations. Yet a subsection of these show clear signs of corresponding word shapes for the same meanings which is generally known as vocal iconicity and sound symbolism. This dissertation explores the boundaries of sound symbolism in the lexicon from typological, functional and evolutionary perspectives in an attempt to provide a deeper understanding of the role sound symbolism plays in human language. In order to achieve this, the subject in question was triangulated by investigating different methodologies which included lexical data from a large number of language families, experiment participants and robust statistical tests.

Study I investigates basic vocabulary items in a large number of language families in order to establish the extent of sound symbolic items in the core of the lexicon, as well as how the sound-meaning associations are mapped and interconnected. This study shows that by expanding the lexical dataset compared to previous studies and completely controlling for genetic bias, a larger number of sound-meaning associations can be established. In addition, by placing focus on the phonetic and semantic features of sounds and meanings, two new types of sounds symbolism could be established, along with 20 semantically and phonetically superordinate concepts which could be linked to the semantic development of the lexicon.

Study II explores how sound symbolic associations emerge in arbitrary words through sequential transmission over language users. This study demonstrates that transmission of signals is sufficient for iconic effects to emerge and does not require interactional communication. Furthermore, it also shows that more semantically marked meanings produce stronger effects and that iconicity in the size and shape domains seems to be dictated by similarities between the internal semantic relationships of each oppositional word pair and its respective associated sounds.

Studies III and IV use color words to investigate differences and similarities between low-level cross-modal associations and sound symbolism in lexemes. Study III explores the driving factors of cross-modal associations between colors and sounds by experimentally testing implicit preferences between several different acoustic and visual parameters. The most crucial finding was that neither specific hues nor specific vowels produced any notable effects and it is therefore possible that previously reported associations between vowels and colors are actually dependent on underlying visual and acoustic parameters.

Study IV investigates sound symbolic associations in words for colors in a large number of language families by correlating acoustically described segments with luminance and

saturation values obtained from cross-linguistic color-naming data. In accordance with Study III, this study showed that luminance produced the strongest results and was primarily associated with vowels, while saturation was primarily associated with consonants. This could then be linked to cross-linguistic lexicalization order of color words.

To summarize, this dissertation shows the importance of studying the underlying parameters of sound symbolism semantically and phonetically in both language users and cross-linguistic language data. In addition, it also shows the applicability of non-arbitrary sound-meaning associations for gaining a deeper understanding of how linguistic categories have developed evolutionarily and historically.

List of original papers

I. Erben Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2020). The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology*. doi: 10.1515/lingty-2020-2034

II. Erben Johansson, N., Carr, J. W., & Kirby, S. (submitted) Cultural evolution leads to vocal iconicity in an experimental iterated learning task. Submitted to *Journal of language evolution*.

III. Anikin, A., & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics, 81*(3), 764-777. doi: 10.3758/s13414-018-01639-7

IV. Johansson, N., Anikin, A., & Aseyev, N. (2019). Color sound symbolism in natural languages. *Language & Cognition*, 1-28. doi: 10.1017/langcog.2019.35

The contribution of the papers:

Study I: Coauthor Erben Johansson conducted the data collection, method evaluation and the writing of the text, as well as the majority of the theoretical and methodological design. Coauthor Anikin contributed to the methodological design and conducted the statistical analysis. Coauthors Carling and Holmer contributed to the theoretical and methodological design. All coauthors were active in the editing and the revision process.

Study II: Coauthor Erben Johansson conducted the data collection and the writing of the text. All coauthors contributed to the theoretical and methodological design and the method evaluation. All coauthors were also active in the editing and the revision process.

Study III: Both coauthors conducted the data collection, the method evaluation and the theoretical and methodological design. Coauthor Anikin wrote the majority of the introductory, methods and results sections, while the discussion and conclusion sections were written by both coauthors. Both coauthors were active in the editing and the revision process. Coauthor Anikin also conducted the statistical analysis.

Study IV: Coauthor Erben Johansson conducted the data collection and wrote the majority of the paper, while coauthor Anikin wrote the methods and results sections. Coauthor Anikin also conducted the statistical analysis. All coauthors conducted the method evaluation and the theoretical and methodological design and all were active in the editing and the revision process.

1. Introduction

When encountering a speaker of a language completely unknown to you, knowing how to initiate verbal communication tends to be difficult. If you are in luck, the language will be either somewhat closely related to one you know, or it uses several similar words due to geographical proximity which would allow you to establish some common ground. However, if this is not the case, decoding and acquiring a new language eventually leads to a demanding task in memorization. Yet, certain words, regardless of which language they come from, just seem to fit with the referents they denote. For example, across languages, words meaning 'round' tend to contain vowels that require the speakers to round their lips during the articulation of the sound which can be aligned with the meaning. This type of intuitive association between sounds and meanings is generally referred to as *sound symbolism*, but also as *(vocal) iconicity*, *non-arbitrariness, phonosemantics, motivatedness*. Throughout this dissertation, the terms *sound symbolism* and *vocal iconicity* are used interchangeably to denote this phenomenon and are not intended to contrast with, for example semiotic *indexicality*. *Iconicity* is used as a general umbrella term for any association between sign and meaning.

Associations between sounds and meanings are confirmed to be cross-linguistically prevalent geographically, synchronically and diachronically in unrelated languages. This suggests that studying the fundamental meanings that all languages utilize to some extent could tell us a great deal about how iconicity and sound symbolism have been, and are, affecting human language. Thus, this dissertation explores sound symbolism from several perspectives in order to better understand how it is established and constrained in language. This is achieved by addressing how large its extent is in the core of the lexicon, how sound symbolic associations emerge and develop under natural language simulation, and how cognitively deep the sound symbolic mappings are grounded.

Study I investigates the phonetic and semantic features involved in sound symbolism from a bottom-up perspective. For this study, a large database was created, consisting of 344 near-universal basic vocabulary concepts gathered from 245 language families. By transcribing the speech sounds and grouping them into phonetically and sound symbolically relevant sound groups, overrepresentations of phonetic features in the investigated meanings could be established. Aside from the 125 robust sound-meaning associations found, semantically and phonetically superordinate concepts (*macro-concepts*) could also be established which were linked to fundamental lexical fields in early human language. In addition, two new types of sound symbolic mappings were described.

Study II looks at how sound symbolic patterns emerge in initially arbitrary words by using an experimental setup which resembles the game of telephone, i.e. people forming a line in which the first person transfers a message to the next one and when the last player in line is reached the word has usually changed considerably. The experiment included two of the most thoroughly investigated semantic opposites in the sound symbolism literature, BIG-SMALL and ROUND-POINTY. 1,500 naïve participants were recruited and divided into five condition groups (BIG, SMALL, ROUND, POINTY and CONTROL) which contained ten chains of 15 participants each. The CONTROL-group received no information about the meaning of the word they were about to hear, while the participants in the other groups were informed that it meant BIG, SMALL, ROUND or POINTY respectively. The first participant in each chain was then audially presented with a word containing a wide range of different segments and asked to repeat it. Thereafter, the recording of the repeated word was played for the next participant in the same chain. After 15 generations, the strongest results had been produced by the SMALL-condition, which correlated with previous studies linking high and/or rising frequencies of vocalizations to small things. The general results were attributed to continuous versus dichotomous mirrorings between semantic and phonetic parameters, semantic poles not being equally iconically charged and the role of transmission and interaction in iconicity.

Study III and IV utilize color words to bridge the gap between cross-modal mappings and sound symbolic mappings in the lexicon. In Study III, the perceptual dimensions that drive sound-color correspondences were investigated by testing cross-modal correspondences between a range of visual (luminance, hue, saturation) and acoustic (loudness, pitch, spectral centroid, F_1, F_2, trill) dimensions through Implicit Associations Task experiments. Circa 20 participants with varying mother tongues were recruited online and were first taught a rule associating the right and left arrow buttons to one color and sound each. They were then presented with either color or sound stimuli and asked to press the correct arrow key as quickly as possible. By measuring the accuracy and reaction time, the results showed that loudness and pitch were implicitly associated with luminance and saturation but also that the actual hue of colors and the formants of vowels did not cause any robust associations. This suggests that underlying parameters are responsible for these associations, rather the characteristics of specific focal colors and phonemes.

Study IV follows up on the findings yielded by Study III but instead looks at phonetically transcribed color name data for eleven color words gathered from 245 language families. Each segment was described acoustically using high-quality IPA recordings and average color coordinates were extracted from a database consisting languages of 110 non-industrialized societies. Then, acoustic parameters (sonority, brightness, spectral centroid, F_1, F_2 and F_3 for vowels and sonority and spectral centroid for consonants) were correlated with the color words' visual parameters (luminance and saturation). Just as in Study I, vowels with high perceived brightness, sonority and F_1 were overrepresented in names of colors with high luminance, but an association between saturation and the sonority of consonants was also found. Evolutionary factors, such as the presence of similar mappings in chimpanzees, are discussed in conjunction with the results. In addition, notable similarities between the results and the cross-linguistic order of how color words are lexicalized suggests a link between which parameters are used for mapping sound to color iconically and which parameters influence how colors are organized in the mental lexicon.

2. Background

Core aspects of modern linguistics can be traced back to the structuralism shaped by Ferdinand de Saussure, in which human linguistic communication is analyzed via the underlying system of language (*langue*) rather than the use of language (*parole*) (Saussure 1959[1916]). The most central element of language, in this view, is the linguistic *sign*, which, simplified, is a linguistic unit that communicates a meaning. The sign is made up of the *signifier* (sound pattern, or phonetic/phonological form of a word) and the *signified* (the concept meaning) which are inseparable. Also central to the linguistic sign is *arbitrariness*, which means that there are no "natural" connections between corresponding sound patterns and concepts. For example, the concept TREE is reflected by the sound patterns [t i:] in English, [u˥˩] in Mandarin and [mti] in Swahili, but the involved sounds are not particularly "tree-like". There is therefore no reason a particular sound pattern should be attached to a particular concept since each of these three languages are equally apt at communicating the meaning TREE. This in essence, is because the language communities have agreed to use these sound patterns consistently for this concept.

However, there are a number of instances where this approach falls short, for example onomatopoeia: phonetically imitative words such as *cuckoo*, which display an obvious direct link between the sound pattern and concept. Indeed, non-arbitrary associations between sounds and meanings have been discussed and debated for more than 2,000 years. For example, in Plato's famous dialogue *Cratylus*, he argued for *the correctness of names* which included that [l] would be better suited for words representing liquid meanings because of its gliding manner of articulation and [o] would be most suitable for imitating roundness, etc. Furthermore, contemporary with Saussure's own most influential work, Jespersen (1922) wrote that "sound symbolism makes some words more fit to survive" since iconic words seem to resist sound change and that semantic domains connected to sensory perception (e.g. size and shape) are more likely to be non-arbitrary.

A few years later, this was followed by the first proper experimental studies on sound-meaning associations. Sapir (1929) constructed two nonsense words that differ only in vowel quality, /mil/ and /mal/ and then asked no less than 500 participants which of the two words meant a large table and which meant a small table. The results showed that an overwhelming majority (80%) thought that /mil/ denoted the small table, and this experiment was later followed up upon by Bentley & Varon (1933) who showed that [a] is perceived as larger and rounder than [i], and by Newman (1933) who found similar results but also investigated consonants and the bright-dark dimension. Similarly, Köhler (1929) constructed slightly more complex nonsense words, /takete/ or /baluma/ (later /maluma/), but instead asked participants which of the words matched best a roundish shape and a jagged shape respectively. The results showed a strong preference for pairing the roundish shape with /baluma/ and the jagged shape with /takete/. Thus, despite that language as a whole may be arbitrary to a large extent, there are notable exceptions suggesting that iconicity could influence how we communicate.

Iconicity also extends across different types of languages regardless of the modalities used to covey meaning. Signed languages primarily use the visual-spatial modality rather than the auditory but are bound by the same linguistic constraints and overarching structures as spoken languages, such as syntax and morphology. Hence, since humans generally communicate about what is visually perceived, signed languages are rich in direct iconic visual-to-visual mappings (Perniss et al. 2010). For example, in British Sign Language the sign for 'cry' is constructed by moving two extended index fingers in an alternating pattern downward from the eyes on the signer's face. In addition, signed languages also systematically and frequently use iconicity for non-manual features, e.g. modulating the mouth, face and eyes to change the size or shape of the reference (puffed cheeks and lip rounding). This further illustrates that the affordances tied to different meanings and modalities affect the distribution of iconicity (Dingemanse et al. 2015). In both spoken and signed languages, abstract concepts are generally hard to convey, while size and repetition are easy to convey. However, the modalities primarily used by spoken languages make expressing sounds and loudness an easy task but spatial relations and visual shapes are more difficult. For signed languages, on the other hand, the relationship is reversed.

2.1 Non-lexical sound-meaning associations

Synesthesia, the perceptual phenomenon in which stimulation from one sense can activate another, is reminiscent of iconicity. The most frequently reported types involve perceiving sequences, such as individual alphabetical letters, numbers, days of the week, etc., as colored. Another common type is perceiving that sounds evoke colors. This type of synesthesia is generally thought to be individual, but there is a tendency for mapping bright-sounding vowels ([i], [e]) to brighter colors, and dark-sounding vowels ([o], [u]) to dark colors (Marks, 1975; Miyahara et al., 2012; Watanabe et al., 2014). In addition, there is a range of less common types of synesthesia which can involve associating sounds with tactile sensations on a specific part of the body, with tastes and so on. Cross-modal associations, i.e. systematic correspondences between different modalities, are also frequently found in non-synesthetes. For example, visual angularity evokes responses from touch, hearing and vision in the form of hardness, pitch, and brightness (Walker, 2012), and high-pitched sounds have been consistently mapped to smallness, brightness, and high elevation (see Spence (2011) for an overview). In addition, there is extensive research showing correspondences between acoustic parameters, such as loudness, pitch and vowel formant levels, and visual parameters, such as luminance and saturation (Marks, 1974, 1987; Mondloch & Maurer, 2004; Moos et al., 2014; Hamilton-Fletcher et al., 2017). However, while it is not completely clear whether synesthesia is qualitatively or quantitively different from the strong associations that non-synesthetes can experience (Lacey et al., 2016; Spence, 2011), both phenomena affect our perception in a similar manner.

One of the more influential theories which could help account for these correspondences is Ohala's (1994) physiologically and functionally grounded *frequency code*, which links the fundamental frequency to body size and thereby maps size onto pitch. The explanation for this correlation is probably rather complex, since more recent studies have shown that the correlation between body size and fundamental frequency is rather weak (Taylor & Reby, 2010). Despite this, listeners still "incorrectly" associate lower pitch with greater size and strength (Bruckert et al., 2006; Collins, 2000; Sell et al., 2010). Accordingly, it is in many cases in animals' interests to appear large to get an advantage in potential confrontations. This can be achieved by erecting feathers or growling with low pitch to exaggerate the apparent size of the animal, and reversely, cowering and whining with high pitch can make an animal seem smaller and thereby indicate submissiveness. Thus, most animals perceive a low and/or falling fundamental frequency to indicate large size, authority, dominance, large distance, statements, etc., while high and/or rising fundamental frequency indicates small size, politeness, submission, proximity, questions, etc.

Moving back to more language-like stimuli, specific phonemes have been associated with a number of meanings. For example, Wisseman (1954) found that participants preferred to use [i] and [u] to imitate high-pitched and low-pitched sounds noises respectively, and voiceless plosives to imitate noises with abrupt beginnings. Likewise, [i] has been connected to acuteness, smallness, lightness, rapidity, speed, friendliness and closeness, [u] to thickness, darkness, sadness, bluntness and strength, stops to hardness, continuants to softness, [r] to roughness, strength and hardness, and [l] to smoothness, weakness and light-weight (Chastaing, 1958, 1965, 1966; Fonagy, 1963). Furthermore, following Sapir (1929) and Köhler (1929) a plethora of different versions of forced-choice matching experiments has shown consistent general associations between close, front, unrounded vowels and voiced obstruents and pointy shapes, and correspondingly between open, back, rounded vowels and voiced sonorants and round shapes (e.g., Davis, 1961; Holland & Wertheimer, 1964; Ahlner & Zlatev 2010; D'Onofrio 2014; Nielsen & Rendall 2011, 2012, 2013). The continuous interest in this subject also ultimately lead to Ramachandran & Hubbard's (2001) famous study which tied maluma-takete/bouba-kiki effect back to synesthesia through sensory features being coded in nearby brain areas. Regardless, it is evident that different senses and modalities are interconnected and can be utilized linguistically to convey meaning.

2.2 Vocal iconicity in the lexicon (sound symbolism)

Evidently, vocal iconicity is not limited to general, low-level cross-modal correspondences and unimodal imitations of surrounding sounds. Sound-meaning associations are, in fact, rather common and intergraded in the phonological and lexical levels of language.

2.2.1 Language-specific vocal iconicity

On a language-specific level, iconicity can in some cases be one of the dominant parts of the lexicon. *Ideophones*, also referred to as *expressives* and *mimetics*, are words that evoke sensory perceptions but usually differ from non-ideophones in the same language in regard to phonotactics and morphosyntax (Dingemanse & Akita, 2016). For example, Japanese *doki doki* can be translated as 'heartbeat' but also 'excitement' and can be used to evoke the feeling of having your heart racing in a heightened situation. This also demonstrates how reduplication can be used iconically for evoking an iterative or intense meaning (Dingemanse, 2011). However, although it has been shown that both adults and children can generalize the meaning of ideophones from unknown languages (Imai et al. 2008; Kantartzis et al. 2011; Lockwood et al. 2016a; Lockwood et al. 2016b; Iwasaki et al, 2017), ideophones are simultaneously highly grammatically integrated and comparable to more traditional word classes such as nouns and verbs. This means that despite ideophones can be understood cross-linguistically, they are ultimately language-specific to a large degree, which in turn illustrates how iconicity can operate in the interface between paralanguage and language (Dingemanse & Akita, 2016).

While ideophones constitute entire separate word classes in some languages and number in the thousands, they are much scarcer in a number of languages, for example in European languages, with the notable exception of Basque (Ibarretxe-Antuñano, 2006, 2017). However, several languages, use *phonesthemes* to evoke similar cross-modal associations that usually relate to hearing, vision and touch. These words include phonemes or phoneme-clusters that can be analogically used to coin new words within languages, are understood by speakers of the same language without prior knowledge (Carling & Johansson, 2014) and have been referred to as *conventional sound symbolism* (Hinton et al, 1994). For example, initial *gl-* in many Germanic words, such as English *glisten* and *glitter* is used for words with light-related meanings. While some phonesthemes correlate with cross-linguistic sound-meaning associations, such as English *-ump* 'rounded object or collection of objects' (Reay, 1994; Abelin, 1999), many, including *gl-*, seem to be less universally understood and could be even more language-specific than most ideophones. Thus, phonesthemes exemplify iconic usage that closely borders arbitrariness since even if one word of a phonestheme cluster is clearly iconically motivated, the link between the referent and the sound could be lost for other words belonging to the same cluster, as the phonemes are passed on primarily via analogy.

2.2.2 Small-scale cross-linguistic studies

Going beyond the complex largely language-specific systems, there are a number of smaller studies that have investigated sound symbolism by including either a larger number of languages or larger number of concepts. Among these, we find several which have shown that speakers of one language can deduce the meanings of oppositional word pairs, such as LARGE-SMALL, DARK-LIGHT, THICK-THIN, etc., from unknown foreign languages above chance level (Tsuru & Fries, 1933; Brown et al., 1955; Brown & Nuttall, 1959; Siegel et al., 1965; Gebels, 1969; Klank et al., 1971; Kunihara, 1971; LaPolla, 1994). In addition, language includes a number of semantically delimited clusters of meanings which are highly functionally interconnected but can also in some cases be iconically motivated. Deictic words and pronouns have gotten a comparatively large amount of attention and have in several cross-linguistic studies of varying scope been demonstrated to have a presence of sound symbolism. By looking at 136 languages, Ultan (1978) found evidence for sound symbolically encoded distance within languages' demonstrative systems. Generally, more close, front and unrounded vowels had a tendency to be found in proximal words, such as 'here' and 'this', as well as in diminutive affixes. Woodworth (1991) correspondingly found that in 13 out of 26 investigated languages, proximal words included vowels with higher second formant frequency than distal words, which was also confirmed by a later study by Traunmüller

(1994) and Johansson & Zlatev (2013). Traunmüller (1994) also found some evidence suggesting that first person singular pronouns tend to contain voiced nasals, while second person singular pronouns contain voiceless stops and/or dentals and sounds involving lip protrusion. Hence, the semantic comprehension of sound symbolic lexemes across languages, suggests that further and systematic study of cross-linguistic comparisons on a grander scale is a promising endeavor.

2.2.3 Basic vocabulary

Until recently, there has been a profound lack of comprehensive comparisons of occurrences of sounds in a larger number of meanings across languages, due to methodological limitations. Iconicity researchers have therefore turned to concepts that are semantically and functionally similar cross-linguistically, rather than looking at language-specific associations, cross-modal correspondences and paralinguistic imitations. These concepts ought to represent the most fundamental, and perhaps also ancient, subsection of the mental lexicon, and some of the most influential works on basic vocabulary have come from the search for true lexical universals. Among the more notable basic vocabulary lists, we find the so-called *Swadesh lists* (Swadesh, 1971), originally consisting of around 200 concepts but later reduced to the commonly used Swadesh-100 list. These lists were constructed to only include concepts that are cross-linguistically relevant and are used for assessing chronological and genealogical relationships between languages. Several shorter adaptations of the Swadesh list have also been constructed, for the purpose of yielding more accurate results when used for lexicostatistic and glottochronological analysis (e.g. Holman et al., 2008). Similarly, Haspelmath & Tadmor (2009) designed an alternative 100-item list based on concepts that were resistant to lexical borrowing. However, these lists' usefulness has ultimately been questioned since linguistic universals are, in the end, very difficult to prove. There have, however, been several attempts at finding the semantic core of language. For example, Goddard & Wierzbicka (2002) have attempted to find true semantic universals, or *semantic primes*, by finding indefinable expressions, i.e. meanings that cannot be reduced to simpler terms. Semantic, and possibly also cognitive, hierarchies among related meanings have been postulated by several others. For example, Berlin & Kay (1969), later developed by Kay & Maffi (1999), have found evidence for that color words are lexicalized according to a similar cross-linguistic order. Likewise, Viberg (2001) found a similar implicational lexicalization order for perception verbs in which those relating to higher (unmarked) modalities, such as 'to see', are more fundamental than relatively lower modalities, such as 'to hear', 'to feel', 'to taste' and 'to smell'. Dixon (1982:1-62) has also proposed a number of (possibly universal) semantic types of adjectives, in which the most fundamental types include DIMENSION ('large', ‘narrow’, etc.), AGE (‘young’, ‘new’, etc.), VALUE e.g. (‘good’, ‘proper’, etc.) and COLOR (‘white’, ‘light’, etc.). In sum, these studies demonstrate that considerable parts of the mental lexicon seem to adhere to - or have a preference for - more or less fundamental patterns, although these effects have been attributed to several factors (Haspelmath, 2008). However, there is evidently still no real consensus regarding which concepts that could be considered universal with certainty, and the items in these lists should therefore be viewed as compilations of universal tendencies. Nevertheless, at least currently, these concepts represent the most fruitful way for studying the core of the lexicon.

2.2.4 Large-scale cross-linguistic studies

As a result of the previous studies on basic vocabulary, combined with digitalization and the more powerful and accessible statistical analyzes in the last decades, increased lexical data availability has allowed some researchers to go far beyond previous small-scale studies. During the last decade, a handful of studies have been able to utilize these new possibilities by more adequately study sound-meaning associations across a very large number of languages and language families. Wichmann et al. (2010) investigated 40 basic vocabulary items in approximately 3,000 of the world's living languages and were able to show that the concepts BREAST, I, KNEE, YOU, NOSE, NAME and WE had non-random word shapes. By looking at the average relative frequencies of each sound for each position in the divergent words, they found several overrepresentations of sounds that word-wise formed interesting sound-meaning correlations. For example, BREAST was rendered as /muma/ in which the labial sounds could relate to the suckling of a child, KNEE, /kokaau/, contained both hard-sounding voiceless stops and rounded vowel, and NOSE, /nani/, unsurprisingly contained two nasal sounds. By expanding on Wichmann et al.’s study, Blasi et al. (2016) added another 3,000 languages and dialects to the dataset, used an improved statistical model and included further controls for interfering areal and genetical effects. Along with confirming the sound-meaning associations found by Wichmann et al., Blasi et al. found sound symbolic effects in a total of 30 concepts, such as rounded back vowels in words for ASH and trills in words for ROUND. In addition, they also analyzed underrepresentations of sound groups which could indicate a clear dispreferences for specific sound groups in certain concepts, such as voiced labials in words for TOOTH and open unrounded vowels in words for NOSE. Joo (2019), similarly investigated 100 basic vocabulary meanings in 66 genetically distinct languages based on the Leipzig-Jakarta List (Haspelmath & Tadmor, 2009). However, as opposed to previous studies, Joo specifically analyzed the words morphemically and the sounds based on phonetic features. Several of the found sound-meaning associations correlated with previous works, but due to the study’s new methodological approach, it also yielded more fine-grained results. Relatedly, Pagel et al. (2013) attempted to link together a number of unrelated Eurasian language families by investigating a set of meanings used frequently in everyday speech. While it could be questioned whether the results actually provided evidence for long-range relationships between these families, the study certainly showed that unrelated languages used phonetically similar word shapes for the same meanings in at least 23 cases. Thus, basic vocabulary items indeed seem to contain large amounts of sound symbolic material. More generally, this suggests that iconicity plays a role in language development, but on an individual level, it also suggests that it is beneficial for language learning. However, our current knowledge of its extent is restricted to just a portion of what could be considered basic vocabulary.

2.3 Iconicity in language evolution and language development

Obviously, iconicity has to be viewed as intuitive in nature, which also leads to the question of why iconicity exists. Several scholars have proposed that non-arbitrary associations might have played a role in how language evolved and how it develops over time. Aside from Ohala's (1994) frequency code, other animals also share more explicit sound-meaning associations with humans. There is evidence showing that chimpanzees are able to consistently map white tiles to high-pitched sounds and black tiles to low-pitched sounds (Ludwig et al., 2011). Furthermore, the same capability has been confirmed in toddlers, synesthetes and non-synesthetes (Mondloch & Maurer, 2004; Moos et al., 2014; Ward et al., 2006). This suggests that luminance-to-pitch mappings, probably along with other similar mappings, must have been present early in human evolutionary history before we split apart from our closest living relatives' lineage. Furthermore, it is therefore also plausible to believe that synesthesia and cross-modal correspondences are qualitatively the same phenomenon and can be linked to origin of sound symbolism (Ramachandran & Hubbard, 2001; Bankieris & Simner, 2015). From a diachronic perspective, some iconic words tend to resist regular sound change. For example, the previously mentioned onomatopoeic word *cuckoo* has not changed its vowel from [u] to [ ] (Jespersen, 1922), nor have the voiceless stops, through Grimm's Law, switched to voiceless fricatives (Joseph, 1987). Instead, it has maintained the pronunciation as [kuku] to keep the sounds close to the bird's call. Likewise, it has been shown that iconicity decays and reemerges historically in meanings prone to be affected by it (Johansson & Carling, 2015; Flaksman, 2017). It has also been shown that iconicity has a range of functional and communicative benefits (Tamariz et al., 2017), since iconic words are easier to learn (Imai & Kita, 2014), Iconic gestures enhance comprehension when used together with speech (Holler et al., 2009; Kelly et al., 2010) and iconic signs are recognized more quickly (Thompson et al., 2012; Vinson et al., 2015). However, in terms of language acquisition, iconicity is very common in child language but tends to be used less in adulthood (Massaro & Perlman, 2017; Fort et al. 2018). This is likely because there are not enough unique individual iconic signals available through sounds or gestures passed a certain vocabulary size (Westbury et al., 2017). This suggests that in less developed and lexically poor stages of linguistic systems, iconicity is crucial for intuitively linking signals to referents but with higher competence, arbitrariness also becomes important as it enables us to cope with our communicative needs (Perniss & Vigliocco, 2014; Dingemanse et al., 2015). Thus, iconicity is important for how language evolves and develops, and by studying iconicity we can also learn great deal about the functions and constraints of human language.

2.4 Research questions

The aim of this dissertation is to explore the boundaries of sound symbolism in the lexicon from a typological, functional and evolutionary perspective. In order to achieve this, the subject in question has be triangulated by being investigated using several different methodologies and from a number of aspects.

Firstly, basic vocabulary, i.e. meanings that denote fundamental concepts relevant for more or less all languages, ought to be the most relevant start off point due to their cross-cultural consistency and possible historical and evolutionary relevance (Swadesh, 1971; Haspelmath & Tadmor, 2009; Goddard & Wierzbicka, 2002). These lexical items could then function as a miniature, albeit not perfect but easily attainable, version of the human mental lexicon. With this in place, it is possible to get an overview of which basic vocabulary meanings are sound symbolic. This approach is crucial, since it will allow us to systematically specify which semantic and phonetic features are involved in these mappings between sounds and meanings. While a number of studies have previously investigated basic vocabulary word lists cross-linguistically, the scope of these studies has been very limited, usually only including less than 50 meanings with some exceptions (Wichmann et al., 2010; Blasi et al., 2016; Joo, 2019). Thus, we do not know if these studies give a completely accurate picture of the degree of sound symbolism in basic vocabulary, since a number of relevant lexical items have not previously been investigated.

Secondly, the next question that needs to be answered is how sound symbolical features are introduced in language and how they interact with it. One of the few universal constants of human language is its tendency to undergo perpetual change. This means that words are in a constant battle for survival which can be greatly influenced by the presence of sound symbolic components (Jespersen, 1922). Thus, in order to keep the motivated mappings between sounds and meanings, sound symbolic items either have to resist sound change, being replaced by another sound symbolic item or decay only to later be reestablished (Johansson & Carling, 2015; Flaksman 2017). It is therefore important to be able to describe the process of how non-sound symbolic words can become sound symbolic through language usage. In order to do this, we have to study how the transfer of linguistic signals (words), in combination with concept meanings that are known to be sound symbolically motivated, can introduce sound-meaning mappings by subconscious bias. This also has to include a large number of language users in an experimental environment that as closely as possible corresponds to a sped-up version of historical (and to some extent evolutionary) natural language sound change processes (Kirby et al, 2015).

The last question to be answered revolves around the cognitive and linguistic level upon which sound symbolic mappings lie. There is a wide array of different types of mappings between senses, ranging from typical sound symbolic mappings to synesthetic and more basic cross-modal mappings (Marks, 1975; Westbury et al., 2017). Furthermore, several of the most fundamental mappings, such as those between light and pitch, have been found in both toddlers and chimpanzees despite their lack of linguistic competence (Mondloch & Maurer, 2004; Ludwig et al., 2011). This suggests that sound symbolism appears to be built on a more rudimentary foundation and it is therefore important to study the interface between underlying processes (cross-modal mappings) and language proper (sound symbolism). In order to investigate this, it is reasonable to start with a smaller set of meanings that can be easily analyzed cross-modally and sound symbolically. Just as in the lexicon as a whole, there are several semantic networks in basic vocabulary, such as kinship terms, deictic terms and, of interest for this dissertation, color terms. In addition, color terms have been shown to be sound symbolic and to constitute one of the main types of synesthesia as well. With a fitting subject of study in place, it is crucial to test for implicit associations between color and sound on a perceptual level. This is because in implicit tasks, the results are less likely to be affected by cultural factors and personal history which could distort the results greatly if not controlled for (Lacey et al., 2016; Miyahara et al., 2012; Parise & Spence, 2012). The results yielded from a study of the implicit, underlying parameters that drive associations between sound and color then make it possible to conduct a follow-up study which accurately investigates how the processes and mappings operate in natural languages. Therefore, this follow-up study had to include a number of cross-linguistically comparable color concepts (with accompanying color coordinates and values) from a large number of sufficiently sampled languages. These two interconnected studies do not only give us an idea about how sound symbolism is connected with lower-level mappings, but they also yield and example of how other non-arbitrary semantic networks in basic vocabulary can be thoroughly investigated. Thus, this dissertation attempts to contribute to a better understanding of the following research questions:

1. To what extent is the core of the lexicon affected by sound symbolism? This is the subject of Study I.
2. How do sound symbolic mappings emerge and develop under natural language simulation? This is the subject of Study II.
3. What is the cognitive depth of sound symbolic mappings? This is addressed by the case studies in Studies III and IV.

3. Methods

This dissertation attempts to get a comprehensive view of sound symbolism in human language. In order to achieve this, the featured studies include a range of different methodological approaches. Human language involves both individual speakers but also the overarching interconnectedness that makes it mutually intelligible for the speakers. In addition, sound symbolism seems to be present in all languages to some degree and we can therefore assume that this was the case historically as well (Blasi et al, 2016). Therefore, this dissertation places equal focus on both language users (experiment participants) and on language systems (natural spoken languages) by including two studies that compare a large number of sampled languages and two studies that investigate language users' sound symbolic bias.

Similarly, it is desirable to achieve a balance between investigating phonetic forms of lexemes and more fundamental acoustic parameters, because while lexemes are the primary bearer of acoustic matter in human communication and interaction, there is a high chance that they will be subject to conventionalization and language-specific phonological constraints over time. Acoustic signals used in human language, such as pitch and loudness, on the other hand, are more closely connected to other animals' signaling system (Ohala, 1994) but also tend to be used paralinguistically to a higher degree than phonemes.

Lastly, since sound symbolism (and iconicity in general) seems thoroughly evolutionarily grounded, its interaction with language over time has to be studied. However, as we only have access to a couple of thousands of years of attested historical language material, elements of evolutionary and developmental simulation of natural languages are included as well in the form of an experiment conducted in the *iterated learning paradigm* (further explained below).

3.1 Language-based data (Studies I and IV)

For Studies I and IV, a large database consisting of sampled basic vocabulary meanings (or concepts) was constructed by systematically including various suggested lists of basic vocabulary and proposed semantic universals. The topic of linguistic universals is, in general, a complex one (Evans & Levinson, 2009) and it should be pointed out that there is no way of knowing exactly which concepts are present in all of the world's languages. Far from all languages are described and only a subsection of those which are described is easily attainable. Furthermore, it must be assumed that there are thousands of now extinct languages that have to have been spoken since the dawn of human language. We do not know anything about these languages and even well-described languages today differ considerably in how they assign semantic boundaries to concepts. However, pooling the existing knowledge of basic vocabulary concepts which at least seem to be present in a great many of the world's languages ought to capture a large portion of the core mental lexicon.

The next step is to elicit these concepts from actual languages. However, when gathering lexical material from a large number of languages, there is always a risk of genetical interference because related languages tend to be similar to each other. Considering that the world's around 6,500 languages are distributed between around 400-450 language families of varying size (Hammarström, 2015), the optimal option was to exclude the genetic component completely by only including one language per language family. There are also various classifications of language families which is why the more conservative approach for grouping languages into families, provided by Glottolog (Hammarström et al., 2017), was deemed the most beneficial for the purpose of these studies. However, due to data availability, only 245 of the 419 language families (58.5%) listed on Glottolog could be attained for the present database. The main drawback of only including one language per language family is that it limits the dataset to a couple of hundred languages. Other large-scale cross-linguistic studies that have included a very large share of recorded languages in their analyses, such as Blasi et al. (2016), have attempted to compensate for genetic and areal bias by a series of statistical tests. While this approach limits biases, it does not exclude them and including one language per language family was therefore deemed the safer option.

The ideal method for obtaining the lexical material would be to conduct fieldwork with native speakers of each language. However, gathering the data in this manner for 245 languages across the world was impossible considering the scope and timeframe of this dissertation. Thus, existing lexical compilations of languages and regular dictionaries were used, as well as grammars and grammar sketches for some under-described languages. While it could be argued that the data obtained from these sources is less reliable than fieldwork data, at least in terms of polysemy and synonymy, the employed method compensates for this by being more time-efficient and thus allows for a larger dataset. And although some concepts could not be retrieved from all languages due to the varying quality of data, the sample remained unbiased since only one language per language family was included.

3.2 Participant-based data (Studies II and III)

Most comparable experimental studies on sound symbolism and cross-modality include around 20 to 30 participants that are recruited on university campuses (e.g. Hamilton-Fletcher et al., 2017) which yields a very homogeneous group of language users that cannot be considered representative for the entire population. Hence, just as in Studies I and IV, large comparative datasets were deemed to be crucial for Studies II and III as well despite that this required the recruitment of a considerable number of participants. This was particularly important for Study II since the methodology has to be considered an uncharted territory, at least when it comes to studying sound symbolism. The most accessible way of achieving this is to recruit participants through online international crowdsourcing platforms and conducting the actual experiments online. This also allows for samples of participants with a range of different mother tongues and from different walks of life. However, while recruiting and conducting the experiments online introduce potential control issues, such as the participants not adequately understanding and executing the tasks, it also makes it possible to include thousands of participants in the same study. The studies were conducted under established ethical standards for Lund University and University of Edinburgh.

3.3 Procedures

With data sources identified for Studies I and IV, the data collection was a fairly simple but time-consuming process. Only a very small portion of the retrieved data was transcribed phonetically which meant that the remainder was collected with incompatible orthographic or database-specific transcription systems. Therefore, a considerable amount of time was devoted to creating a unified transcription system that was able to both capture the diversity of various phonemic systems and quantify these systems to be comparable. The created system is similar to International Phonetic Alphabet (IPA) with some minor but crucial differences. For example, sounds that incorporate more than one place of articulation (e.g. [k]) were split into two segments in order to be quantified separately. This was done primarily to correspond to how sound symbolism is used in language, i.e. being able to capture separate features of sounds instead of viewing them as a single sound. This allows these features (which are partly phonemes, partly acoustic representations) to be grouped and quantified in a more appropriate way. These features could then be grouped according to salient articulatory parameters in conjunction with distinctive acoustic features which are sound symbolically relevant.

For experiments (Studies II and III), the main aim for the data collection was to be minimalistic, i.e. it had to be brief and easy to understand from the participants' point of view. Study II was designed to investigate how words can be shaped by cognitive biases through cultural evolution. The most fitting methodology for studying this type of biases is the *iterated learning paradigm* (Kirby et al, 2015) which involves some form of information (for example a word) to be transmitted from one participant to another over multiple *generations* which, in turn, forms a *transmission chain*. This experiment only required the participants to, via his/her phone or computer, be told a certain meaning ('big', 'small', 'round' or 'pointy'), then listen to a word and repeat it. This simple setup, combined with the large number of included participants, allowed both for very efficient data collection and robust results. Study III, on the other hand, was designed as a more conventional set of experiments and involved matching two pairs of colors and sounds to see if the participants preferred to match certain visual parameters to certain acoustic parameters. However, the experiments tested for implicit associations, instead of explicit associations, in order to prevent the participants from being aware of the investigated cross-modal correspondences. As stated above, implicit tasks have the advantage of yielding results which are less likely to be affected by cultural norms or idiosyncratic personal preferences and are therefore also used by, for example, psychologists to study social prejudices and biases. In addition, the sound stimuli were created using formant synthesis to make them as natural-sounding as possible which also enabled us to manipulate one acoustic feature at a time (Anikin, 2018). The main disadvantage of this design was that only two pairs of colors and sounds could be compared in each experiment. However, this was compensated for with the large number of participants included.

3.3 Analyses

Bayesian mixed (multilevel) regression models were judged the most adequate method of analysis for all four studies. This was particularly true for Study I due to the very large dataset and number of analyzed sound groups. The main reason for this is that, compared to frequentist mixed models, Bayesian mixed models are more flexible in terms of model structure, can incorporate meaningful prior knowledge and can provide unambiguous estimates of uncertainty (Kruschke & Liddell, 2018).

The goal for Study I's data analysis was to identify words with overrepresented sound groups. However, calculating the absolute number of segments occurring within a word could skew the results through, for example, reduplication and effects of word length. Thus, the data was analyzed as proportions rather than absolute counts of sound groups calculated separately for vowels and consonants. The proportions for each of the ten evaluated sound groups were then analyzed using Bayesian generalized linear models. Then, in order to identify cases of over- or underrepresentations, fitted average proportions of each sound group across all words (concepts) were extracted and compared to per-word estimates.

The statistical model for Study II was very similar to the one used for Study I. Proportional values were used, and vowels and consonants were calculated separately since it is possible that some transmission chains might utilize vowels iconically, while others might utilize consonants. For example, if a particular meaning is mapped to high frequency sounds, the sound could be voiceless consonants, front unrounded vowels, or both. Binomial mixed models were then used and to account for non-independent nature of observations from the same transmission chain.

For Study III, the data generated from the implicit association tests was analyzed both for accuracy and response time. Two Bayesian mixed models of the same structure were fit for each experiment. A logistic model predicted the accuracy and a log-normal model predicted response time in the correct trials.

Similarly, statistical analyses for Study IV were also performed using Bayesian mixed models in which the unit of analysis was a single segment from the word for a particular color in one of the sampled languages. The task was to predict the acoustic characteristics of each segment from the luminance or saturation of the color. Thus, for each acoustic characteristic, the trend driven by a visual predictor could be estimated while also allowing individual colors to be associated with acoustic properties and allowing the effect of the visual predictor to be language-specific.

4. The conducted studies for the dissertation

The four empirical studies that constitute the dissertation attempt to give a deeper understanding of the role sound symbolism plays in human language. Study I investigates basic vocabulary items in a large number of language families in order to establish the extent of sound symbolic items in the core of the lexicon, as well as how the sound-meaning associations are mapped and interconnected. Study II explores how sound symbolic associations emerge in arbitrary words through sequential transmission over language users. Study III and IV use color words to investigate differences and similarities between low-level cross-modal associations and sound symbolism in lexemes. Study III explores the driving factors of cross-modal associations between colors and sounds by experimentally testing implicit preferences between several different acoustic and visual parameters. Study IV investigates sound symbolic associations in words for colors in a large number of language families by correlating acoustically described segments with luminance and saturation values obtained from cross-linguistic color-naming data.

4.1 Study I

Study I investigates the extent of sound symbolism in basic vocabulary cross-linguistically, but also which types of sound symbolism are used and how these findings can relate to human cognition more broadly. Despite the resurgence of interest in the field of iconicity and the consensus that this is a universal phenomenon, very few cross-linguistic studies include more than a handful of languages or more than twenty meanings. Many also lack phonetic distinctions that are commonly utilized in sound symbolic mappings. Study I therefore focuses on the phonetic and semantic features involved in sound symbolism by gathering 344 meaning concepts with claimed universal tendencies, from 245 language families. The concepts are sampled to represent the core of the lexicon based on existing lists of basic vocabulary and to completely control for genetic bias in the language sample, only one language per language family is included. The segments of the linguistic forms are then systematically grouped according to phonetically and sound symbolically relevant sound groups. Then, by using a set of Bayesian generalized linear models, the data can be analyzed to establish cross-linguistic over- and underrepresentations of sound features in the investigated concepts.

The results show that, using a conservative estimate, 125 robust associations between sounds and meanings spanning 59 concepts adhering to a range of different semantic domains can be established. Thus, several concepts can be linked to more than one sound group through sound symbolism which means that the combinations of associated sound groups play a key role for grounding the mappings between sounds and meanings. Among the established associations, many are onomatopoetic concepts (e.g. BLOW) but interestingly, many other types of concepts are found to be equally robust. This study therefore illustrates that placing focus on correlations between semantic and phonetic features, rather than on specific words and phonemes, is a more appropriate way of investigating sound symbolism's universal yet flexible structure.

In addition, the established associations can be correlated with at least 16% of the items from commonly used basic vocabulary lists. Hence, large parts of these lists are affected by sound symbolism which could greatly impair their ability to determine genetic relationships. The high incidence of sound symbolism found in core vocabulary could be explained by the functional and communicative benefits of sound symbolism and iconicity (Tamariz et al., 2017), since it has been shown that iconic words are easier to learn (Walker et al., 2010; Imai & Kita 2014; Massaro & Perlman, 2017) and iconic gestures used together with speech can enhance comprehension (Holler et al., 2009; Kelly et al., 2010). Hence, iconicity seems to have a strong scaffolding or bootstrapping effect on language, and even though iconicity is more common early in language acquisition (Massaro & Perlman, 2017; Perry et al., 2017), sound symbolism in core lexicon remains prevalent throughout adulthood.

Furthermore, the established sound symbolic concepts with noteworthy overrepresentations can be grouped into 20 semantically and phonetically superordinate concepts, referred to as *macro-concepts*. The actual mappings can also be analyzed and grouped into four types of sound symbolism (Dingemanse, 2011; Carling & Johansson, 2014). These include *onomatopoeia* (unimodal imitative mappings), *vocal gestures* (cross-modal imitation through other senses than hearing), *relative* sound symbolism (grounded in relational mappings between the ends of parallel semantic and phonetic scales) and *circumstantial* mappings (based on associations between referent in specific events in which some sounds are frequently produced).

Study I, thus, acts as a foundation for this dissertation by yielding a very large dataset and a thorough sound symbolic analysis of the featured lexemes. It shows that, using a robust feature-based approach that kept genetic and areal bias to a minimum, sound symbolism is an influential force in in the core of the lexicon. In addition, grouping the semantic and phonetic features involved in the sound-meaning associations into macro-concepts gives us an idea of which lexicalized semantic domains could have been present at the dawn of human language. Furthermore, aside from the two previously well-described types of sound symbolism; onomatopoeia and relative sound symbolism, two new types; vocal gestures and circumstantial sound symbolism, are established. Lastly, the results yielded by Study I show that different types of sound symbolism tend to involve different mappings between sounds and meanings. This suggests that despite the richness of linguistic variation in human language, sound symbolism is a crucial and substantial part of our communicative system which can help us understand the mental lexicon and language diversity more broadly.

4.2 Study II

While we know that non-arbitrary associations between sound and meaning occur throughout the core of the lexicon (Dingemanse et al., 2015; Blasi et al., 2016), we do not know how these patterns enter languages. Study II, therefore, uses a simple experimental approach to study how the cultural transmission of a single artificial and arbitrary word can introduce sound symbolic elements that correspond to cross-linguistic sound-meaning associations.

The earliest proper experiments on sound symbolism have involved *oppositional* (or *relative*, see Study I) sound symbolism. For example, among the earliest studies, Sapir (1929) showed that people have a preference for associating SMALL with /i/ and BIG with /a/ and Köhler (1929) showed that voiced and rounded sounds are preferred when referring to round shapes, while unvoiced and unrounded sounds are preferred when referring to pointy shapes. During the almost one hundred years of studying the so called *bouba-kiki phenomenon*, a number of aspects have been investigated (Lockwood & Dingemanse, 2015), including the involved cognitive semiotic processes (Ahlner & Zlatev, 2010) and underlying biases of the actual experiments (Nielsen & Rendall, 2011, 2012). However, the experimental setup has remained much the same, in which participants were asked to associate meanings with a set of words or syllables that were predefined. Since each individual participant was asked to combine meanings with sounds that may or may not adequately fit his or her intuition or phonology, the results could yield a much coarser picture than what is required for understanding how sound symbolism emerges. In our study, we investigate these involved cognitive biases through a methodological approach that focuses on the transmission of sound symbolism through the language filters of participants, but which also excludes orthographic influence as much as possible.

In order to achieve this, methods that are specifically designed to study how languages change over time are selected, i.e. the *iterated learning paradigm* (Kirby et al, 2015). In iterated learning studies, information, in this case words, is transmitted from one participant (the teacher) to another (the learner). This is then done for a number of *generations* of learners which together form a *transmission chain*. Throughout the transmission process, some information tends to be lost, which causes the information to change depending on the learner's cognitive biases. Thus, since sound-meaning associations seem to occur in all languages, iconicity must be regarded as a cognitive bias and iterated learning experiments can therefore be used efficiently for studying this phenomenon.

Thus, the methodological setup used for Study II is fairly simple. The participants are divided into five conditions (CONTROL, BIG, SMALL, ROUND and POINTY), presented with a recording of a single arbitrary seed word and asked to repeat it. The seed word is designed to not carry iconic biases in any established semantic or phonetic direction but is complex enough to be somewhat difficult to remember, to ensure that the word can evolve phonetically. The repetitions of the participants' utterances are recorded and then used as stimuli for the next participant in the same transmission chain. This process is then repeated for 15 generations of participants per transmission chain, there are 20 transmission chains for each of the five conditions and the participants are recruited online from across the world. In the CONTROL-condition, the word is passed down 15 generations without any introduced bias stimuli, but in the other conditions the participants are primed with a meaning connected to the word they hear. The meanings for the BIG- and SMALL-conditions are conveyed in text and the biases for the ROUND- and POINTY-conditions are conveyed through shapes presented visually. All audio recordings are transcribed into IPA and the transcribed sounds are then categorized according to six binary sound parameters; HIGH-LOW, FRONT-BACK and ROUNDED-UNROUNDED vowels, and GRAVE-ACUTE, VOICED-VOICELESS and SONORANT-OBSTRUENT consonants. Thereafter, binomial mixed models are used on the proportion of vowels or consonants of each particular sound parameter out of the total number of vowels or consonants in the word for generation 0 through 15.

Study II's results show that FRONT vowels decrease in the CONTROL-, POINTY- and ROUND-conditions, while ROUNDED vowels increase in the ROUND-condition and decrease in the SMALL-condition. In addition, GRAVE consonants decrease in all conditions and VOICED consonants increase slightly in the ROUND-condition.

However, when the stimuli-condition values are compared to CONTROL-condition values, FRONT vowels increase in the SMALL-condition which is also mirrored by a decrease of ROUNDED vowels, and there is a slight decrease of GRAVE consonants in the POINTY-condition. Thus, both vowels and consonants can be confirmed to be involved in size and shape iconicity (Ahlner & Zlatev, 2010; Nielsen & Rendall, 2013; D'Onofrio, 2014).

The strongest results are produced by the SMALL-condition whose associated sound groups align well with Ohala's (1994) *frequency code* which predicts that smallness, as well as related meanings, is evoked by high and/or rising frequencies of vocalizations. Another interesting finding is that the SMALL-condition, which belongs to the continuous SIZE-domain, is appropriately mapped to the continuous frequency scale, while the POINTY-condition is associated with sounds through non-continuous tactile mappings. Hence, the mappings incorporate felicitous and correlating semantic and phonetic features. Furthermore, the SMALL- and POINTY-conditions produce several iconic effects while the BIG- and ROUND-conditions do not, which could be explained by the fact that semantic poles might not be equally iconically charged (Nielsen & Rendall, 2011; Jones et al., 2014; Tamariz et al., 2017) and that semantic markedness might aid learnability (de Villiers & de Villiers, 1978; Paradis, Willners & Jones, 2009).

The perhaps most important finding, however, is that transmission of signals (words) seem to be sufficient for iconicity to emerge. Some previous studied have indeed shown that iconicity emerges through transmission but only with the use of text-based artificial languages or forced-choice experimental design (Jones et al., 2014). Others have argued that iconicity emerges through communicative interaction because of the increased number of possible innovations which could introduce iconic labels for meanings (Tamariz et al., 2017). Study II therefore shows that while interaction could provide an even more advantageous environment for iconicity and sound symbolism, it is not a prerequisite. In sum, by flipping the classic bouba-kiki experiment on its head and including a very large number of participants, as well as an auditorily modest linguistic environment, we are able to get a deeper understanding of how iconicity operates within the semantic SIZE- and SHAPE-domains.

4.3 Study III

Study III looks at low-level cross-modal associations between the underlying parameters of sounds and colors. This is the first of two case studies on non-arbitrary mappings between sounds and colors, the other being Study IV. Words for colors were deemed particularly relevant for a case study since they belong to a conveniently delimited semantic network and are among the most fundamental descriptors in basic vocabulary. Furthermore, they are shown to be sound symbolic and are frequently involved in synesthesia which makes them a suitable subject for studying both sound symbolism and cross-modal associations.

There is extensive previous work on how people seem to automatically and consistently correlate colors with sounds and other senses, such as high-frequency sounds with brightness and auditory loudness with visual luminance (Marks, 1974; Root & Ross, 1965; Marks, 1987). For example, pitch has been reliably associated with luminance (Mondloch & Maurer, 2004; Ward, Huckstep, & Tsakanikos, 2006), however, these associations could, at least in part, be caused by accompanying changes in loudness. What makes this matter even more complicated is that the experimental designs of many previous studies have involved matching language-specific focal colors to phonemes, both of which carry a number of intertwined underlying parameters. Study III attempts to resolve some of these remaining issues through three methodological advances.

Firstly, Study III controls for visual confounds by not working with focal colors or subjective color space but instead using perceptually accurate CIE-Lab color space (Kim et al., 2017), since this preserves subjective distances between colors while also separating lightness, hue and saturation. For example, while several studies have linked higher pitch to yellow (Orlandatou, 2012), focal yellow is also the brightest color (Witzel & Franklin, 2014) which makes it unclear whether the hue, luminance or saturation of the color is the cause of the association. Thus, uniform color pairs are selected which only differ along one dimension in the CIE-Lab space; high-low luminance, green-red hue, yellow-blue hue or high-low saturation.

Secondly, Study III similarly controls for acoustic confounds by breaking down speech sounds to a set of independent acoustic properties. In addition, in order to control for idiosyncratic effects from human recordings as well as to be able to control for each acoustic feature separately, synthesized, yet speech-like, sounds are created (Hamilton-Fletcher et al., 2017; Kim et al., 2017). The investigated acoustic features are chosen based on previously reported sound-color correspondences and include loudness, pitch and energy spectrum, as well as the frequencies of the first two vowel formants and the typologically most common trill.

Thirdly, Study III tests for implicit, rather than explicit, associations between sounds and colors because an explicit matching of sounds to colors fails to look for sound-color associations at a lower perceptual level. Thus, a web-based association test (Parise & Spence, 2012) is implemented for each combination between acoustic and visual parameters. The task is to learn a rule associating the left arrow key with one color and sound and the right arrow key with another color and sound. Then, through a series of blocks of trials, the participants are presented with one color or sound at the time and asked to press the corresponding arrow as quickly as possible. The colors and sounds corresponding to each key are then switched around randomly for each block of trials and both accuracy and response time is recorded. Around 20 participants are recruited online for each experiment, the study is performed online, and the statistical analysis is conducted in the form of two Bayesian mixed models of the same structure fit for the accuracy and the response time.

The results of Study III show that higher luminance (light vs. dark gray on white background) is associated with lower loudness, higher pitch, higher spectral centroid, and the presence of a trill. However, there is no reliable association between luminance and the frequency of the first two formants, or the green-red and yellow-blue hue contrasts. In addition, high (vs. low) saturation is associated with greater loudness, higher pitch, and higher spectral centroid, and the trill is weakly associated with low saturation based on the response time. One of the more surprising results is that light gray is associated with low loudness and dark gray with high loudness. However, this seems to be an effect of the contrast between the two gray stimuli and the white background rather than of lightness or luminance as such (Hubbard, 1996; Marks, 1974, 1987). In addition, the association between higher pitch and light rather than dark gray suggests that one mechanism is responsible for cross-modal correspondences between luminance and loudness, but another for luminance and pitch. Thus, Study III indicates that luminance-loudness associations are *prothetic* (quantitative) and driven by visual and auditory saliency, while luminance-pitch associations are *metathetic* (qualitative). In contrast to luminance, high saturation is associated with high loudness and high frequency, possibly because mappings between auditory frequency and different modalities could vary in strength. This would mean that stronger associations could override weaker ones. Perhaps the most crucial finding is that formant frequencies and specific color hues are not found to have any associations the acoustic or visual parameters. Study III thus demonstrates the strength of breaking down cross-modal associations to their components in order to tease apart the driving factors.

In sum, Study III shows that by isolating relevant visual and acoustic parameters, while at the same time keeping them fairly close to natural speech sounds and the colors, the interaction between perception, language, and cognition can be studied in greater detail. Most of the findings correlate with previous studies, but the study is able to establish relatively simple quantitative dimensions (luminance, saturation, loudness and frequency), rather than qualitative ones (hue and vowel quality) as the primary components in sound-color mappings. In addition, the distinction between prothetic and metathetic mappings is shown to play a crucial role in how we associate sounds to colors, which deepens our understanding of how iconic associations are grounded and operate on semantic, phonetic, semiotic, and cognitive levels.

4.4 Study IV

Study IV investigates how sound symbolism affects color words in a large number of genetically and areally sampled languages. This is done by building on results from Study III which show that underlying acoustic (luminance and saturation) and visual parameters (loudness and frequency) are the driving forces behind cross-modal sound-color correspondences. Thus, it is likely that these correspondences play a role in lexical color sound symbolism as well.

Previous cross-linguistic studies on sound-meaning associations in color lexemes are few. In addition, sound-color associations yielded from these studies have been limited to phonetic similarity across several language families in words for BLACK (Pagel et al., 2013) and NIGHT (Wichmann et al., 2010), and overrepresentations of rhotics in RED (Blasi et al., 2016). However, some languages have well-developed, language-specific color sound symbolism systems. For example, in Korean (Rhee, 2019) base color words (i.e. WHITE, BLACK, RED, GRUE and YELLOW) can be expanded to hundreds of color words through systematic alterations between vowel harmony, consonant tensing and morphological processes which changes the luminance and saturation of the color words.

Since study III shows that loudness and frequency are the acoustic properties with the greatest effect in sound-color correspondences, these parameters have to be translated into speech sounds. However, while frequency is used in several ways throughout languages, loudness is not used phonemically, and we therefore need to find a suitable proxy. For this purpose, sonority or perceived loudness is used, since it is one of the most salient properties of segments cross-linguistically. Furthermore, since we also want to see if luminant colors contain segments that are perceived to be relatively "brighter" (Ludwig & Simner, 2013; Walker et al., 2010; Walker, 2012), a small pilot study is conducted to understand what acoustic measure corresponds best to perceived "brightness". In the pilot, the participants are shown IPA symbols with corresponding audio in a random order and have to arrange them along a horizontal scale from darkest-sounding to brightest-sounding. The results show that perceived brightness is based on upward shifts of the first three formants for vowels and that spectral centroid seems to be the best proxy of perceived brightness for both vowels and consonants.

For the primary investigation of Study IV, lexical data is obtained from the same database as Study I. From this data, eleven color concepts are selected based on color opponency, RED-GREEN, YELLOW-BLUE, BLACK-WHITE, as well as GRAY and the semantically related concepts NIGHT-DAY and DARK-LIGHT. Then, in order to make the data in text form comparable with acoustic measurements, IPA recordings of individual typologically common and frequently occurring segments are obtained (Lawson et al., 2015).

The luminance and saturation values of the color words are obtained from CIE-Lab coordinate averages based on cross-linguistic color-naming data (Regier et al., 2005), and NIGHT-DAY and DARK-LIGHT are arbitrarily assigned luminance values while excluded from the analyses involving saturation. Statistical analyses are then performed using Bayesian mixed models for the acoustic parameters (sonority, brightness rating, spectral centroid and first, second and third formants for vowels, and sonority, brightness rating, spectral centroid for consonants) and color parameters (luminance and saturation).

The results of Study IV reveal that the luminance of a color and the sonority of vowels in the word for this color is significantly associated. Likewise, luminance predicts the subjective brightness of vowels, as well as frequency of the first vowel formant. Reversely, sonorous consonants are overrepresented, albeit weaker than the effects for vowels, in words for both luminant and saturated colors. Thus, there is a tendency for both bright and sonorous vowels to occur in the words for light colors, while there also appears to be a tendency for sonorous consonants to occur in words for more luminant and saturated colors.

Study IV shows that acoustic characteristics of vowels and color luminance produce the strongest associations which could be attributed the fact that vowels are articulatorily more gradient than consonants and thus more similar to visual parameters. Furthermore, the results show some interesting similarities with lexicalization patterns (the process of adding lexemes to the lexicon) of color words. It has been shown that the primary color distinction present in a large number of sampled languages separates light/warm colors (WHITE, YELLOW, RED) from dark/cool colors (BLACK, BLUE, GREEN) (Kay & Maffi 1999). The next distinction, which was also present in almost all of the investigated languages, separated the light color WHITE from the warm (saturated) colors, YELLOW and RED. These patterns correlate well with the results in Study IV, as luminance produced the strongest sound symbolic results and is the most fundamental parameter for distinguishing colors, followed by saturation which is used for the split between the most luminant color and the warm colors. In addition, albeit uncertain, Study IV demonstrates a tendency for vowels and consonants to have different sound symbolic functions (similar to Korean color sound symbolism). This suggests that primary acoustic and articulatory affordances are used to distinguish between perceptual contrasts and could therefore influence lexicalization processes.

Taking a step back, there is strong evidence that associations between luminance and phonetic dimensions are among the most fundamental types of cross-modal mappings, occurring in synesthetic and non-synesthetic people (Moos et al., 2014; Ward et al., 2006), toddlers (Mondloch & Maurer, 2004) and chimpanzees (Ludwig et al., 2011). Similarly, infants can distinguish between long and short wavelength colors (Adams, 1987), which, in turn, can correspond to saturation distinctions (Witzel & Franklin, 2014). It is thus plausible that fundamental sound-color correspondences evolved before our and our closest living relatives' lineages split apart, since they are probably utilized to discern important distinctions between features of objects quickly and easily. These findings further connect to the increased learnability created by iconic mappings (Imai & Kita, 2014; Massaro & Perlman, 2017; Nygaard, 2009). Accordingly, it can be assumed that the cross-linguistic prevalence of sound symbolism in color words has been perpetuated because it aids lexical acquisition, leading to a cultural transmission bias. In sum, Study IV shows that color sound symbolism seems to be grounded in evolutionary, environmental, biological and developmental constraints, and that color sound symbolism can help us understand how linguistic categories evolve and develop since semantic processing seems to be affected by fundamental cross-modal associations.

5. Conclusion and future work

This dissertation uses several different methodological approaches to get a more thorough understanding of why and how humans associate sounds with meanings. The first research question asks how much of the core of the lexicon is affected by sound symbolism. While additional studies need to be conducted in order to arrive at the exact answer to this question, Study I demonstrates that by including a larger number of basic vocabulary items than previous studies, several new sound-meaning associations can be established. However, the clearest indication for the actual extent in the core of the lexicon is that by aggregating the most frequently used basic vocabulary lists and correlating them with the results, at least a sixth of the total number of items is found to be sound symbolic. In addition, looking at the phonetic and semantic features that sounds and meanings are made up by not only makes it possible to define two new types of sounds symbolism, but it has also shown to be more suitable for studying sound symbolism than language-specific words and phonemes. In turn, this approach also makes it possible to establish macro-concepts which we argue can be used for identifying the first broad lexical fields in early human language. Study I thus contributes broadly to increasing research on language evolution and language acquisition.

The second research question asks how sound symbolic mappings emerge and develop under natural language simulation. Mimicking language development over time is notoriously difficult, however, by turning the typical bouba-kiki experiment on its head and letting an arbitrary seed word to be transmitted through speakers with conditioned semantics, Study II demonstrates that it takes very little for iconic effects to emerge. In essence, it shows that while interaction seems to offer the most felicitous environment for sound symbolism and iconicity, transmission of signals through disconnected language users is enough. Furthermore, the study also shows that meanings which are potentially more semantically marked produce the strongest effects and there are similarities between the internal semantic relationships of each oppositional pair and their respective associated sounds. Thus, these findings help us to understand how to study sound symbolism and how lexemes develop and interact historically, as well as evolutionary.

The third research question asks what the cognitive depth of sound symbolic mappings is. In order to address this issue, we use color words as a case study since these have been confirmed to produce both cross-modal and sound symbolic associations, as well as they belong to a neatly delimited semantic field. By investigating a number of visual and acoustic parameters through a series of implicit associations tests, Study III finds several associations that are in line with previous reports. However, when controlling for interconnected visual and acoustic confounds, neither specific hues nor specific vowels produce any notable effects. It is thus possible that previously reported associations between vowels and colors are dependent on luminance, saturation and general energy levels in the spectrum. Study IV then investigates similar visual and acoustic measurements in color words cross-linguistically. The study shows that luminance produces the strongest results and is primarily associated with vowels, while saturation is primarily associated with consonants. We argue that the correlation between these results and the visual traits is important for the way color words are lexicalized cross-linguistically and that these findings can have a great significance for our understanding of how linguistic categories have developed.

Thus, sound symbolism, at least in the case of color, can be reliably linked to cross-modal correspondences. In addition, the correlating results produced by these studies contribute to the large field of cross-modal studies by showing the benefits of implicit rather than explicit experiment, as well as controlling for the confounding factors many previous studies have overlooked. Furthermore, the results also help us to understand how lexical iconicity can be linked to low-level perceptual processes, and more broadly, they contribute to our understanding of the mental lexicon and how to study similar delimited iconic semantic domains.

These studies provide ample ground for a range of future studies. Concerning data, this dissertation attempts to include as large datasets as possible given time limitation. However, this is an area that could always be improved by for example including a larger number of languages without introducing genetical bias, and involving more participants, preferably with an even greater range of mother tongues. Study I shows that by including more basic vocabulary items, more sound-meaning associations are found, which suggests that it could be beneficial to increase the semantic breadth of future large-scale studies even further. There is, however, probably a limit for what could be considered cross-linguistically comparable concepts, but, on the other hand, we know that iconicity occurs in language-specific concepts as well. Therefore, it could be interesting to test the effects of iconicity in the transmission of loanwords to see if lexical forms carrying iconic material spread more easily. Furthermore, future studies should also focus on suprasegmentals, specifically on describing the relation between iconic segments and stress patterns and tone. As for future studies utilizing the iterated learning paradigm for studying sound symbolism, it would be interesting to study the development and interaction of individual segments across speakers in greater detail. This could be done by using a range of different arbitrary seed words or even letting some seed words start off as iconic in one direction (e.g. BIG) and then associating them with the semantically opposite meaning (e.g. SMALL) and studying the effects. The two case studies on color (Studies III and IV) illustrate how cross-modal correspondences can be linked to sound symbolism and thus provide a potential way forward to study sound-meaning associations in other modalities, such as taste, smell and touch, as well as in location/proximity. Furthermore, the role vowels and consonants (with associated acoustic parameters) play in these associations is still rather unclear and should therefore be investigated further.

References

Abelin, Å. (1999). *Studies in sound symbolism*. Gothenburg: University of Gothenburg dissertation.

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies, 38*(1/4), 298-348.

Anikin, A. (2018). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavoir Research Methods, 1-15*. doi: 10.3758/s13428-018-1095-7

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition, 136*, 186-195.

Bentley, M., Varon, E. (1933). An accessory study of phonetic symbolism. *American Journal of Psychology, 45*, 76-86.

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkley, CA: University of California Press.

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.

Brown, R. W., Black, A. H., & Horowitz, A. E. (1955). Phonetic symbolism in natural languages. *The Journal of Abnormal and Social Psychology, 50*, 388-393.

Brown, R., & Nuttall, R. (1959). Methods in phonetic symbolism experiments. *Journal of Abnormal and Social Psychology, 59*, 441-445.

Bruckert, L., Liénard, J.-S., Lacroix, A., Kreutzer, M., & Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society of London B: Biological Sciences, 273*(1582). 83-89.

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.

Chastaing, M. (1958). Le symbolisme de voyelles: significations des "i" I& II. *Journal de Psychologie, 55*, 403-423 & 461-481.

Chastaing, M. (1965). Pop - fop - pof – fof. *Vie et Langage, 159*, 311-317.

Chastaing, M. (1966). Si les *r* étaient des *l. Vie et langage, 173*, 468-472.

Collins, S. A. (2000). Men's voices and women's choices. *Animal behaviour 60*(6), 773-780.

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

Davis, R. (1961). The fitness of names to drawings: A cross-cultural study in Tanganyika, *British journal of psychology, 52*, 259-268.

de Villiers, J. G., & de Villiers, P. A. (1978). *Language Acquisition*. Cambridge MA: Harvard University Press.

Dingemanse, M., & Akita, K. (2016). An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics, 53*(3), 501-532.

Dingemanse, M. (2011). Ezra pound among the Mawu. In Michelucci, P., Fischer, O., & Ljungberg, C. (eds.), *Semblance and signification. Iconicity in Language and Literature, 10*, 39-54. Amsterdam: John Benjamins.

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.

Dixon, R. M. W. (1982). Where have all the adjectives gone? In Dixon, R. M.W. (ed.), *Where have all the adjectives gone? and other essays in Semantics and Syntax*, 1-62. Amsterdam: Mouton.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and brain sciences*, *32*(5), 429-448.

Flaksman, M. (2017). Iconic treadmill hypothesis. In M. Bauer, A. Zirker, O. Fischer, & C. Ljungberg (eds.), *Dimensions of Iconicity. Iconicity in Language and Literature, 15*, 15-38. Amsterdam: John Benjamins.

Fónagy, I. (1963) *Die metaphern in der phonetik*. The Hague: Mouton.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental science, 21*(5), e12659. doi: 10.1111/desc.12659

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behavior, 8*, 310-312.

Goddard, C., & Wierzbicka, A. (eds.) (2002). *Meaning and Universal Grammar: Theory and Empirical Findings* (2 volumes). Amsterdam & Philadelphia: John Benjamins.

Hammarström, Harald. (2015). Ethnologue 16/17/18th editions: A comprehensive review. *Language*, *91*(3). 723-737.

Hammarström, H., Forkel, R., & Haspelmath, M. (eds.) (2017). *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. http://glottolog.org

Hamilton-Fletcher, G., Witzel, C., Reby, D., & Ward, J. (2017). Sound properties associated with equiluminant colours. *Multisensory Research, 30*(3-5), 337-362.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive linguistics, 19*(1), 1-33.

Haspelmath, M., & Tadmor, U. (eds.) (2009). *Loanwords in the World's Languages: A Comparative Handbook*. Berlin &New York: De Gruyter Mouton.

Hinton, L., Nichols, J., & Ohala, J. J. (1994). Introduction: Sound-symbolic processes. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Holland, M. K., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills*, *19*, 111-117.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context?. *Journal of Nonverbal Behavior, 33*(2), 73-88.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., & Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica, 42*(3-4). 331-354.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology, 109*(2), 219-238.

Ibarretxe-Antuñano, I. (2006). Estudio lexicológico de las onomatopeyas vascas: El Euskal Onomatopeien Hiztegia: Euskara-Ingelesera-Gaztelania. *Fontes Linguae Vasconum, 101*, 145-159.

Ibarretxe-Antuñano, I. (2017). Basque ideophones from a typological perspective. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 62*(2), 196-220.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B*, *369*(1651). doi: 10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H.. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007). What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking. *Japanese-language education around the globe, 17,* 53-78.

Jespersen, O. (1922). *Language - its nature, development and origin*. London: Allen and Unwin.

Johansson, N., & Zlatev, J. (2013). Motivations for Sound Symbolism in Spatial Deixis: A Typological Study of 101 languages. *The Public Journal of Semiotics, 5*(1), 3-20.

Johansson, N., & Carling, G. (2015). The De-Iconization and Rebuilding of Iconicity in Spatial Deixis: A Indo-European Case Study. *Acta Linguistica Hafniensia, 47*(1), 4-32.

Jones, J. M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J., & Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 2459-2464.

Joo, I. (2019). Phonosemantic biases found in Leipzig-Jakarta lists of 66 languages. *Linguistic Typology, 0*(0). doi: 10.1515/lingty-2019-0030

Joseph, B. D. (1987). On the use of iconic elements in etymological investigation: some case studies from Greek. *Diachronica, 4*, 1-26.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science, 35*(3), 575-586.

Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American anthropologist, 101*(4), 743-760.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*(2), 260-267.

Kim, H. W., Nam, H., & Kim, C. Y. (2017). [i] is lighter and more greenish than [o]: Intrinsic association between vowel sounds and colors. *Multisensory Research, 31*(5), 419-437.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.

Klank, L. J. K., Huang, Y.H., & Johnson, R. C. (1971). Determinants of success in matching word pairs in tests of phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 10*, 140-148.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review, 25*(1). 178-206.

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

Kunihara, S. (1971). Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 10*, 427-429.

Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synesthesia strengthens sound-symbolic cross modal correspondences. *European Journal of Neuroscience, 44*(9), 2716-2721.

LaPolla, R. J. (1994). An Experimental Investigation into Sound Symbolism as it Relates to Mandarin Chinese. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Lawson, E., Stuart-Smith, J., Scobbie, J. M., Nakai, S., Beavan, D., Edmonds, F., Edmonds, I., Turk, A., Timmins, C., Beck, J., Esling, J., Leplatre, G., Cowen S., Barras, W., & Durham, M. (2015). *Seeing Speech: an articulatory web resource for the study of Phonetics.* University of Glasgow. http://www.seeingspeech.ac.uk/

Lockwood, G., & Dingemanse, M. (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

Lockwood, G., Dingemanse, M., & Hagoort, P. (2016a). Sound-symbolism boosts novel word learning. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 42*(8), 1274-1281.

Lockwood, G., Hagoort, P., & Dingemanse, M. (2016b). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra, 2*(1). doi: 10.1525/collabra.42

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *PNAS, 108*(51), 20661-20665.

Ludwig, V. U., & Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex, 49*(4), 1089-1099.

Marks, L. E. (1974). On associations of light and sound: Themediation of brightness, pitch, and loudness. *The American Journal of Psychology, 87*(1-2), 173-188.

Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin, 82*(3), 303-331.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. Journal of Experimental Psychology: *Human Perception and Performance, 13*(3), 384-394.

Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication, 2*(4). doi: 10.3389/fcomm.2017.00004

Miyahara, T., Koda, A., Sekiguchi, R., & Amemiya, T. (2012). A psychological experiment on the correspondence between colors and voiced vowels in non-synesthetes. *Kansei Engineering International Journal, 11*(1), 27-34.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 133-136.

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Crossmodal associations in synaesthesia: Vowel colours in the ear of the beholder. *i-Perception, 5*(2), 132-142.

Newman, S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology, 45*, 53-75.

Nielsen, A. K., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition, 4*(2), 115-125.

Nielsen, A. K., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67*(2), 153-163.

Nielsen, A. K., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 65*(2), 115-124.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*(1), 181-186.

Ohala, John J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Orlandatou, K. (2012). The role of pitch and timbre in the synaesthetic experience. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece* (pp. 751-758).

Pagel, M., Atkinson, Q. D., Calude, A. S., & Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences, 110*(21), 8471-8476.

Paradis, C., Willners, C., & Jones, S. (2009). Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon, 4*(3), 380-429. Amsterdam: John Benjamins.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*(3-4), 319-333.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B,* 369(1651). 20130300. doi: 10.1098/rstb.2013.0300

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology*, *1*(227), 1-15.

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science, 21*(3). doi: 10.1111/desc.12572

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Reay, I. E. (1994). Sound Symbolism. In Asher, R.E. (ed.), The *Encyclopedia of Language and Linguistics*, *8*, 4064-4070. Oxford: Pergamon Press.

Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences, 102*(23), 8386-8391.

Rhee, S. (2019). Lexicalization patterns in color naming in Korean. In I. Raffaelli, D. Katunar & B. Kerovec (eds.), *Lexicalization patterns in color naming: a cross-linguistic perspective*, 109-128. Amsterdam: John Benjamins.

Root, R. T., & Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American Journal of Psychology, 78*(2), 285-289.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of experimental psychology, 12*(3), 225-239.

Saussure, F. de 1959[1916]. *Course in General Linguistics*. New York: The Philosophical Library.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., Krauss, A., & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society of London B: Biological Sciences*. doi: 10.1098/rspb.2010.0769

Siegel, A., Silverman, I., & Markel, N. N. (1965). On the effects of mode of presentation on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior, 6*, 171-173.

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

Swadesh, M. (1971). *The origin and diversification of language*. Edited postmortem by Joel Sherzer. London: Transaction Publishers.

Tamariz, M., Roberts, S. G., Martínez, J. I., & Santiago, J. (2017). The interactive origin of iconicity. *Cognitive science, 42*(1), 334-349.

Taylor, A. M., & Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology 280*(3), 221-236.

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological science, 23*(12), 1443-1448.

Traunmüller, H. 1994. Sound symbolism in deictic words. In H. Auli & P. af Trampe (eds.), *Tongues and Texts Unlimited. Studies in Honour of Tore Jansson on the Occasion of his Sixtieth Anniversary*, 213-234. Department of Classical Languages, Stockholm University.

Tsuru, S., & Fries., H. S. (1933). Sound and Meaning. *Journal of General Psychology, 8*, 281-284.

Ultan, R. 1978. Size-sound symbolism. In J. Greenberg (ed.), *Universals of human language: phonology* (Vol. 2), 525-568. Stanford: Stanford University Press.

Viberg, Å. (2001). Verbs of perception. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), *Language typology and language universals: An international handbook*, 1294-1309. Berlin & New York: Walter de Gruyter.

Vinson, D., Thompson, R. L., Skinner, R., & Vigliocco, G. (2015). A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language, 82*, 56-85.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. Attention, *Perception, & Psychophysics, 74*(8), 1792-1809.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex, 42*(2), 264-280.

Watanabe, K., Greenberg, Y., & Sagisaka, Y. (2014). Sentiment analysis of color attributes derived from vowel sound impression for multimodal expression. In Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific (pp. 1-5).

Westbury, C., Hollis, G., Sidhu, D. M., & Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language, 99*, 122-150.

Wichmann, S., Holman, E. W., & Brown, C. H. (2010). Sound Symbolism in Basic Vocabulary. *Entropy, 12*(4), 844-858.

Wisseman, H. (1954) *Untersuchungen zur Onomatopoie,* Part I: Die sprachpsychologischen Versuche. Heidelberg: Winter.

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful" *JOSA A, 31*(4), A365-A374.

Woodworth, N. L. (1991). Sound symbolism in proximal and distal forms. *Linguistics, 29*, 273-299.

Study I

Niklas Erben Johansson*, Andrey Anikin, Gerd Carling
and Arthur Holmer

The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features

https://doi.org/10.1515/lingty-2020-2034
Revised August 27, 2019; accepted May 21, 2018

Abstract: Sound symbolism emerged as a prevalent component in the origin and development of language. However, as previous studies have either been lacking in scope or in phonetic granularity, the present study investigates the phonetic and semantic features involved from a bottom-up perspective. By analyzing the phonemes of 344 near-universal concepts in 245 language families, we establish 125 sound-meaning associations. The results also show that between 19 and 40 of the items of the Swadesh-100 list are sound symbolic, which calls into question the list's ability to determine genetic relationships. In addition, by combining co-occurring semantic and phonetic features between the sound symbolic concepts, 20 *macro-concepts* can be identified, e. g. basic descriptors, deictic distinctions and kinship attributes. Furthermore, all identified macro-concepts can be grounded in four types of sound symbolism: (a) unimodal imitation (*onomatopoeia*); (b) cross-modal imitation (*vocal gestures*); (c) diagrammatic mappings based on relation (*relative*); or (d) situational mappings (*circumstantial*). These findings show that sound symbolism is rooted in the human perception of the body and its interaction with the surrounding world,

Notes: Contributions. Niklas Erben Johansson: text writing, data collection, theoretical and methodological design, graphics design, evaluation, text revision. Andrey Anikin: methodological design, statistical analysis, evaluation, text revision. Gerd Carling: supervision, theoretical and methodological design, text revision. Arthur Holmer: supervision, theoretical and methodological design, text revision.

***Corresponding author: Niklas Erben Johansson [nɪklas æʁbɛn juʊhan:sɔn],** Division of General Linguistics, Center for Language and Literature, Lund University, Helgonabacken 12, SE-223 62, Lund, Sweden, E-mail: niklas.erben_johansson@ling.lu.se
Andrey Anikin [ɐndrʲˈej ɐnʲˈɪkʲɪn], Division of Cognitive Science, Department of Philosophy, Lund University, Helgonavägen 3, SE-221 00, Lund Sweden, E-mail: andrey.anikin@lucs.lu.se
Gerd Carling [jæʁd kʰɑ:ʁlɪŋ]: E-mail: gerd.carling@ling.lu.se, **Arthur Holmer [ɑ:ʁtøʁ hɔlməʁ]:** E-mail: arthur.holmer@ling.lu.se, Division of General Linguistics, Center for Language and Literature, Lund University, Helgonabacken 12, SE-223 62, Lund, Sweden

and could therefore have originated as a bootstrapping mechanism, which can help us understand the bio-cultural origins of human language, the mental lexicon and language diversity.

Keywords: Sound symbolism, iconicity, lexical typology, phonetic typology, language evolution, Swadesh list, origin of language, semantic typology

1 Pulling iconicity off the sidelines

This paper contributes to the increasingly popular research area of sound symbolism, by looking at 344 basic vocabulary concepts from 245 independent language families. The main purpose of the paper is to answer the following questions:

(a) What is the cross-linguistic extent of sound symbolism in basic vocabulary?
(b) Which types of sound symbolism can be distinguished?
(c) What does sound symbolism reveal about fundamental categories of human cognition?

Cross-linguistic sound symbolic patterns in basic vocabulary are particularly interesting since they entail cognitively universal associations, which were present early in our evolutionary history and must have impacted the formation of human language. Thus, defining the sound-meaning associations that belong to the core of sound symbolism, i. e. the most fundamental and language-independent associations and their accompanying semantic and phonetic features, is a way of looking into the most basic meanings in language and elucidating how lexical fields are related to each other and develop over time. In addition, mapping out correspondences between sound and meaning provides a valuable source of testable hypotheses for future perceptual studies, and this data can help us understand how humans classify concepts. The present paper achieves this by excluding genetic bias and including a wider range of investigated concepts compared to previous comparable studies. It also includes a sound feature system designed to facilitate analysis of lexical sound symbolism and demonstrates how sound-meaning associations can be arranged into semantically and phonetically superordinate concepts, referred to as *macro-concepts*.

Over the roughly twenty-year period of renewed interest in non-arbitrary associations between sound and meaning, referred to as *iconicity*, *non-arbitrariness*, *motivatedness*, and here, *(lexical) sound symbolism*, the area has gone from a poorly understood field residing on the fringes of linguistics and semiotics to an area extensively studied from a range of perspectives and through a wide array of methods (Perniss et al. 2010; Dingemanse et al. 2015). There have been several attempts to describe various sound-meaning associations and their causes, although the vast majority of studies have based their findings on only a few languages and concepts (Köhler 1929; Sapir 1929; Newman 1933; Fónagy 1963; Diffloth 1994; Sereno 1994; Ramachandran & Hubbard 2001, etc.).

There is also renewed interest in typological studies of *phonesthemes* – language-specific morpheme-like phoneme clusters that lack compositionality – and *ideophones* – words that evoke sensory perceptions (Hinton et al. 1994; Ibarretxe-Antuñano 2006; Iwasaki et al. 2007; Akita 2009, Akita 2012; Dingemanse 2012, Dingemanse 2017, Dingemanse 2018; Dingemanse & Akita 2016; Ibarretxe-Antuñano 2017).

Increasingly, studies have investigated the role that sound symbolism, and iconicity in general, play in language acquisition and language evolution (Kita et al. 2010; Fay et al. 2013; Perlman & Cain 2014; Perlman et al. 2015; Perniss & Vigliocco 2014; Lockwood et al. 2016a, Lockwood et al. 2016b). Other research has focused on how specific sound symbolic domains operate (Nielsen & Rendall 2013; Cuskley et al. 2015), or on more general, underlying, more or less universal causes and structural features of sound symbolism. Among these, the most famous example is probably Ohala's (1994) physiologically and functionally grounded *frequency code*, which states that the fundamental frequency depends on body size and thereby maps size onto pitch.

More recent comparative research has shown that the correlation between body size and fundamental frequency is actually rather weak and mostly found in species with highly variable body sizes, such as domestic dogs, whereas formant dispersion is a more reliable predictor of size (Taylor & Reby 2010). Nevertheless, listeners erroneously associate lower pitch of human voices with size (Bruckert et al. 2006; Collins 2000) and physical strength (Sell et al. 2010). These correlations are further utilized in various ways to evoke properties related to size: for example, if an animal wants to seem threatening, it can erect feathers or growl with low pitch to exaggerate its apparent size. Reversely, cowering and whining with high pitch suggests smaller size and thereby indicates submissiveness. Thus, most animals perceive a low and/or falling F0 to indicate large size, authority, dominance, large distance, etc., and a high and/or rising F0 to indicate small size, politeness, dependence, proximity, etc.

Despite the progress made in the field of sound symbolism and iconicity, which has greatly contributed to the reevaluation of the Saussurean principle of arbitrariness of the linguistic sign (Saussure 1983[1916]), our understanding of sound symbolism and its mechanisms remains patchy. One way of bridging the gaps in our knowledge of universal sound symbolism is to conduct large-scale cross-linguistic comparisons of basic vocabulary (Swadesh 1971; Goddard & Wierzbicka 2002) to establish sound symbolic realizations.

In addition to establishing sound symbolic associations, such inquiries can contribute more extensive examinations of interdependent semantic and phonetic correlations and patterns that can help to explain which properties of human (spoken) language are affected by sound symbolism, and possibly why. A few studies of this type have been conducted, including up to thousands of languages and a greater number of concepts. By examining 37 languages, Traunmüller (1994) found that words for first person singular personal pronoun tend to contain nasals, while its second person counterpart tends to contain stops. The same study, along with Ultan (1978) and Woodworth (1991), which included 136 and 26 languages, respectively, found that deictic proximal words such as 'this' often contain high, front, unrounded vowels, while words meaning 'that' contain low, back, rounded vowels.

Johansson (2017) found several semantic groupings and relations based on phonological contrasts, e. g. SMALLNESS, LARGENESS, deictic distinctions, MOTHER-FATHER, and several oppositional perceptual concepts relating to shape, such as WARMTH, LIGHT and CONSISTENCY, when 56 fundamental oppositional concepts were compared over 75 genetically and areally distributed languages.

Wichmann et al. (2010) compared a 40-item subset of the Swadesh list (Swadesh 1971), normally used for establishing genealogical relationships between languages, in around 3,000 languages. They found seven phonologically distinctive words, including associations between BREAST and labial sounds (reflecting the suckling of a child), between phonemes associated with hard and round qualities and KNEE, and between nasal sounds and NOSE. Wichmann et al. (2010) also reported symbolically coded deictic distinctions between I, YOU, WE and NAME.

Blasi et al. (2016) ambitiously expanded on the Wichmann et al. study by investigating the same lexical items in over 6,000 languages and dialects. By sifting out all possible combinations of their investigated meanings and sounds that occurred in at least ten language families and three out of six geographical macro-areas, they were able to define 74 positive and negative sound-meaning associations, over 30 concepts and 23 sound groups, making it the most extensive study on typological sound symbolism so far. These results show the potential extent of sound symbolism in some of our most basic lexemes, but they also illustrate that there seems to be a link between sound symbolism and the origin of human language. Pure imitation, or onomatopoeia, can refer to a range of referents that produce sounds, but most sound-meaning associations found in basic vocabulary involve concepts which are difficult to mimic acoustically, such as deictic concepts. These concepts must therefore be grounded in some other way, which probably requires more effort than unimodal imitation. This suggests that if a basic vocabulary concept is sound symbolic, despite the extra effort necessary to establish the mapping, it likely plays an important role in language as well. For that reason, sound symbolism seems to be one way of establishing fundamental lexical fields.

Concurrently, categorization of distinct types of sound symbolic mappings has increasingly been brought to the fore. Already a hundred years ago, Jespersen (1922) constructed seven rudimentary yet broad categories of sound symbolism, which included direct imitation, originator of the sound, movement (inseparable from sound), things and appearances, states of mind, size and distance, and length and strength of words and sounds. The categories of this taxonomy are, however, only based on semantics and are neither mutually exclusive nor exhaustive, as pointed out by Abelin (1999). In a more recent context, Dingemanse (2011) built on the work of Pierce (1931–1958) and Bühler (1934), as well as on his extensive work on ideophones, to describe two primary types of sound symbolic mappings (Table 1), or *iconicity* (i. e. non-arbitrary rather than iconic in a strict semiotic sense).

Table 1: Types of sound symbolism described by Dingemanse (2011) and Carling and Johansson (2014) with respective involved modalities, semiotic grounds, emergence and examples.

Type of mapping		Modality	Semiotic ground (Emergence)	Dingemanse (2011)	Carling and Johansson (2014)	Example
Imitative/Absolute		Uni-modal	Iconic (Direct)	Imagic	One-to-one	Onomatopoeia
Diagrammatic	Word-internal	Cross-modal	Iconic/ indexical (Structural)	Gestalt	–	Reduplication
	Word-relational	Cross-modal	Indexical (Structural)	Relative	Oppositional/ Relational	Frequency code
Associative	Language-internal	Cross-modal	Indexical (Analogical)	–	Complex	Phonesthemes

The first and semiotically simplest form is *imagic iconicity* (referred to as *absolute iconicity* by Dingemanse et al. 2015 and *imitative sound symbolism* by; Hinton et al. 1994), which involves pure iconic imitation of real world sounds, or onomatopoeia. Since humans are bound by their articulatory filters, this type of imitation is generally far from perfect and ranges from recognizable to approximate. The second type, *diagrammatic iconicity*, associates relations between forms with relations between meanings, which allows all types of sensory attributes of speech, such as tone and volume, to establish sound-meaning associations, and can be further divided into two subtypes. *Gestalt iconicity* includes resemblance between word structure and structure of the perceived event which evokes iterated or intense events. The most telling example of this is reduplication, as shown in Japanese *doki-doki* 'heartbeat, excitement'. *Relative iconicity*, on the other hand, involves relations between multiple sounds or sound combinations and multiple meanings. This is perfectly exemplified by Ohala's (1994) frequency code, which conjoins the two respective oppositional poles of the phonetic parameter FREQUENCY and the semantic parameter SIZE by correlating high-frequency sounds with small size and low-frequency sounds with large size.

In the same spirit, Carling and Johansson (2014) tried to establish a similar taxonomy based on a range of semiotic and sound symbolic parameters. Firstly, the Peircian sign distinction was used to disentangle *iconic signs* (resemblance based on likeness, such as representing a human through a stick figure), *indexical signs* (resemblance based on contiguity in time and space, such as representing fire through smoke) and *symbolic signs* (convention) (Ahlner & Zlatev 2010; Johansson & Zlatev 2013). Secondly, realizations of sound symbolic mappings on the form side were divided into four types. (a) a motivated connection between meaning and qualitative aspects of linguistic form (*qualitative iconicity*), such as phonematic or phonotactic structure as in *mil-mal* 'small-big'); (b) a motivated connection between meaning and quantitative aspects of linguistic form (*quantitative iconicity*), such as word length or reduplication, as in the difference in perceived descriptive length between *long* and *looooooong;* (c) a motivated connection between meaning and parts of lexeme(s) (*partial iconicity*), as in the *gl-* section of the phonesthemes *glisten*, *glitter*, *glimmer* etc.; and (d) a motivated connection between meaning and whole lexeme(s) (*full-word iconicity*), as in the bird name *cuckoo*.

Lastly, organization and type of emergence of mappings were divided into (a) a motivated connection based on one-to-one correlations between forms and meaning, which is grounded in an obvious association with an acoustic signal, i. e. *one-to-one iconicity* of *direct emergence;* (b) two or more meanings in oppositional or relational semantic positions with corresponding linguistic forms which are grounded in a preconditioned structure and not directly related to other linguistic material within the language, i. e. *oppositional/relational iconicity* of *structural emergence;* (c) complex networks of meaning(s) and linguistic form(s) which are grounded in an association to other sound symbolic words within the language, i. e. *complex iconicity* of *analogical emergence* (see also Hinton et al.'s [1994] *conventional sound symbolism*).

The structure of the paper is as follows: Section 2 is a general overview of the aims of the paper. Section 3 presents the methodology used for this paper and includes descriptions of how the featured concepts and languages were sampled, how the data was collected and transcribed and how the phonetic categorization and data analysis were conducted. Section 4 includes general results along with plausible explanations for the sound-meaning associations found. Section 5 features discussion about the role iconicity could have played in the evolution and development of language. Section 6 includes some final remarks.

2 Amending unresolved issues by adapting them to sound symbolism

Based on previous findings, it is evident that sound symbolism is a rather common phenomenon, but its true extent in the linguistic system is still not completely known. Likewise, it remains unclear which sounds are involved and how they interact with different concepts. We believe that research on sound symbolism can benefit from three methodological advances: expanding the number of analyzed lexemes, improving transcription systems, and sampling unrelated language families to avoid genealogical bias.

Firstly, for every study where the scope of investigated meanings and sounds is increased, more sound symbolic mappings are discovered. This suggests that analyzing a larger number of lexemes should significantly improve our ability to formulate and define different types of sound symbolism. Therefore, we investigated a much larger number of basic vocabulary items than previous studies. This enables a deeper understanding of the semantic and phonetic relationships that sound symbolic mappings adhere to from a functional, communicative or embodied perspective, where embodiment refers to the shaping of the human mind by the human body (Clark 2006; Zlatev 2007; Ziemke 2016; Johansson 2017). In addition, it also provides a proper assessment of the origin of sound symbolic mappings, e. g. imitation.

Secondly, the rather coarse transcription system used in Wichmann et al.'s and Blasi et al.'s large studies fails to capture several distinctions essential for sound symbolic associations (Ohala 1994), e. g. contrasts between places of articulation, some contrasts between manners of articulation (stops and fricatives) and, most crucially, voicing distinctions between several sounds.

Consequently, in the present study we first transcribed sounds according to a close approximation of the International Phonetic Alphabet. We then grouped the sounds into a more principled classification of sound groups based on systematic divisions of salient phonetic parameters which have been shown to be relevant for sound symbolism. Lastly, investigating typological patterns in the vast majority of the world's languages introduces the problem of genealogical bias, although sound symbolism and cognacy do not necessarily rule each other out. Previous studies have attempted to solve this by using Levenshtein distances as a proxy for cognacy (Blasi et al. 2016), but this method has poor genealogical predictability (Greenhill 2011) and can be influenced by borrowings, sound change, or even sound symbolism (!) (Campbell & Poser 2008). Thus, genealogical bias has been completely eliminated from the present study by means of including only one language per language family and spreading the chosen languages geographically to exclude areal bias.

With these issues and solutions in mind, the present study focuses on the phonetic and semantic features involved in sound symbolism, narrowing down the definition of a sound symbolic *association* (referred to as *signal* by Blasi et al.) to (near-)universal, non-arbitrary and flexible associations between sounds and meanings that are statistically detectable across languages when genetic and areal biases are excluded. This approach may shed new light on the core of sound symbolism by contributing to our understanding of the cross-linguistic extent of sound symbolism in basic vocabulary, which types of sound symbolism can be distinguished and what sound symbolism can reveal about fundamental categories of human cognition.

3 Method

3.1 Establishing near-universal vocabulary

When searching for sound symbolic patterns, basic vocabulary is especially suitable since it consists of concepts that are supposed to be salient for all speakers regardless of the language, culture and era. These concepts broadly relate to the fundamental categories of the mind (e. g. emotions, senses, tastes, perceivable physical properties), the body (e. g. body part terms, mental and bodily functions), society (e. g. kinship terms, human categories), the surrounding world (e. g. natural entities) and reference (e. g. deictic concepts, determiners, spatial relations). Furthermore, to account for language-specific delimitation of semantic fields, boundaries between concepts were generalized according to prevailing typological and physiological patterns. For example, singular-plural distinctions were included for pronouns but not dual, paucal etc., and individual terms for ‘hand’ and ‘arm’ were included rather than a term for the entire limb. Thus, the selection of the 344 featured concepts (see Table 2) was based on:

(a) near-universality, i. e. presence in the majority of the world’s languages,
(b) strong linguistic typological patterns as a basis for drawing the borders between concepts,
(c) physiological and natural constraints as a basis for drawing the borders between concepts,
(d) lists of basic vocabulary for high comparability with similar studies.

To begin with, we included all of the 56 fundamental oppositional concepts that were shown to have great sound symbolic potential by Johansson (2017), mostly based on proposed lexical universals (Dixon 1982; Goddard 2001; Goddard & Wierzbicka 2002; Koptjevskaja-Tamm 2008; Paradis et al. 2009), namely I-YOU, BIG-SMALL, GOOD-BAD, THIS-THAT, MANY-FEW, BEFORE-AFTER, ABOVE-BELOW, FAR-NEAR, MAN-WOMAN, BLACK-WHITE, HOT-COLD, HERE-THERE, LONG-SHORT, NIGHT-DAY, FULL-EMPTY, NEW-OLD, ROUND-FLAT, DRY-WET, WIDE-NARROW, THICK-THIN, SMOOTH-ROUGH, HEAVY-LIGHT, DARK-LIGHT, QUICK-SLOW, HARD-SOFT, DEEP-SHALLOW, HIGH-LOW and MOTHER-FATHER.

Associations between colors and sounds are perhaps among the most commonly studied sound symbolic areas. Even though synesthetes experience these associations more strongly than non-synesthetes (Ward et al. 2006), hue, chroma and lightness are also associated with auditory frequency and loudness (Spence 2011; Walker 2012; Hamilton-Fletcher et al. 2017) and other senses, such as touch (Ludwig & Simner 2013), in the general population. What is more, Berlin and Kay (1969) famously demonstrated strong cross-linguistic regularities in the lexicalization patterns of monolexemic color terms. However, monolexemic terms are not guaranteed to be free of lexical interference from non-color concepts, as they can often be traced back to old derivations of a referent in nature having that particular color. For example, ‘green’ is ultimately derived from Proto-Indo-European _*ǵ^{h}reh$_{1}$-ni-_, meaning ‘to grow’ (Kroonen 2010–), i. e. ‘plant-colored’. Selecting color concepts based on these lexicalization patterns may therefore not be ideal for the purposes of this paper. Therefore, selecting these concepts based on color opponency (Kay & Maffi 2013) was judged a more suitable choice since it offers a more neutral division of the color spectrum which is also cross-linguistically grounded. Thus, we included two pairs of fundamental opponent chromatic colors (RED-GREEN and YELLOW-BLUE), a single pair of fundamental achromatic colors (BLACK-WHITE), and GRAY, the combination of the most basic colors. Number concepts were also narrowed down in a similar manner to accommodate most of the world’s numeral systems (Comrie

Table 2: The 344 investigated concepts. Kinship concepts are abbreviated according to the standard kinship terminology: M = mother, F = father, Z = sister, B = brother, D = daughter, S = son, O = older, Y = younger, _MS = male speaking, _FS = female speaking.

ABOVE	COME	GIRL	LONG	RIPE	SUCK	WHERE?	MoB _MS
AFRAID	CORRECT	GIVE	LOUD	RIVER	SUN	WHITE	MyB _FS
AFTER	COUGH	GO	LOUSE	ROOT	SWEAT	WHO?	MyB _MS
AIR	CROOKED	GOOD	LOW	ROPE	SWEET	WHY?	FoB _FS
ALL	CRUSH	GRASS	LUNG	ROTTEN	SWIM	WIDE	FoB _MS
ANGRY	CRY	GRAY	MAN	ROUGH	TAIL	WING	FyB _FS
ANIMAL	DARK	GREASE	MANY	ROUND	TAKE	WIND	FyB _MS
ANT	DAY	GREEN	MAYBE	RUN	TASTE	WOMAN	oZ _FS
ARM (UPPER)	DEEP	GROW	MILK	SAD	TEN	WORD	oZ _MS
ARM (LOWER)	DEFECATE	HAIR	MOON	SALTY	TESTICLE	WRONG	yZ _FS
ASHES	DIE	HALF	MOUNTAIN	SAME	THAT	YAWN	yZ _MS
BACK	DIRTY	HAND	MOUTH	SAND	THAT YONDER	YEAR	oB _FS
BAD	DO	HAPPY	NAME	SAY	THEN	YELLOW	oB _MS
BARK	DOG	HARD	NARROW	SEA	THERE	YESTERDAY	yB _FS
BEAUTIFUL	DRINK	INTERCOURSE	NAVEL	SEE	THERE YONDER	YOUNG	yB _MS
BECAUSE	DRY	HEAD	NEAR	SEED	THICK	1SG	oZD _FS
BEFORE	DUST	HEAR	NECK	SEMEN	THIGH	1PLI	oZD _MS
BEHIND	EAR	HEART	NEW	SEVEN	THIN	1PLE	yZD _FS
BELLY	EARTH	HEAVY	NIGHT	SHADOW	THINK	2SG	yZD _MS
BELOW	EAT	HERE	NINE	SHALLOW	THIS	2PL	oBD _FS
BESIDE	EGG	HIDE	NIPPLE	SHARP	THREE	3SG	oBD _MS
BETWEEN	EIGHT	HIGH	NO(THING)	SHORT	THROAT	3PL	yBD _FS
BIG	ELEVEN	HIT	NOSE	SIT	THUNDER	MM _FS	yBD _MS
BIRD	EMPTY	HORN	NOT	SIX	TIE	MM _MS	oZS _FS
BITE	EYE	HOT	NOW	SKIN	TOE	FM _FS	oZS _MS
BITTER	FALL	HOUSE	OLD (AN)	SKY	TONGUE	FM _MS	yZS _FS

BLACK	FAR	HOW?	OLD (INAN)	SLEEP	TOOTH	MF _FS	yZS _MS
BLOOD	FART	IF	OLD MAN	SLOW	TOUCH	MF _MS	oBS _FS
BLOW	FEATHER	IN FRONT OF	OLD WOMAN	SMALL	TREE	FF _FS	oBS _MS
BLUE	FEW	IN(SIDE)	ONE	SMELL	TURN	FF _MS	yBS _FS
BLUNT	FINGER	KILL	OTHER	SMOKE	TWELVE	M _FS	yBS _MS
BODY	FINGERNAIL	KNEE	OUT(SIDE)	SMOOTH	TWENTY	M _MS	D _FS
BONE	FIRE	KNOW	PART	SNEEZE	TWO	F _FS	D _MS
BOY	FISH	LAUGH	PATH	SNORE	UGLY	F _MS	S _FS
BRAIN	FIVE	LEAF	PENIS	SOFT	URINATE	MoZ _FS	S _MS
BREAST	FLAT	LEFT	PERSON	SOME	VOMIT	MoZ _MS	DD _FS
BREATHE	FLESH	LOWER LEG	POINTY	SOUR	VULVA	MyZ _FS	DD _MS
BURN	FLOWER	LIE (DOWN)	QUICK	SPIT	WANT	MyZ _MS	SD _FS
BUTTOCKS	FLY (N)	LIGHT (NOT DARK)	QUIET	STAND	WATER	FoZ _FS	SD _MS
CARRY	FLY (V)	LIGHT (WGT)	RAIN	STAR	WEAK	FoZ _MS	DS _FS
CLEAN	FOOT	LIP	RAW	STONE	WET	FyZ _FS	DS _MS
CLOUD	FOUR	LIVE	RED	STRAIGHT	WHAT?	FyZ _MS	SS _FS
COLD	FULL	LIVER	RIGHT	STRONG	WHEN?	MoB _FS	SS _MS

2013), namely *decimal* (base 10), *vigesimal* (base 20), *restricted* (individual terms up to around 'five' which are combined to create higher numbers) and *extended body-part* (based on individual words for body part without an arithmetic base). The final selection included numbers ONE through TWELVE, as well as TWENTY.

Among the deictic concepts, first, second and third personal pronouns in the singular and plural, as well as inclusive and exclusive first-person plural (1SG, 2SG, 3SG, 1PLI, 1PLE, 2PL, 3PL). These general concepts were included to account for the various strategies that languages use to divide pronouns and nouns into noun classes and grammatical genders (Corbett 2013) as the alternative would force us to include separate slots for all possible categories (such as masculine, feminine, neuter, common, animate, inanimate, human, non-human, countable, uncountable etc.) for each pronoun concept.

Proximal, medial and distal location adverbs (HERE, THERE, THERE YONDER) and demonstratives in the singular (THIS, THAT, THAT YONDER) were chosen for being the most common types cross-linguistically (Diessel 2014), with the addition of two temporal deictic concepts (NOW, THEN) and six interrogative pronouns, which incorporate the notions of human (WHO?), non-human (WHAT?), location (WHERE?), time (WHEN?), manner (HOW?), and reason (WHY?).

Closely related to demonstratives, location concepts seem to be arranged to be maximally informative within languages, i. e. languages seem to categorize objects in a way that favors accurate mental reconstruction by a listener of a speaker's intended meaning rather than basing it on other natural or salient categories (Khetarpal et al. 2013). Despite this, there does not seem to be a clear-cut set of universal categories (Levinson & Meira 2003; Burenhult & Levinson 2008; Khetarpal et al. 2010). Thus, the selected concepts were only meant to convey immediate relational positions to objects rather than directions (e. g. ABOVE but not UP), and belonged to four types: horizontal (LEFT-RIGHT, BEHIND-IN FRONT OF, BESIDE), vertical (ABOVE-BELOW), time (BEFORE-AFTER), and object-related (INSIDE-OUTSIDE, BETWEEN). In addition, a universal (ALL), existential (SOME) and negatory (NOTHING) quantifier were included, as well as an equal (SAME) and contrastive (OTHER) determiner.

Linguistic variation in age categories creates similar issues when working with cross-linguistic data, e. g. the Austroasiatic language Khmu [kgj] distinguishes about twice as many categories as English [eng] (children, teenagers, young adults, adults and elders); hence, the selected concepts only included a general term (PERSON) and three age-coded groups of concepts in order to fit most languages: elderly (OLD MAN, OLD WOMAN), adult (MAN, WOMAN) and child (BOY, GIRL).

Surprisingly, one of the most studied anthropological subjects, kinship systems, which are organized in complex and varying manners, seem to exhibit an almost optimal tradeoff between simplicity and informativeness (Kemp & Regier 2012). However, in contrast to the location concepts, kinship terms have multiple vectors that could be sound symbolically encoded. Hence, the main criterion used to select these concepts was to capture as many kinship terms as possible and included all blood relations two steps from the ego, with relative age distinctions when applicable, e. g. *younger sister's son*, while excluding more distant relations, non-blood relations and umbrella terms, e. g. *grandparent* and *sibling*. For a complete list of the 64 selected kinship concepts see Table 2.

Body part concepts are perhaps some of the most fundamental linguistic concepts, but the linguistic segmentation of the body is highly language-specific. While it is easy to assume that body part nomenclature is primarily determined by visual features, there is evidence that proprioceptive (Enfield et al. 2006), developmental (Andersen 1978), and neurological (Penfield & Boldrey 1937; Penfield & Rasmussen 1950) factors also make important contributions. Furthermore, it has been proposed that most languages adhere to a possibly universal hierarchy of lexicalized body parts (Andersen 1978), for the most part corroborated by the fact that joints act as boundaries between body parts in distance judgements (Enfield et al. 2006; de Vignemont et al. 2009). Thus, body part concepts considered fundamental according to these criteria were included (ARM, BACK, BODY, BREAST, CHEST/TRUNK, EAR, EYE, FACE, FINGER, FOOT, HAND, HEAD, LEG, MOUTH, NECK, NOSE, TOE). However, CHEST/TRUNK was replaced by the more distinctive BELLY, FACE was excluded in favor of HEAD, and ARM and LEG were further divided into UPPER ARM, LOWER ARM, THIGH and LOWER LEG. In addition, body part concepts with distinctive appearances and/or many nerve endings were included (BUTTOCKS, FINGERNAIL, HAIR, NAVEL, TOOTH and SKIN, LIP, THROAT, TONGUE, PENIS, VULVA, NIPPLES, TESTICLES), as well as the most distinctive internal organs: HEART, LUNGS and BRAIN. We also included all salient bodily and mental functions related to eyes (CRY), mouth (BITE, BLOW, BREATHE, COUGH, DRINK, EAT, LAUGH, SAY, SNORE, SPIT, SUCK, VOMIT, YAWN), nose (SNEEZE), genitals/excrement (DEFECATE, INTERCOURSE, SEMEN, URINATE), skin (BLOOD, MILK, SWEAT), mind (KNOW, SLEEP, THINK), movement (FALL, GO, LIE, RUN, SIT, STAND, TURN) and living (DIE, LIVE). Generally, the verb related to the function was chosen, except for BLOOD, MILK, SEMEN and SWEAT. HICCUP, BURP and MENSTRUATE were excluded due their similarities to COUGH and BLOOD, respectively.

According to several studies (Viberg 1983, Viberg 2001), sensory concepts are hierarchically lexicalized, sight being the most fundamental, possibly because there are more occasions to talk about visual objects than about objects related to other senses. Furthermore, touch, taste and smell words often lexically overlap with other senses (San Roque et al. 2015), and not much work has been done on the sound symbolic aspects of these concepts. Accordingly, all typical sense words were selected (SEE, HEAR, TASTE, TOUCH, SMELL). Similarly, despite criticism of the four traditional basic taste distinctions (Erickson 2008) and various lexical conflations of taste terms occurring throughout languages, cross-linguistic data does support BITTER, SALTY, SOUR and SWEET as fundamental taste concepts (Majid & Levinson 2008), which were therefore included in the list. Basic emotion concepts were, on the other hand, somewhat generalized following Jack et al. (2014), resulting in HAPPY, SAD, AFRAID and ANGRY.

We also selected a number of natural entities that would be salient features in the surrounding world of pre-agrarian societies (BONE, FIRE and SAND), general plant and animal concepts (DOG), as well as concepts relating to weather, heaven (e. g. SKY and SUN), (bodies of) water, DAY and NIGHT, but not ICE and SNOW, as they are unknown in many parts of the world. In addition, in order to make the sample of concepts comparable to many other studies which incorporate basic vocabulary and to estimate base frequencies of each sound, it is crucial to include a substantial number of concepts which are likely not affected by sound symbolism. We therefore also added the remaining concepts present in the Swadesh-100 and Swadesh-207 lists (Swadesh 1971), the Leipzig-Jakarta list (Haspelmath & Tadmor 2009) and Goddard and Wierzbicka’s (2002) *semantic primes*. Altogether, this included AIR, ANIMAL, ANT, ASHES, BARK, BECAUSE, BIRD, BLUNT, BONE, BURN, CARRY, CLEAN, CLOUD, COME, CORRECT, CROOKED, CRUSH, DIRTY, DO, DOG, DUST, EARTH, EGG, FART, FEATHER, FIRE, FISH, FLESH, FLOWER, FLY (n), FLY (v), GIVE, GRASS, GREASE, GROW, HALF, HIDE, HIT, HORN, HOUSE, IF, KILL, KNEE, LEAF, LIVER, LOUD, LOUSE, MAYBE, MOON, MOUNTAIN, NAME, NOT, PART, PATH, PERSON, POINTY, QUIET, RAIN, RAW, RIPE, RIVER, ROOT, ROPE, ROTTEN, SAND, SEA, SEED, SHADOW, SHARP, SKY, SMOKE, STAR, STONE, STRAIGHT, STRONG, SUN, SWIM, TAIL, TAKE, THUNDER, TIE, TREE, WANT, WATER, WEAK, WIND, WING, WORD, WRONG, YEAR and YESTERDAY.

3.2 Capturing linguistic diversity without genetic bias

Undoubtedly, simulating the diversity of human language as a whole by selecting a number of spoken languages is a complicated matter. Not only are languages incredibly diverse in terms of phonology, morphology, lexicon, semantics and syntax, but they also differ widely in the number of speakers, geographical spread and in the number of genetic relatives. However, the main point of selecting a sample of languages is to represent diversity and, for this particular research question, lexical diversity. As relations between languages are largely determined based on lexical differences, a more cautious approach for grouping languages into families is preferable. Therefore, Glottolog's (Hammarström et al. 2017) approach was adopted in favor of the other currently largest language database, Ethnologue (Simons & Fennig 2017), except in the few cases when Ethnologue's language division was more conservative. Even if complete datasets for all the world's documented languages were available, we would still not get the complete picture of what human language is capable of, as most languages are already dead.

The aim of sampling is therefore to include one representative from all the world's living and extinct documented language families (and isolates) with sufficient and reliable data for at least one member, spread geographically as widely as data availability allowed. In addition, this also compensates for the concepts that lacked data for some languages, since the language sample remains genetically balanced regardless of the number of included languages. Thus, after excluding artificial, sign, unattested and unclassifiable languages, as well as creoles, mixed languages, pidgins and speech registers, because they are mostly based on already existing languages, 245 languages and language families were selected (58.5% of the 419 featured on Glottolog), of which 68 were isolates (Figure 1 and Online Appendix 2). This sample of languages yielded almost 70,000 lexemes.

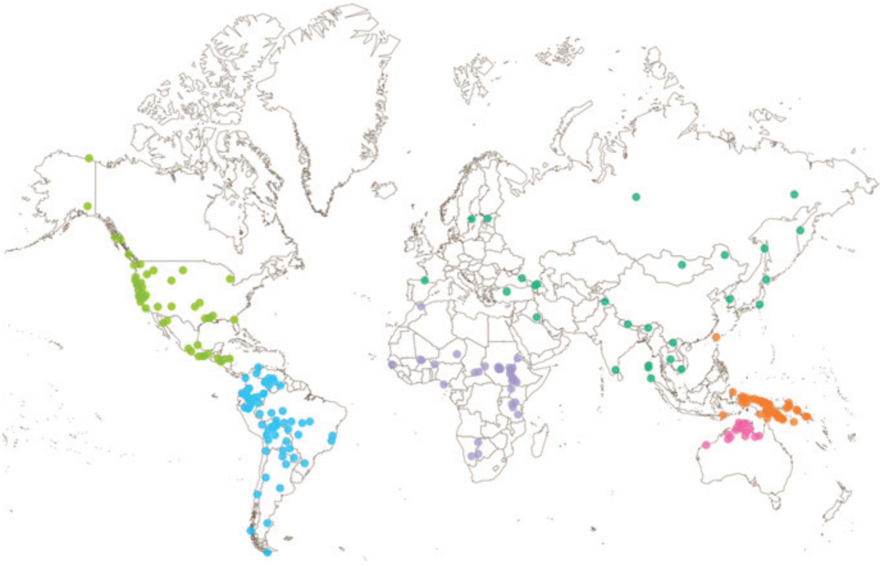


Figure 1: Featured languages divided into the same geographical macro-areas as used by Blasi et al. (2016) and shown by color (olive green: North America, blue: South America, forest green: Eurasia, purple: Africa, orange: Papunesia, pink: Australia).

3.3 Data collection

One of the challenges of compiling cross-linguistic data is data collection. For languages with many speakers or long histories as literary languages, comprehensive and reliable sources such as databases or comprehensive dictionaries make the collection of data straightforward. For many poorly documented languages, on the other hand, only a handful of sources have ever been produced, and usually only one or two of those are available. Data availability was thus an important consideration guiding language sampling. Furthermore, due to the varying quality of data, some concepts were not retrieved from all languages, but since only one language per language family was included, the sample remained unbiased.

Moreover, even when obstacles related to data availability have been overcome, differences between languages, such as grammatical marking, still pose problems. For example, when concepts were found to have multiple forms (e. g. gender inflections), only the unmarked form was selected to ensure comparability across languages, as long as relevant information about the meaning was provided through the lexical entries or grammatical descriptions, i. e. in the singular nominative for accusative systems, in the singular absolutive for ergative systems, and so forth. In many languages, the same concept can have a number of different roots or versions, e. g. in classificatory verbs in native North American languages (Kibrik 2012), which makes it difficult to know which form of a group of words is the unmarked one. Likewise, throughout languages, most concepts also have several synonyms. Therefore, all phonemes from all forms in these cases were combined into a single string rather than selecting only one of the forms to represent the concept in question. For example, the three English forms of the third person singular personal pronoun (*he*, *she* and *it*) were analyzed as a single word with six phonemes [hiːʃiːɪt]. Conversely, when the same term is used for more than one concept, both slots were filled with the same form. For example, in Pirahã [myp] ‘I’ and ‘we’ are both referred to as *ti*^{3}. In addition, large bodies of water are of great import for all speech communities, and thus the concept SEA naturally belongs in the list of featured concepts. However, since many cultures lack contact with oceans and thus have no specific word referring to SEA, in these cases LAKE was added instead. This was the only replacement of this kind.

Although including borrowed linguistic forms in cross-linguistic comparisons might seem counterintuitive, it does not by default result in an areal bias. A description of a language is only a snapshot of an ever-changing dynamic system, which means that if a word is borrowed and used, it is also part of the language. And in time, the borrowed words usually adapt to the semantic and phonological framework of the language. However, detected late loans from languages with a strong influence on other cultures, namely Arabic [ara], English, French [fra], Malay [msa], Mandarin Chinese [cmn], Portuguese [por] and Spanish [spa], were removed since the same loans from these languages often occur in a great number of languages and could therefore be mistaken for overrepresentations of sounds. Among these loans, we find e. g. PENIS, MILK, SALT, numerals, several color concepts, ANIMAL, BODY, IF and YEAR. All featured languages with large Sino-Xenic vocabularies have different lexical registers, and thus native words for a concept were selected when available. In the cases without native forms, many of the loans were kept as the vast majority were borrowed more than eight centuries ago and have undergone extensive phonological and often semantic change, unless the linguistic form showed considerable similarity across the Sino-Xenic languages. Likewise, loan words from less culturally influential languages that were only found in one target language were kept when no native form was found, especially if the word was borrowed within a language family.

3.4 Data transcription model

The inconsistency, quality and granularity of the sources also cause ripple effects for transcription of the collected data. Furthermore, the poor quality of many orthographies, especially of less studied languages, combined with the fact that different sources describing the same language often use different kinds of orthographies and are frequently based on the mother tongue of the data compiler, adds to the overall disarray. In other words, it is nearly impossible to make a dataset with a larger number of languages completely comparable without employing a unifying transcription system. Therefore, all sounds were transcribed into The International Phonetic Alphabet as accurately as the sources for the featured languages allowed for, albeit with some minor yet crucial differences. After the lexemes had been collected, we obtained phonological and orthographical descriptions from the same sources when available in order to convert the text into IPA. When this information was not available, which was the case for several older sources, we consulted available grammatical and phonological sketches and articles. We also utilized phonological data from databases and compilations of phoneme inventories, such as *PHOIBLE* (Moran et al. 2014), that described the languages in question. As for the featured extinct languages, most of them went extinct in recent times and are therefore quite well-described. However, the few included ancient languages, such as Sumerian [sux], obviously entail more phonetic uncertainty despite the amount of research that has gone into describing them.

The main aim of the current paper requires a quantification and statistical measuring of sound symbolic associations from a cross-linguistic perspective. This aim demands a model of data transcription that is capable of 1) capturing the diversity of various phonemic systems, 2) quantifying these diverse systems in a manner that is representative, comparable, and relevant to the research theme of the paper, *sound symbolism*. While IPA provides a detailed description of speech sounds, it can be too fine-grained for comparing such a diverse range of languages with highly dissimilar sound systems. Therefore, some sounds needed to be grouped together or segmented in order to make them statistically analyzable. In addition, these classifications should also correspond to how features of speech sounds are observed to behave with respect to sound symbolic mappings in languages. To begin with, all original IPA oral and nasal vowels were included, as well as all pulmonic, doubly articulated consonants and consonants with secondary articulation and non-pulmonic consonants. Voicing was also distinguished by contrasting complete voicelessness with all degrees of voicing, i. e. also including partial, weak and short voicing (Cho & Ladefoged 1999), in accordance with how voicing is mapped sound symbolically (Lockwood et al. 2016a).

Sounds that incorporate more than one place of articulation were split into two segments in order to quantify them separately. This for several reasons: a labialized velar stop, [k^w], might be used sound symbolically to indicate abruptness through the stop, or to indicate a round shape through the rounding of the lips. Thus, it cannot be equated only with [k] or [w] since the other feature would have been left unnoticed in the data. This model, which is based on how sound symbolism is observed to be reflected in language, may vary with respect to how precisely the phonemic systems of languages are rendered (some languages may have richer systems). However, the model captures the crucial acoustic features important for sound symbolism and allows these features (which are partly phonemes, partly acoustic representations) to be grouped in a more appropriate way than dedicated labels for combinations of phonemes. Hence, diphthongs and triphthongs were transcribed as sequences of vowels, and affricates as combinations of plosives and fricatives because of the shared closure phase between affricates and plosives, and the shared friction phase of affricates and fricatives (Sidhu & Pexman 2018).

Furthermore, the meanings that are sound symbolically associated with affricates are usually semantically similar to both meanings associated with stops and meanings associated with fricatives (Abelin 1999: 37–41). The same principle was applied to ejective affricates and consonants with double and secondary articulation, such as consonantal release types, as well as consonants with aspiration (including preaspiration), labialization and palatalization. For example, [tsʼ], [k͡p], [p^m], [k^h], [k^w] and [k^j] were transcribed as /tʼsʼ/, /kp/, /pm/, /kh/, /kw/ and /kj/, respectively. In contrast, breathy (murmured) vowels and nasalized and creaky voiced sounds were coded as separate phonemes since the involved features are difficult to distinguish. Plain click consonants, such as [ʘ], were considered voiceless and contrasted with voiced variants such as [^{g}ʘ]. While aspirated and glottalized click consonants were segmented as described above, nasalized clicks and voiceless nasalized clicks were considered separate phonemes, and clicks with a velar, velar ejective, uvular and uvular ejective fricative release were transcribed as a click followed by /x/, /xʼ/, /q/ and /qʼ/. Stress and tones were not recorded since information about stress patterns was generally lacking or poorly described for most languages, and tones occurred only in a fraction of the language sample, which would lead to very low comparability.

Phonetic length was recorded in the form of a double occurrence of the same phoneme: for example, [a:] resulted in /aa/. While this is a simplification, it does retain the perceptual length, which languages with long vowels in their phonological systems could utilize for either quantitative iconicity or for emphasizing sound symbolic segments grounded in qualitative iconicity. Coding long and short segments (such as [a] and [a:]) as different sounds would, on the other hand, fail to record the qualitative similarity between them (for example [a] and [a:] coded as /a/ and /a:/), and coding them as the same sound (for example [a] and [a:] both coded as /a/) would not record potential qualitative iconicity.

3.5 Phonetic categorization

Sound-meaning associations are seldom restricted to one specific phoneme; sounds with similar phonetic characteristics are often used for the same meaning in different languages depending solely on what sounds are accessible for the languages in question. If the purpose is to find statistical evidence for cross-linguistic sound-meaning associations, it would seem unwise to count a plain bilabial [m], a creaky voice bilabial [m̰] and a plain labiodental [ɱ] as separate phonemes due to their phonetic similarity. In addition, sound symbolic associations may not necessarily be grounded in phonemes as such, but rather in the acoustic and/or motor features that define them (Sidhu & Pexman 2018). For example, an association between [m] and MOTHER might only be based on its nasal, and not labial, quality. Thus, as this element has not been incorporated in previous studies, it is crucial to systematically group phonetic parameters to pinpoint the features responsible for each sound symbolic association. There are various ways of analyzing and grouping the features of human speech sounds, but cross-linguistic frequencies of sounds as well as phonetic and phonological similarity are generally the most informative parameters that can be used for this purpose (Mielke 2012). Similarly, sounds can be reduced to a set of distinctive features which can be used to describe most sound classes (Mielke 2008). However, while most of these distinctions are appropriate for describing languages phonetically and phonologically, several distinctions are not relevant for studying cross-linguistic sound symbolism and in some cases can even muddy statistical analyses. For example, typologically uncommon distinctions are by definition difficult to compare across languages, but more importantly, several distinctive features sometimes have to be grouped in order to expose sound symbolic relationships caused by a more general feature. Therefore, all human speech sounds were grouped according to salient articulatory parameters in conjunction with distinctive acoustic features which have been shown to evoke sound symbolic associations in experimental and cross-linguistic studies.

3.5.1 Vowel groups

In contrast to consonants, vowels are completely gradient in nature and therefore easily colored by neighboring sounds (Lindblad 1998: 111–112). Additionally, vowels can be realized with a lot more individual variation (Fox 1982) and thus benefit from being divided into larger, more general groups than consonants. Vowels were divided according to their main articulatory dimensions, namely height (*[high]*, *[mid]*, *[low]*), backness (*[front]*, *[central]*, *[back]*) and roundedness (*[−round]*, *[+round]*) (Lindblad 1998: 87–110; Stevens 1998: 257–322; Ladefoged 2001: 40–62).

In addition, vowels were also grouped into four groups that correspond more closely to the movement of the tongue and to the principal vowel distinction important for sound symbolism (Lockwood et al. 2016a). Back vowels were divided into high-back or raised (including close central to close back vowels and close back to true mid back vowels, as well as schwa) and low-back or retracted (including open central to open back vowels and open back to open-mid back vowels). Front vowels were aligned with the back vowel groups by splitting them into high-front (including close front to true mid front vowels) and low-front (including open front to open-mid front vowels). Including this four-way distinction is important since height and backness force sound symbolically distinct sounds to be conflated with each other. For example, the confirmed sound symbolically charged sound [i] would always be grouped with either [u] or [a], while these sounds are usually treated as oppositions to [i] in relative iconicity (e. g. Sapir 1929; Newman 1933). For the same reason, a final sound class consisting of the same four extreme vowel positions with added distinctions for unrounded and rounded variants was included as well. Roundedness of vowel groups are indicated by *[−r]*/*[+r]*, e. g. [high-front, −r] ‘high-front unrounded’ and [high-front, +r] ‘high-front rounded’ (see Table 3).

Table 3: Sound classes, with sound groups and corresponding cardinal sounds.

Sound class			**Sound group**	**Cardinal sounds**
Vowel	Simple	Height	[high]	i, y, ɨ, ʉ, ɯ, u, ĩ, ỹ, ɨ̃, ʉ̃, ɯ̃, ũ
			[mid]	e, ø, ə, ɵ, ɤ, o, ẽ, ø̃, ə̃, ɵ̃, ɤ̃, õ
			[low]	a, ɶ, ä, ɒ̈, ɑ, ɒ, ã, ɶ̃, ä̃, ɒ̈̃, ɑ̃, ɒ̃
		Backness	[front]	i, y, e, ø, a, ɶ, ĩ, ỹ,
			[central]	ɨ, ʉ, ə, ɵ, ä, ɒ̈, ɨ̃, ʉ̃, ə̃, ɵ̃, ä̃, ɒ̈̃
			[back]	ɯ, u, ɤ, o, ɑ, ɒ, ɯ̃, ũ, ɤ̃, õ, ɑ̃, ɒ̃
		Roundedness	[−round]	i, ɨ, ɯ, e, ə, ɤ, a, ä, ɑ, ĩ, ɨ̃, ɯ̃, ẽ, ə̃, ɤ̃, ã, ä̃, ɑ̃
			[+round]	y, ʉ, u, ø, ɵ, o, ɶ, ɒ̈, ɒ, ỹ, ʉ̃, ũ, ø̃, ɵ̃, õ, ɶ̃, ɒ̈̃, ɒ̃
	Aggregated	Extreme	[high-front]	i, y, e, ø, ĩ, ỹ, ẽ, ø̃
			[low-front]	a, ɶ, ã, ɶ̃
			[high-back]	ɨ, ʉ, ɯ, u, ə, ɵ, ɤ, o, ɨ̃, ʉ̃, ɯ̃, ũ, ə̃, ɵ̃, ɤ̃, õ
			[low-back]	ä, ɒ̈, ɑ, ɒ, ä̃, ɒ̈̃, ɑ̃, ɒ̃
		Extreme-roundedness	[high-front, −r]	i, e, ĩ, ẽ
			[high-front, +r]	y, ø, ỹ, ø̃
			[low-front, −r]	a, ã
			[low-front, +r]	ɶ, ɶ̃
			[high-back, −r]	ɨ, ɯ, ə, ɤ, ɨ̃, ɯ̃, ə̃, ɤ̃
			[high-back, +r]	ʉ, u, ɵ, o, ʉ̃, ũ, ɵ̃, õ
			[low-back, −r]	ä, ɑ, ä̃, ɑ̃
			[low-back, +r]	ɒ̈, ɒ, ɒ̈̃, ɒ̃
Consonant	Simple	Manner	[nas]	m̥, m, n̥, n, ɲ̊, ɲ, ŋ̊, ŋ
			[stop]	p, b, t, d, c, ɟ, k, g, ʔ
			[cont]	f, v, s, z, ç, j, x, ɣ, h, ɦ
			[vib]	ʙ̥, ʙ, r̥, r, ɽ̊, ɽ, ʀ̥, ʀ, ʜ, ʢ
			[lat]	ɬ, l, ʎ̥, ʎ, ʟ̥, ʟ
		Place	[lab]	m̥, m, p, b, f, v, ʙ̥, ʙ
			[alv]	n̥, n, t, d, s, z, r̥, r, ɬ, l
			[pal]	ɲ̊, ɲ, c, ɟ, ç, j, ɽ̊, ɽ, ʎ̥, ʎ
			[vel]	ŋ̊, ŋ, k, g, x, ɣ, ʀ̥, ʀ, ʟ̥, ʟ
			[glot]	ʔ, h, ɦ, ʜ, ʢ

(*continued*)

Table 3: (*continued*)

Sound class		Sound group	Cardinal sounds
	Voicing	[–voice]	m̥, p, f, ʙ̥, n̥, t, s, ɾ̥, ɬ, ɲ̊, c, ç, ɽ̊, ʎ̥, ŋ̊, k, x, ʀ̥, ʟ̥, ʔ, h, ʜ
		[+voice]	m, b, v, ʙ, n, d, z, r, l, ɲ, ɟ, j, ɽ, ʎ, ŋ, g, ɣ, ʀ, ʟ, ɦ, ʢ
Aggregated	Manner-voicing	[nas, –v]	m̥, n̥, ɲ̊, ŋ̊, ŋ
		[nas, +v]	m, n, ɲ, ŋ
		[stop, –v]	p, t, c, k, ʔ
		[stop, +v]	b, d, ɟ, g
		[cont, –v]	f, s, ç, x, h
		[cont, +v]	v, z, j, ɣ, ɦ
		[vib, –v]	ʙ̥, ɾ̥, ɽ̊, ʀ̥, ʜ
		[vib, +v]	ʙ, r, ɽ, ʀ, ʢ
		[lat, –v]	ɬ, ʎ̥, ʟ̥
		[lat, +v]	l, ʎ, ʟ
	Place-voicing	[lab, –v]	m̥, p, f, ʙ̥
		[lab, +v]	m, b, v, ʙ
		[alv, –v]	n̥, t, s, ɾ̥, ɬ
		[alv, +v]	n, d, z, r, l
		[pal, –v]	ɲ̊, c, ç, ɽ̊, ʎ̥
		[pal, +v]	ɲ, ɟ, j, ɽ, ʎ
		[vel, –v]	ŋ̊, k, x, ʀ̥, ʟ̥
		[vel, +v]	ŋ, g, ɣ, ʀ, ʟ
		[glot, –v]	ʔ, h, ʜ
		[glot, +v]	ɦ, ʢ

3.5.2 Consonants groups

Consonants, on the other hand, fall into more distinct types of sounds. Since the manner of articulation of consonants involves a greater variety of active articulators than that of vowels (Lindblad 1998: 111–112), the boundaries between consonant groups are more easily defined than the boundaries between vowels. Thus, the consonants were divided into five places of articulation and five manners of articulation. The groups based on place of articulation were further subdivided based on passive articulators, which include a general grave-acute distinction (Jakobson et al. 1951).

This distinction between perceptually sharper versus perceptually duller sounds, generated by the hard palate on one side and the soft palate and lips on the other, can be of great import from a sound symbolic point of view

(LaPolla 1994). The oral passive articulators can naturally be dived into two regions: the hard palate, which corresponds to acute sounds, and the lips and the area behind the hard palate, which correspond to grave sounds. Two of these regions are rather large, but sound symbolic mappings can involve more specific places of articulation. For example, palatals are much more frequent than alveolars in diminutives (Alderete & Kochetov 2017), although both sounds are acute, and while velars are often associated with means such as 'hard' and 'bent' (Bolinger 1950; Wichmann et al. 2010), this does not apply to glottals. Thus, we divided these coarser regions further. While the labial articulator cannot be easily subdivided, sounds articulated with the area behind the hard palate can be subdivided into those which are pronounced using the soft palate and those pronounced using the throat.

Likewise, sounds articulated at the hard palate can be subdivided into those pronounced using the alveolar ridge and those pronounced behind it. This further division produces five sound groups: *[lab]ials* (bilabials, labiodentals, linguolabials, labio-palatals, labio-velars), *[alv]eolars* (dentals, alveolars, palato-alveolars), *[pal]atals* (retroflexes, alveolo-palatals, palatals), *[vel]ars* (velars, uvulars) and *[glot]tals* (pharyngeals, glottals). Retroflexes are lower in acoustic frequency than alveolo-palatals and palatals, but since they are typologically rare, placing them in a separate group would hinder statistical analysis. Placing the retroflexes with the dentals, alveolars and palato-alveolars would also be undesirable since those sounds are also higher in acoustic frequency than retroflexes, and it would furthermore deplete the [pal]atal sound group of values, which would also hinder statistical analysis. Another option would be to place them in one of the grave sound groups, but since tactile factors are also central in sound symbolism (Imai et al. 2008; Watanbe et al. 2012; Ludwig & Simner 2013), this is not viable.

Several sounds, such as retroflexes, also affect adjacent sounds by lowering the formants of neighboring vowels, which could be of great sound symbolic import. However, since the present dataset is compiled in text form, studying including effects from acoustic interactions such as these has to be saved for future studies. As for manner of articulation, consonants were divided into five sound groups with distinct sound symbolic functions: nasals, stops, continuants, vibrants and laterals (Hinton et al. 1994; Wichmann et al. 2010; Blasi et al. 2016; Johansson 2017; Westbury et al. 2018). Occlusives produced nasally were placed in the *[nas]al* sound group since nasals have been shown to evoke a number of sound symbolic associations, ranging from nasal and ringing sounds to pronominal meanings (Hinton et al. 1994; Traunmüller 1994). Occlusives produced orally were grouped in the *[stop]* sound group, which is often associated with visual and tactile unevenness or spikiness. While somewhat similar to ejectives, clicks were also grouped under [stop] because the ingressive mechanism tied to the production of clicks can only be used for stops and affricates (Ladefoged & Maddieson 1996: 247). Thus, the [stop] group is a more fitting affiliation for clicks than any other of the major manners of articulation. Likewise, ejectives, which for the most part are voiceless, were grouped with their plain voiceless stop counterparts, implosives with voiced stops as they usually are voiced (Ladefoged 2001: 147–150), and creaky and nasalized consonants with the plain versions of the same phoneme.

Despite having different acoustic profiles, all *[cont]inuants* with the exception of laterals, i. e. fricatives and approximants, were kept as a unitary group since the type of obstruction involved is comparable, as well as rather simple compared to, for example, vibrants. Furthermore, the shared continuant, central, oral features of approximants and fricatives argue in favor of treating them as similar when it comes to sound symbolic utilization (Hinton et al. 1994; Abelin 1999; Westbury 2005; Sidhu & Pexman 2015). In addition, there is no reason to expect a qualitative difference between a true approximant and voiced fricatives with a low degree of turbulent airflow. The varying degree of obstruction on a perceptual level may instead correlate with voicing, since voicelessness increases air flow and turbulence, which again unites voiceless approximant and fricatives. All *[vib]rants*, which are sound symbolically perceived to be wild, rolling, rough and hard (Fónagy 1963; Chastaing 1966), were grouped together since they behave similarly, although they can be pronounced using a single pulse, as in the case for taps/flaps, or with up to five periods, as in the case of trills (Ladefoged & Maddieson 1996: 215–232). Furthermore, there is usually only one rhotic phoneme per language, and it is therefore frequently reanalyzed to fit the native phonology, e. g. Brazilian Portuguese [peʁu] from Spanish [pero]. Lateral sounds, which can occur as fricatives, approximants or vibrants, were grouped separately as *[lat]eral* because of the unique way the airstream travels along the sides of the tongue rather than in the middle of the mouth and because of their recorded associations with smoothness, liquidness and the tongue (Chastaing 1966; Blasi et al. 2016). All consonant groups were further divided based on voicing, and analyses were repeated with and without this voicing distinction. Voicing of consonant groups is indicated by *[−v]*/*[+v]*, e. g. [stop−v] 'voiceless stop' and [stop+v] 'voiced stop' (see Table 3). Although the nature of voicing may differ between sonorants and obstruents, a binary distinction was judged to be the most suitable option for studying sound symbolism in such a large number of diverse languages (Lockwood et al. 2016a). Lastly, a general sound class of voiced and voiceless consonants (*[−voice]*, *[+voice]*) was included (see Table 3).

3.5.3 Cardinal sounds

A drawback of the present sound group-based method is the loss of phonetic granularity. An association between a concept and a sound group does not necessarily mean that all sounds within the sound group are equally overrepresented. In order to compensate for this, we attempted to recapture the sounds which could be the driving factors behind sound-meaning associations by dividing all speech sounds into *cardinal sounds*. For vowels, the three levels of height and backness were combined into nine points of articulation. These nine points could be unrounded, rounded, oral or nasal, e. g. [i], [y], [ĩ] and [ỹ], amounting to 36 cardinal vowels. Likewise, the five generalized levels of consonantal manner and place of articulation were combined and divided into voiceless and voiced versions, e. g. [p] and [b]. As several of the consonant combinations are impossible to articulate, this resulted in 43 cardinal consonants (see Table 3).

3.6 Data analysis

The goal of data analysis was to identify words with over-represented sound groups – for example, words that contain an unexpectedly high proportion of high vowels across the sampled languages. We started with the assumption that each language has a typical distribution of vowels by height (and other features of interest listed in Table 3) and estimated this distribution by looking at all 344 sampled words from that language. If, in many of the sampled languages, a particular word contained a markedly higher proportion of rounded vowels than the average for each language, we interpreted this as evidence that some force, such as sound symbolism, was driving this non-arbitrary word form.

Calculating the absolute number of phonemes occurring within a word could skew the results through, for example, reduplication and effects of word length. Previous comparable studies did not include concepts which often involve reduplication (kinship concepts, numerals, etc.); hence, reduplication and similar phenomena were not controlled for, even though these phenomena also affect a range of other basic vocabulary items. Furthermore, the aim of this study was to investigate the occurrence of phonemes across languages, not their occurrence within specific linguistic forms. To avoid this problem, we chose to analyze proportions rather than absolute counts of sound groups. These proportions were calculated separately for vowels and consonants. So, for example, the word /mantu/ (“belly” in the Ngarinyin language [ung]) contains 66% of voiced and 33% of unvoiced consonants; 50% of high, 0% of mid, and 50% of low vowels; and so on. A hypothetical complete reduplication into /mantu-mantu/ would have no effect on these proportions, and it would therefore remain “invisible” to the model. In contrast, a partial reduplication of one syllable (e. g. /mantu-tu/) would affect the proportions of sound groups. Likewise, because long vowels and diphthongs were coded as two separate phonemes (e. g. /ma:ntu/ would be coded as /maantu/), sound symbolic prolongation of vowels was captured by the models we used. As a “bonus”, this approach also solves the problem of some concepts being represented by more than one linguistic form, e. g. the English *he*, *she*, *it*.

A transformed dataset of proportions was prepared and modeled separately for each of 10 evaluated sound groups: backness, height, roundedness, extreme and extreme-roundedness for vowels; manner, manner-voicing, place, place-voicing and voicing for consonants (Table 3). One row in the dataset corresponded to one word in one language, and the response variable was a vector of proportions that summed to one – in mathematical terms, a simplex. We modeled these simplex responses with the Dirichlet distribution in the framework of Bayesian generalized linear models (GLM) as implemented in the R package *brms* version 2.9.0 (Buerkner 2017), with default conservative priors.

Using vowel height as an example, the model included a population-level intercept corresponding to the overall distribution of vowels by height across all words and languages, a group-level (random) intercept per language corresponding to the typical distribution of high, low and mid vowels in each particular language, and a group-level (random) intercept per word. This random intercept per word was the measure of interest, since it showed deviations from the typical distribution of vowels by height in particular words. As usual with multilevel models, representing proportions for each word and language as drawn from a single distribution imposed shrinkage – that is, drew the estimates closer to the group mean. The amount of shrinkage was controlled adaptively by the data itself, which is a great advantage of multilevel models and the reason why the effect of word was modeled as a random rather than fixed effect. Shrinkage was stronger when the outcome variable had many levels and more moderate for outcomes with two levels, such as voicing and roundedness; it was also stronger for rare sound groups with relatively few observations (e. g. voiced glottals), where the apparent outliers were driven by only a few languages (Online Appendix 3).

The output of interest from these Dirichlet models was a list of fitted proportions of sound groups (e. g. of high, low and mid vowels) in each of 344 words. To identify cases of over- or underrepresentation, we also extracted fitted average proportions of each sound group (e. g. high vowels) across all words and then compared per-word estimates to these average values. To propagate uncertainty of model estimates, this comparison was performed for each step in the Markov chain Monte Carlo, resulting in a posterior distribution of how much each word deviated from the typical distribution of sound groups.

One way to compare distributions would be to look at simple differences of proportions of each class. For example, if the typical proportion of high vowels is 50% and a word contains (on average across all languages) 55% of high vowels, this constitutes a 5% overrepresentation. The problem with this approach is that it does not scale very well for proportions that are close to 0% or 100%. For example, if the frequency of a rare sound group jumps from a base rate of 5% to 10% in a particular word, this is substantively a greater change than from 50% to 55%. To account for this, we compared odds ratios (OR): an increase from 5% to 10% corresponds to OR = 1:9/1:19 = 2.1, while an increase from 50% to 55% gives OR = 11:9/1:1 = 1.2.

Since we employed a Bayesian analysis, we did not test the statistical significance of any effects. Instead, we defined a region of practical equivalence (ROPE), symmetric on a logarithmic scale, around the null effect of no overrepresentation (log-odds ratio = 0 or, equivalently, OR = 1). The width of the ROPE corresponded to a change of OR by a factor 1.25 1 * 1.25 = 1.25, or +25%; 1/1.25 = 0.8, or −20%). This ROPE was set to represent the smallest substantively interesting effect size: a 25% increase of OR corresponds to an increase in the proportion of a sound in a word from 10% to 12%, 50% to 55.5%, 90% to 92%, etc. Following the guidelines for decision making in this analytical framework (Kruschke & Liddell 2018), we distinguished between three types of outcome:

(1) "Strong association": if the 95% credible interval (CI) for the OR fell completely outside the ROPE, we concluded that the distribution of sound group in this word substantively deviated from the distribution expected by chance.

(2) "No association": if the 95% CI was completely contained inside the ROPE, we concluded that there was no over- or underrepresentation.

(3) If the 95% CI partly overlapped with the ROPE, the result was treated as ambiguous. Because there was a substantial number (~9%) of such cases, we further distinguished between two subtypes. If the 95% CI excluded zero and the median of posterior distribution (our "best guess") was outside the ROPE, the association was treated as "weak" but potentially interesting; otherwise it was treated as too uncertain for being considered further.

It is worth emphasizing that the ROPEs refer to fitted rather than observed values. In most models and categories, shrinkage of regression coefficients to zero was very pronounced (see Online Appendix 3), thus producing very conservative estimates of the degree of under- or overrepresentation. As a result, the number of associations reported below (225, or ~1.3%) is vastly lower than the number of cases for which the observed OR lies outside the same ROPE (6708, or ~36% of all possible associations).

4 Results and analysis

4.1 General results

The total number of potential associations was very large, varying from 344 in models with two sound groups (e. g. voiced or unvoiced consonants, rounded or unrounded vowels) to 3096 in the models with ten sound groups (Place-voicing and Manner-voicing), for a total of 17,888 possible associations across ten models. However, an overwhelming majority of associations was classified as absent (90.8%) or doubtful (7.9%), leaving only 176 (1.0%) weak and 49 (0.3%) strong associations (Figures 2 and 3). These numbers exclude cases of underrepresentation of sound groups with two levels (vowel roundedness and consonant voicing) since these were redundant mirror images of overrepresentations. For example, if rounded vowels are overrepresented in a particular word, unrounded vowels register as equally underrepresented. Cases of underrepresentation could be of some interest for oppositional concepts of binary or continuous domains. For example, sounds overrepresented in BIG might be underrepresented in its opposite SMALL to emphasize the contrast. However, the results yielded few clear sound symbolic antonyms, making the negative associations difficult to interpret: there could be many reasons why some sounds seldom occur in a specific word. In some cases, underrepresented sounds could be a consequence of other classes being strongly overrepresented, particularly in short words. However, we did not find any correlation between word length and the probability of a word being sound symbolically affected. Thus, we only focus on the overrepresented associations in the following discussion.

If we compare the found associations with previous similar studies, it becomes evident that the investigated concepts have varied considerably, see Table 4. For the present study, the associated sound group with the highest specificity is listed, e. g. if a concept was associated with [high], [high-back] and [high-back, +r], only [high-back, +r] was listed. 19 concepts and 20 associations (ASHES-[back], BONE-[−voice], BREAST-[nas, +v], F_FS/F_MS-[lab]/[low-front, −r], 1SG-[nas, +v]/[−round], KNEE-[+round], M_FS/M_MS-[nas, +v], NOSE-[nas, +v], SKIN-[−voice], TONGUE-[alv, +v], 1PLI/1PLE-[nas, +v], 2SG-[nas, +v]/[−round]) also clearly correlate with those reported by Johansson (2017), Wichmann et al. (2010) and Blasi et al. (2016) and therefore ought to be considered very robust. In addition, several of the associations were also found to be similar to previous findings. For example, HARD was found to be associated with voiceless alveolars by Johansson (2017) and with (mostly voiceless) stops in the present paper.

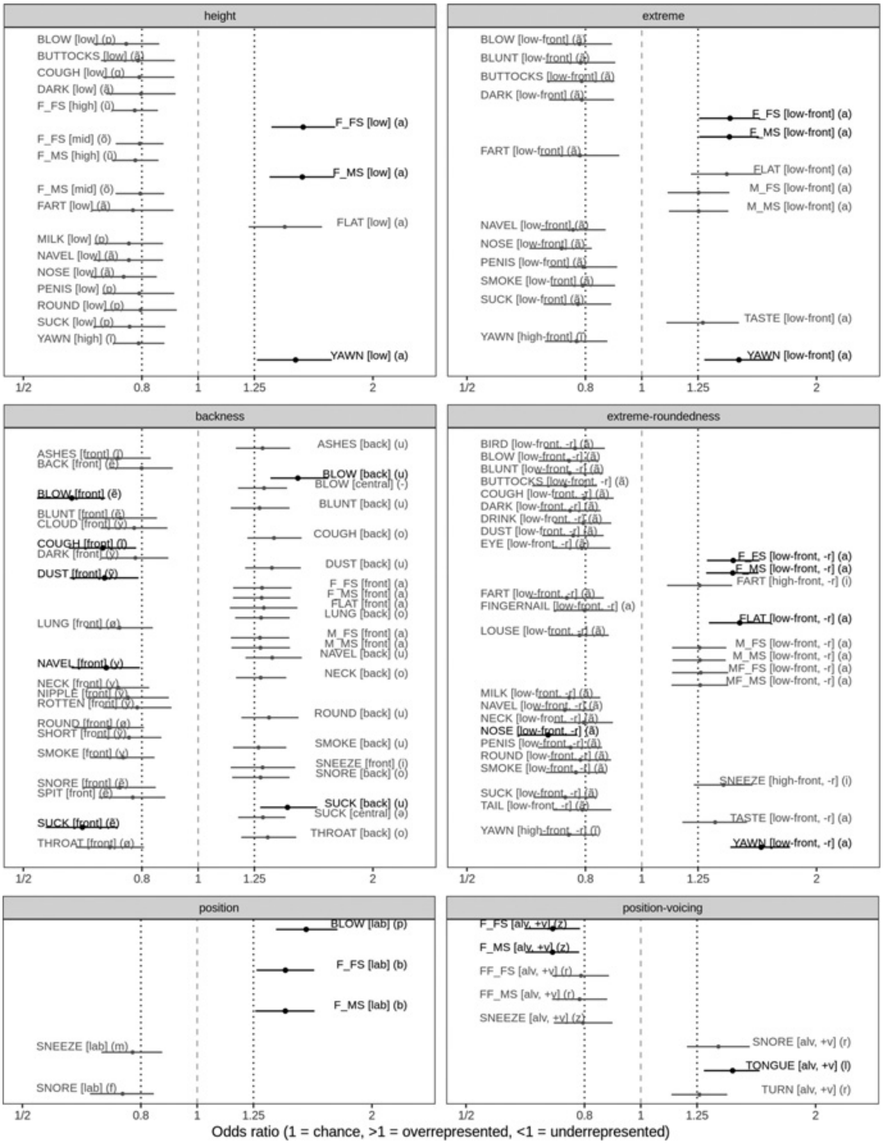


Figure 2: Over- and underrepresented sound groups in 344 concepts: strong (black) and weak (gray) associations. Each point shows the median of posterior distribution of the ratio of observed to expected odds, with 95% CI. Text labels show the concept, associated sound group, and the most strongly over- or underrepresented cardinal sound in parentheses. The marked region of practical equivalence (ROPE) of [0.8, 1.25] was used to select substantively relevant findings.

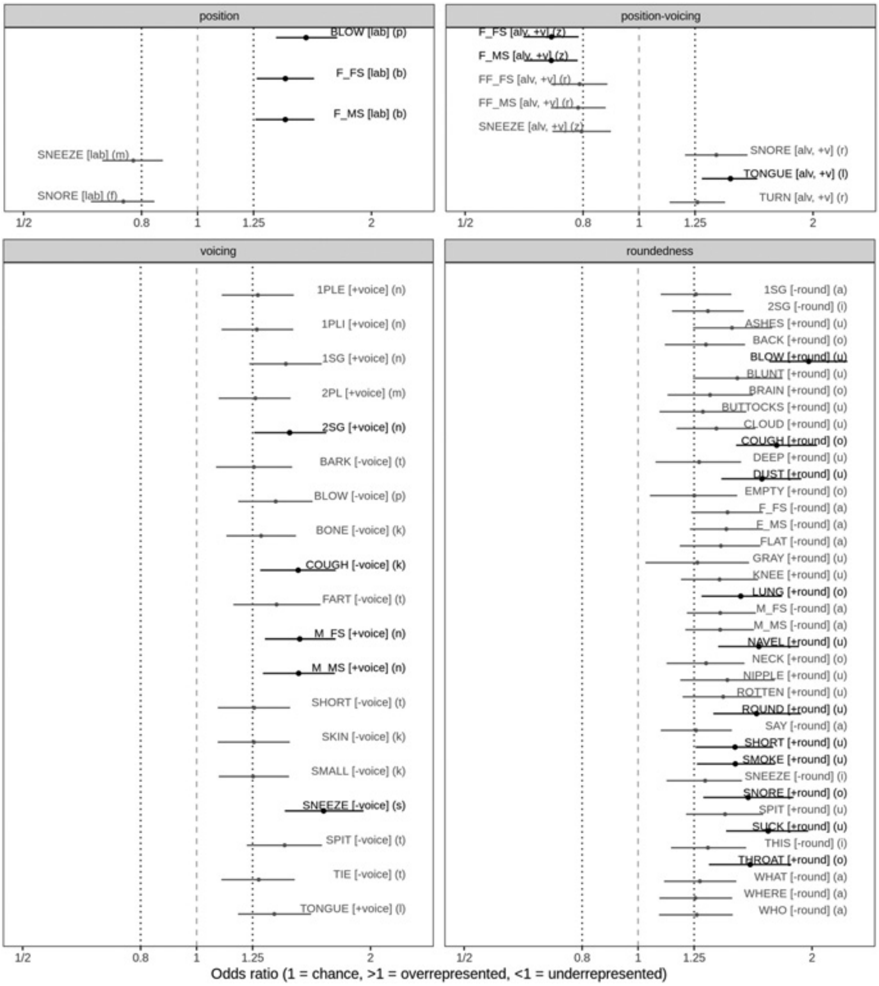


Figure 3: Over- and underrepresented sound groups in 344 concepts: strong (black) and weak (gray) associations. Each point shows the median of posterior distribution of the ratio of observed to expected odds, with 95% CI. Text labels show the concept, associated sound group, and the most strongly over- or underrepresented cardinal sound in parentheses. The marked region of practical equivalence (ROPE) of [0.8, 1.25] was used to select substantively relevant findings.

Likewise, while SHORT and SMALL were found to be associated with voiceless alveolars, /i/ and /C/ by Johansson (2017) and Blasi et al. (2016) but with [stop, −v] (as well as [rounded]) in the present results, all sound groups involve high frequency sounds. Furthermore, the present study found another 39

Table 4: Comparison of all positive associations (overrepresentations) between sounds and meanings identified in previous large-scale cross-linguistic studies on sound symbolism and the current study. Gray indicates that the concept was not included in the study and a dash (-) that the concept was investigated but no association was found. The results from the present study include the associated sound groups with the highest specificity, followed by the most commonly occurring cardinal sound from that sound group in parentheses.

Study	**Wichmann et al. (2010)**	**Blasi et al. (2016)**	**Johansson (2017)**	**Present study**
Languages (families)	3,000 + (170)	4,000 + (359)	75 (39)	245 (245)
Lexemes	40	40	56	344
ASH(ES)		u		[back] (u)
BITE		k		–
BONE	–	k		[–voice] (k)
BREAST(S)	muma	u m		[nas, +v] (m)
COLD		–	voiceless velar	–
DAY			lateral	–
DEEP		–	vibrant, lateral	[+round] (u)
DOG	–	s		–
EAR	–	k		–
FATHER			/a/-like, voiceless labial	[lab] (b), [stop] (t), [low-front, –r] (a)
FEW			voiceless alveolar	–
FISH	–	a		–
FLAT			voiceless labial, lateral	[low-front, –r] (a)
FULL	–	p b	voiceless alveolar, voiceless labial	–
HARD			voiceless alveolar, vibrant	[stop] (k)
HEAR	–	N		–
HORN	–	k r		–
I	naa	5	nasal	[nas, +v] (n), [–round] (a)
KNEE	kokaau	o u p k q		[+round] (u)
LEAF	aaaa	b p l		–
LIGHT (not DARK)			vibrant	–
LONG		–	voiced velar, lateral	–
MOTHER			nasal	[nas, +v] (n), [low-front, –r] (a)
NAME	nani	i		–

(*continued*)

Table 4: (*continued*)

Study	Wichmann et al. (2010)	Blasi et al. (2016)	Johansson (2017)	Present study
NARROW	▨	▨	voiceless alveolar	–
NOSE	nani	u n	▨	[nas, +v] (n)
OLD	▨	▨	vibrant	–
ONE	–	t n	▨	–
RED	▨	r	▨	–
ROUGH	▨	▨	voiceless alveolar, fricative, vibrant	–
ROUND	▨	r	vibrant	[back] (u)
SAND	▨	s	▨	
SHORT	▨	▨	voiceless alveolar	[stop, −v] (t)
SKIN	kaaa	–	▨	[−voice] (k)
SMALL	▨	i C	voiceless alveolar	[−voice] (k)
SMOOTH	▨	▨	vibrant, lateral	–
STAR	–	z	▨	–
STONE	–	t	▨	–
THIS	▨	–	voiced palatal	[nas] (n), [−round] (i)
THAT	▨	–	–	–
TONGUE	–	e E l	▨	[alv, +v] (l)
WE	–	n	▨	[nas, +v] (n), [−round] (a)
WET	▨	▨	voiceless alveolar	–
WHITE	▨	–	vibrant	–
WIDE	▨	▨	lateral	–
YOU	nin	–	–	[nas, +v] (n), [−round] (i)

concepts and 105 associations which are described in Section 4.2. There were, however, also several discrepancies between the present and previous studies. Johansson (2017) found several associations to sound groups that generally contain few sounds, i. e. vibrants, laterals and voiced palatals in DEEP, FLAT, HARD and THIS. This is likely a result of the considerably smaller and less balanced sample of languages and the less robust statistical analysis. It is possible that the overrepresentation of voiceless labials in FLAT found by the same study is a similar case. Both Blasi et al. (2016) and Johansson (2017) also found associations between ROUND and vibrants, while the present study found associations to [back] (as well as [+round]), mainly represented by /u/. The association between ROUND and rounded sounds is further discussed in Section 4.2.2 but is rather straightforward to understand. However, the "lack" of overrepresentation of vibrants could be attributed to the strict modeling used in the present study which might have created a higher confirmation threshold for the investigated sound-meaning associations compared to previous studies. Our cautious approach could also therefore have resulted in the loss of several potential associations. What is more, the semantically similar concept TURN was found to be associated with [alv, +v] (mainly represented by /r/), which also suggests a connection between circular shapes and vibrants. A full list of all associations and concepts is found in Online Appendix 1 along with cardinal sounds, overall occurrence, as well as the type of sound symbolic mapping and associated macro-concept, as explained below.

4.2 Macro-concepts based on semantic and phonetic common denominators

Overall, all of the discovered sound-meaning associations belonged to bodily functions, body parts, deixis, descriptors, kinship terms, logical concepts, or natural entities. More interestingly, however, the concepts with noteworthy overrepresentations could in turn be grouped into semantically and sometimes phonetically superordinate concepts, here referred to as *macro-concepts*. Arranging the discovered associations in this manner has several benefits: a) it provides an overview of the rather long list of confirmed sound-meaning associations in an exploratory study such as the present one, and b) it makes it possible to use observable semantic and phonetic regularities to further understand how sound symbolism could be used to define fundamental lexical fields in human language. The macro-concepts should therefore be regarded as preliminary classifications, but could still act as a stepping stone for future studies. This grouping required that the confirmed sound-meaning associations shared both semantic and phonetic features and was defined as follows.

Strong macro-concepts had to include at least one of the strong sound-meaning associations or at least two weak sound-meaning associations. For strong macro-concepts consisting of more than one sound-meaning association, the included associations also had to share one or more concepts that share at least one semantic feature and one or more concepts that share at least one associated sound.

Weak macro-concepts had to include one of the weak sound-meaning associations which could be corroborated by a qualitatively parallel association or macro-concept (e. g. associations between LARGE and low-frequency sounds, and SMALL and high-frequency sounds) or by a plausible sound symbolic explanation in line with known associations reported in other studies on sound symbolism and iconicity. When evaluating shared sound-meaning associations, the most commonly occurring cardinal sounds were taken into account as these are informative in regard to the driving factors behind associations. For example, the effect of an association between a concept and [stop, −v] and [lab, −v] could be driven by an intersecting /p/ in both cases.

This further means that a concept, particularly concepts associated with more than one sound group, can belong to several macro-concepts, and macro-concepts can include various sound groups as long as those sound groups share relevant phonetic features. In addition, the interaction between semantic and phonetic features, as well as cardinal sounds, also makes it possible to trace which type of sound symbolic mapping grounded each sound-meaning association. As the study was designed to be explorative, all possible types were of interest. However, basing the calculated results on relative frequencies of sounds washed away internal word structure patterns, making it impossible to analyze gestalt iconicity, i. e. cross-modal, iconic or indexical mappings of word-internal structural emergence. For example, reduplication occurs frequently in some languages but is almost absent in others. To complicate things even more, words can be reduplicated either completely (e. g. Basque [eus] *zapla-zapla* ‘slap’) or partially (e. g. Pangasinan [pag] *toó* ‘man’ and *totóo* ‘people’). Phenomena such as phonesthemes, i. e. associative, cross-modal, indexical mappings of language-internal analogical emergence, also had to be excluded since they are not detectable due to their language-specific character.

A complete list of all macro-concepts, their contained concepts and associated sound groups, as well as the most frequently occurring cardinal sounds in each sound group association and sound symbolic mapping types, are provided in Table 5.

In total, the results revealed 134 sound-meaning associations. We did not find any plausible explanation for the associations between BACK and [+round], EMPTY and [+round], THINK and [nas, +v] and TIE and [−voice], while those between BLOW and [central] and SUCK and [central] were only found in 11 and 14 languages, respectively. Therefore, these associations were judged as doubtful. Furthermore, SHORT was unexpectedly associated with [+round] which would be the reverse of the expected pattern of ‘small’-high frequency and ‘large’-low frequency (see Section 4.2.3). However, as this association does not correlate with any previous findings, it is quite possible that it is a result of noise in the source materials. This association was thus also judged as doubtful. These were therefore excluded from further analysis, resulting in a grand total of 125

Table 5: Macro-concepts with contained concepts (possibly involved concepts in parentheses), their associated sound groups (the most commonly occurring cardinal sounds in parentheses) and the type of sound symbolic mapping (certainty in parentheses) which are defined and discussed throughout Section 4.

Macro-concept	Contained concepts: certain (possible)	Associated sound groups	Primary cardinal sounds	Mapping (certainty)*
AIRFLOW	ASHES, BLOW, CLOUD, DUST, SMOKE, (GRAY)	[–voice], [lab], [+round], [back]	p, u	O (strong)
PHARYNGEAL	COUGH, LUNG, SNORE, THROAT	[–voice], [+round], [back]	k, o	O (strong)
EXPULSION	FART, SNEEZE, SPIT	[–voice], [–round], [front], [high-front, +r]	t, s, i	O (strong)
GAPING	TASTE, YAWN	[low], [low-front], [low-front, –r]	a	O/V (strong)
UNEVEN	BARK, SKIN, SNORE	[–voice], [alv+v]	k, t, r	O/V (strong)
ROUNDNESS	BLUNT, BUTTOCKS, KNEE, NAVEL, NECK, NIPPLE, ROUND	[+round], [back]	o, u	V (strong)
FLAT	FLAT	[–round], [front], [low], [low-front], [low-front, –r]	a	V (strong)
TONGUE	TONGUE	[+voice], [alv, +v]	l	V (strong)
NOSE	NOSE	[nas, +v]	n	V (weak)
TURN	TURN	[alv, +v]	r	V (weak)
SMALLNESS	SHORT, SMALL	[–voice], [stop], [stop, –v]	t, k	R (strong)
DEEP	DEEP	[+round]	u	R (weak)
HARDNESS	HARD, BONE	[–voice], [stop]	k	R (strong)
SOFTNESS	BRAIN, BUTTOCKS, ROTTEN	[+round]	o, u	R (strong)
QUESTION	WHAT, WHERE, WHO, (SAY)	[–round]	a	R (strong)
MOTHER	M_FS, M_MS	[voiced], [nas], [nas, +v], [–round], [front], [low-front], [low-front, –r]	n, a	C (strong)
FATHER	F_FS, F_MS	[lab], [stop], [–round], [front], [low], [low-front], [low-front, –r]	b, t, a	C (strong)
RELATIVE	MF_FS, MF_MS	[low-front, –r]	a	C (weak)
INFANCY	BREAST, M_FS, M_MS, MILK, NIPPLE, SUCK	[+voice], [nas], [nas, +v], [+round], [back]	m, n, u	C/V (strong)
DEIXIS	1SG, 2SG, 3SG, 1PLI, 1PLE, 2PL, THIS	[+voice], [nas], [nas, +v], [–round]	m, n, a, i	C/R (strong)

*O: onomatopoeia, V: vocal gestures, R: relative, C: circumstantial.

relevant associations involving 59 concepts. In addition, since an association can be grounded in more than one way simultaneously, e. g. both through visual and acoustic motivations, there were in total 140 sound symbolic motivations. In turn, these motivations were found across four types of mappings, of which two, vocal gestures and circumstantial mappings, have not previously been explicitly described in the sound symbolic literature (summarized in Figure 4).

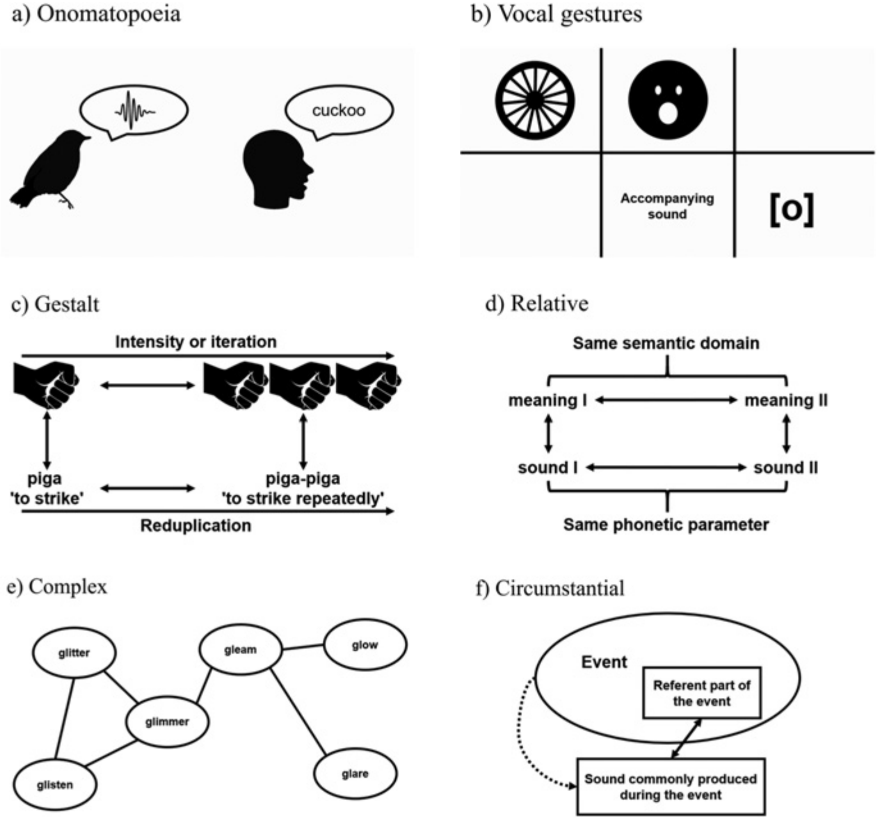


Figure 4: Illustrated simplifications of types of sound symbolism described by Dingemanse (2011), Carling and Johansson (2014) and the present paper. a) Onomatopoeia (imitative): acoustic approximations using the human vocal apparatus. b) Vocal gestures (imitative): cross-modal imitations in which the acoustic signals are only accompanying the gesture. c) Gestalt (diagrammatic): mappings between event structures and word structures, e. g. Swahili *piga* 'to strike' and its reduplicated form *piga-piga* 'to strike repeatedly'. d) Relative (diagrammatic): relational mappings between semantic and phonetic scales or poles. e) Complex (associative): language-internal mappings which emerge through analogy. f) Circumstantial (associative): mappings based on circumstantial associations between referents which are part of an event and sounds which are frequently expressed during the same event.

Summarized, of the 140 motivations, 37 (26.4%) were defined as onomatopoeia, 31 (22.1%) as vocal gestures, 16 (11.4%) as relative, 57 (40.7%) as circumstantial and 7 (5%) remain doubtful. Furthermore, macro-concepts consisting of a single concept could in fact be members of yet undefined larger macro-concepts that remain opaque since they include concepts not featured in the present sample.

4.2.1 Primarily onomatopoeic mappings

Several of the concepts related to bodily functions were often found to have full-word onomatopoeic forms, i. e. uni-modal, iconic mappings of direct emergence based on sound imitation (Hinton et al. 1994; Dingemanse 2011; Dingemanse et al. 2015; Carling & Johansson 2014), in which manner and place of articulation as well as function were featured in their sound symbolic mappings. BLOW and the semantically related concepts ASHES, CLOUD, DUST and SMOKE all involve air moving or fine material moving through air. Phonetically, these concepts were associated with vowel sound groups ([+round] and [back]) in which the most commonly occurring cardinal sound was /u/, as well as [lab] and [–voice] in which the most commonly occurring cardinal sound was /p/. The associated sounds seem to all involve labial components and the macro-concept AIRFLOW could therefore be onomatopoeically grounded in the fact that lip rounding regulates the amount of air that is passed through the mouth and thereby intensifies friction on both acoustic and tactile levels. Colors that are lexicalized late, such as 'gray', 'purple', 'pink' and 'orange', tend to be derived from concrete referents. Thus, it is also possible that GRAY belongs to AIRFLOW indirectly since it also contains rounded vowels and is often derived from words for 'ashes'.

COUGH, LUNG, SNORE and THROAT were also associated with [+round] and [back], but instead of /u/, the most commonly occurring cardinal sound was /o/ in all cases. In addition, COUGH was also associated with [–voice] which was represented by the cardinal sound /k/. This seems to suggest that the common phonetic denominator in the macro-concept PHARYNGEAL involves the back of oral cavity and possibly also a somewhat more open mouth than the vowels of AIRFLOW.

In contrast to AIRFLOW and PHARYNGEAL, FART, SNEEZE and SPIT were associated with vowel sound groups ([–round], [front], [high-front, +r]) in which the most commonly occurring cardinal sound was /i/. These concepts, which constitute the macro-concept EXPULSION, were also associated with [–voice] represented by the cardinal sounds /t/ and /s/. Thus, this onomatopoeic macro-concept can be explained by the associated sounds' energy distribution in high frequencies and the sounds produced by FART, SNEEZE and SPIT (Taitz et al. 2018).

In a similar fashion to how rounded vowels represent AIRFLOW, the macro-concept GAPING (consisting of TASTE and YAWN) was represented by its association to [low], [low-front] and [low-front, −r], which of course mainly involved /a/. Furthermore, it is possible that the associated sounds are indirectly associated, while the gesture producing them is the fundamental ground for this association (see Section 4.2.2).

The concepts with UNEVEN semantic features (BARK, SKIN and probably SNORE) were associated with sound groups with turbulent, pulsating airflow, probably grounded in the shared features of sounds produced when running an object over an uneven surface and the tactile unevenness. Among these sound groups, we find [alv, +v] which mainly consisted of the pulsating trill, /r/, (Ladefoged & Maddieson 1996: 215–232). We also find [−voice], in which the most commonly occurring cardinal sounds were /k/ and /t/. This association might be grounded in the irregular, noisy airflow created by many typologically common voiceless obstruents. The apparent tactile sensation produced by vibrating sounds further suggests that this macro-concept could be motivated through both onomatopoeia and *vocal gestures* (see Section 4.2.2).

4.2.2 Primarily vocal gesture mappings

Several more macro-concepts appear to be based on imitation, in which the referents are perceived cross-modally and indexically through other senses than hearing (here referred to as *vocal gestures*). In these mappings, the articulatory gesture is mapped to the referent and the sounds produced are only secondarily associated. For example, the noticeably round concepts of the macro-conept ROUNDNESS – BLUNT, BUTTOCKS, KNEE, NAVEL, NECK, NIPPLE and ROUND – were associated with the vowel groups [+round] and [back], which mainly consisted of the rounded cardinal sounds /u/ and /o/. The ground for this association could lie in the rounded shape that the mouth assumes when producing rounded sounds and not in the acoustic signals themselves. Therefore, the acoustic signals are simply accompanying the articulatory gesture and are associated with the referent only by being attached to the articulatory gesture. Thus, rounding one's lips to denote that something is round is indeed iconic, but the accompanying sound is not. For example, if the articulatory gestures of a [u] could produce the acoustic properties of an [i], the sound symbolic mapping between [u] and the meaning round would still be functioning (Jones et al. 2014).

FLAT was associated with several vowel sound groups of varying specificity ([−round], [front], [low], [low-front], [low-front, −r]), but in all of them the most commonly occurring cardinal sound was /a/. The ground for this association could lie in the appearance and sensation produced by having the tongue level and extended at the bottom of the mouth.

The body part macro-concept TONGUE could be established through its association with [+voice] and [alv, +v] which mostly involved /l/. This association could be explained by the fact that the tongue can be made visible when alveolar laterals are continuously produced, as opposed to alveolar stops, nasals, sibilants and vibrants, and that alveolar laterals are typologically more common than [θ] and [ð]. The weak body part macro-concept NOSE could be established through its association with [nas, +v] (the sounds produced using the nose).

The connection between (rapid) movement or continuity and vibrants was in the present sample represented by the associations of [alv, +v], primarily involving /r/, with TURN, and mentioned in some of the earliest studies on sound symbolism (Plato’s *Cratylos* [Sedley 2003], Humboldt 1838; Jespersen 1922; Fónagy 1963). Vibrants are made of a series of pulses (Ladefoged & Maddieson 1996: 215–232), which are individually distinguishable, but too rapid to be counted, and bear similarities to e. g. quick steps.

4.2.3 Primarily relative mappings

Intensity is a common cross-modal dimension applied to the oppositional poles of light, sound, smell, taste, pain, emotion, etc. and clearly visible in linguistic labels. For example, sounds and lights can be bright or dull, and ‘long’ and ‘short’ can refer to physical objects and durations (Levinson & Majid 2014). It therefore comes as no surprise that the results revealed macro-concepts that were descriptive in nature or even adjective-like, based on relative sound symbolism such as the thoroughly studied mapping between small-large and high-low frequency in pitch (Sapir 1929; Ohala 1994). SHORT and SMALL were associated with [−voice], [stop] and [stop, −v], which consisted of the high-pitched sounds /t/ and /k/ and thus constituted the SMALLNESS macro-concept (Dolscheid et al. 2012). Conversely, DEEP was associated with [+round] and driven by /u/, which generally corresponds to low-frequency sounds.

Similarly to SMALLNESS, the macro-concept HARDNESS could be established by grouping the phonetic features shared by HARD and BONE: [−voice] and [stop] (consisting of /k/) (compare also the association between *bone* and *k* reported by Blasi et al. 2016). In contrast, the corresponding macro-concept SOFTNESS could be formed through BRAIN, BUTTOCKS and ROTTEN and their associations with [+round], driven by /o/ and /u/. It should furthermore be noted that markedness might play an important role in relative sound symbolism (compare de Villiers and de Villiers' 1978: 139–141 work on semantic markedness and learnability). For example, the unmarked pole of oppositional meanings, such as 'hard' and 'soft', are generally understood earlier by children than the marked pole. Thus, it is also possible that only one of the poles is more sound symbolically charged since the other pole could be defined primarily by contrasting with the first.

The associations between the question concepts WHAT, WHERE and WHO (possibly also along the semantically related concept SAY) and [–round], i. e. mostly /a/, could be explained by the fact that interjections such as *huh?* occur cross-linguistically as a conversational repair initiator, as they often contain a a mid-to-low and front-to-central vowel with rising intonation (Dingemanse et al. 2013). Dingemanse et al. mainly attributed this cross-linguistic similarity to convergent evolution shaped by interactional selective pressures rather than being based on some sort of innate human grunting sound. However, it should be mentioned that, according to the frequency code (Ohala 1994), high frequency sounds and rising intonation indicate insecurity, questioning, etc.

4.2.4 Primarily circumstantial mappings

The results also exposed *circumstantial sound symbolism*, an associative language-external mapping which has less to do with how the association operates and more to do with its circumstantial emergence, in many ways similar to complex iconicity (Carling & Johansson 2014) since it is cross-modally and indexically mapped. For example, if infants were able to produce other sounds while breastfeeding, the macro-concept MOTHER (M_fs, M_MS) would probably not be associated (only) with [+voice], [nas] and [nas, +v] (/m/ being the most overrepresented and /n/ the most common cardinal sound), and [–round], [front], [low-front] and [low-front, –r], which were all represented by the cardinal sound /a/. Thus, this type of sound symbolism appears to be grounded in the sounds that are produced in very specific situations tied to our life world (Gibson 1977).

The concepts including the notions of FATHER, F_FS and F_MS, were associated with a similar set of vowel sound groups ([–round], [front], [low], [low-front], [low-front, –r]), which were also represented by the cardinal sound /a/. They were also associated with [lab] and [stop], which featured /b/ and the more typologically common sound /t/ as the most commonly occurring cardinal sound. All remaining sound-symbolic kinship terms referred to grandparents

(MF_FS, MF_MS) and were also associated with [low-front, −r], represented by /a/, and were grouped under the macro-concept RELATIVE. Despite the fact that lexical and phonological influences create language-specific differences in language development, the consonants first acquired by infants generally tend to be [m], [n] and [p], followed by [b] and [w], and the first acquired vowel is [a] (Sander 1972). At the same time, these sounds are cross-linguistically very common (Maddieson 1984; Moran et al. 2014). However, phonetic acquisition explains only parts of these associations, at least in the case of nasal sounds.

The macro-concept INFANCY was established by including M_FS and M_MS, as well as BREAST and MILK, which were all associated with the nasal sound groups [nas] and [nas, +v]. A possible explanation is that nasal sounds are commonly produced by infants while breastfeeding since their mouths are obstructed, hindering breathing through the mouth and oral sound production (Swadesh 1971: 191–199; Traunmüller 1994; Jakobson 1962; Wichmann et al. 2010; Johansson 2017). Furthermore, the semantically related concepts SUCK and NIPPLE were associated with [+round] and [back], driven by /u/. These associations resemble the connection between AIRFLOW and labial sounds, but the motivation is different. Instead of causing friction to amplify the sound of air leaving the body, the rounded vowels in INFANCY appear to be mapped through the suckling motion involved in breastfeeding and other acts involving sucking via vocal gestures.

Pronouns, alongside other deictic concepts (Traunmüller 1994), were also found to be extensively affected by sound symbolism. Six of the seven featured personal pronoun concepts were associated with [+voice], [nas] and [nas, +v], represented by /m/ and /n/. Nasal sounds therefore seem to be associated with indexicality beyond the ego and personal pronouns (Johansson 2017). In addition, this macro-concept, DEIXIS, was also associated with [−round], driven by /a/ and /i/, which also correlate with SMALLNESS and DEEP. SMALLNESS was associated with sounds with energy distribution in high frequencies while DEEP was associated with a low-frequency sound group. Thus, it seems plausible that the deictic concepts correlate with other sound symbolic concepts that denote small size, as it can easily be translated into small distance and proximity.

Linguistic forms such as *mama*, *nana* etc., relating to ‘mother’, ‘breast’ or similar, have often been explained by baby talk or babbling, despite their cross-linguistic salience (Nichols 1999; de l’Etang et al. 2008; Bancel et al. 2013). However, there could be a concrete motivation for their associations with nasals which cannot be attributed to imitation or relative mappings. Social interaction is one of the most important components in early language acquisition (Fromkin et al. 1974), which also applies to non-human vocal learners (Beecher 2017). Theofanopoulou et al. (2017) suggests that oxytocin plays a major role in social motivation and vocal learning. Oxytocin also facilitates language learning since it regulates biological processes related to childbearing and bonding, such as breastfeeding, and it has been linked to semantic integration in speech comprehension (Ye et al. 2016), verbal communication (Zhang et al. 2016) and directed singing in songbirds (Pedersen & Tomaszycki 2012). As stated above, infants tend to produce nasal sounds while breastfeeding, which also constitutes a considerable amount of the infants' time spent awake. Thus, the high emissions of oxytocin combined with the frequent production of nasals during breastfeeding could explain the typological prevalence of nasal sounds in infancy-related concepts despite their atypical mappings.

5 Scaffolding effects of iconicity on the lexical core of language

Perhaps unsurprisingly, imitative mappings involving either conventionalized onomatopoeia or vocal gestures constituted the most commonly reoccurring type of mapping (Figure 5) in our study. For example, the association between BARK and voiceless sounds does not correspond perfectly to the sound produced by running something over an uneven surface, but it is one of the closest approximations producible by the human vocal apparatus. Since everything we perceive is filtered in some sense, there is a lot of room for sensory idiosyncrasies, such as color blindness and synesthesia. Thus, due to sound symbolism's probabilistic rather than deterministic nature (Dingemanse 2018), some degree of phonetic flexibility is required on the level of both individual speakers and languages. Correspondingly, several sound groups were associated with more than one single concept and/or macro-concept, which created unique but not dichotomous combinations of associations. This extensive overlap does not only indicate that sound symbolism can be rather fine-tuned despite its flexibility, but it also alludes to the different grounds responsible for the associations.

But why then is imitative sound symbolism the most common mapping found in basic vocabulary? Concepts of binary semantic relationships, and some other types of oppositional semantic relationships, are the best fit for sound symbolic mappings based on relative sound symbolism, but generally only represent a limited share of typical basic vocabulary. Circumstantial sound symbolic mappings, on the other hand, are based on very salient language-external factors of the surrounding world, and are rare in general. Thus, imitative sound symbolism (48.8% of all mappings, of which 26.4% is

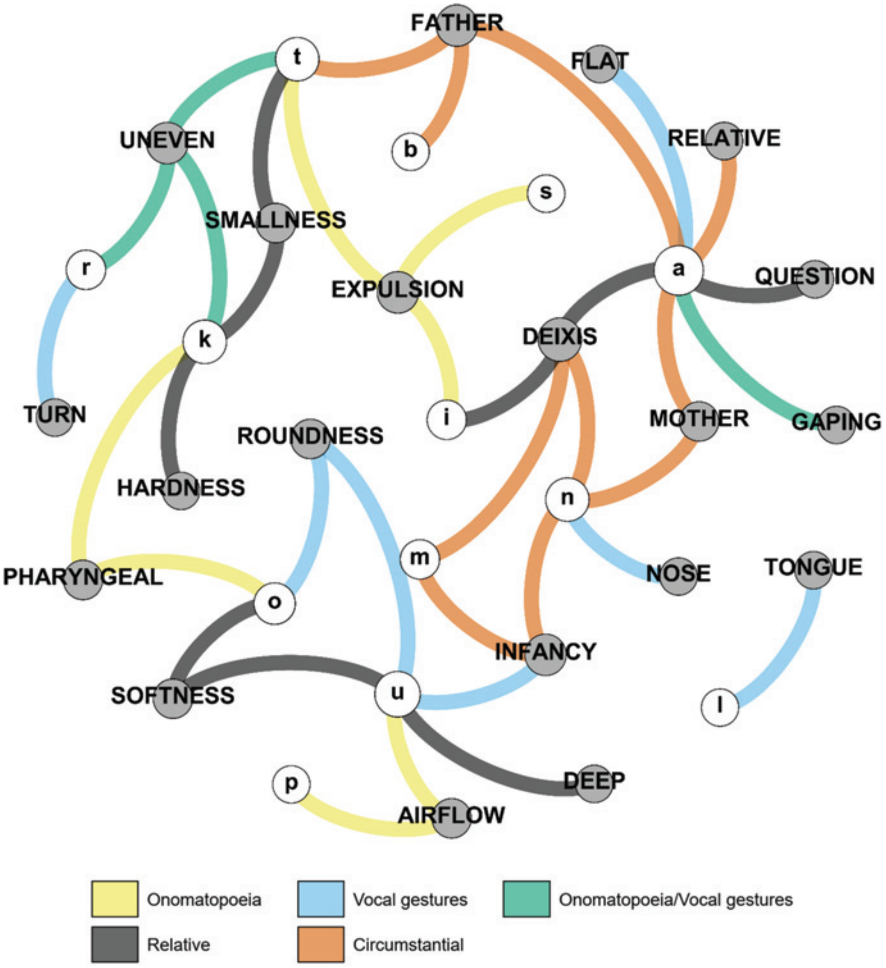


Figure 5: Macro-concepts and the most commonly occurring overrepresented cardinal sound from sound groups with the highest specificity per concept.

onomatopoeia and 22.1% is vocal gestures) may be so common because it is the most accessible type of mapping for basic vocabulary, and arguably also the simplest and most salient one, despite a considerable amount of indirectness (Edmiston et al. 2018).

The high incidence of sound symbolism found in basic vocabulary also brings us back to the lists of words and concepts that are meant to consist of vocabulary items so fundamental that they could be considered universals, and can therefore be used to determine genetic relationships between languages.

Among these, we find the frequently used 100 and 207-item Swadesh lists (Swadesh 1971), shorter adaptations of the Swadesh lists, which have been claimed to have similar or even more accurate lexicostatistical and glottochronologial explanatory power (Starostin 1991; Holman et al. 2008; Pagel et al. 2013), and the Leipzig-Jakarta list based on resistance to lexical borrowing (Haspelmath & Tadmor 2009). The present results showed that, when these lists are combined, at least one sixth of the items can be correlated with the 38 of the 59 sound symbolically affected concepts found in the present study. If semantically related concepts that could cause sound symbolic interference are included as well (e. g. 'rough' could influence words for bark of trees because of bark's often rough surface), this proportion rises to more than one third of all items (Table 6). This could potentially cause subsequent complications for reconstructions of hypothetical long-distance language families, such as Nostratic, as well as e. g. the mostly poorly documented Papuan languages, which are primarily genetically grouped based on their pronominal forms (Ross 2005), of which all were found to be sound symbolic in the current study. Thus, it is necessary to replace these lists by something completely different, amend them by removing the affected item, or, at the very least, use them with extreme caution.

Table 6: The proportion of words included in the most frequently used basic vocabulary lists that may be affected by sound symbolism, with or without semantically related concepts.

Basic Vocabulary Lists	**Items**	**Including semantically related concepts**	
		No	**Yes**
Swadesh-207 (Swadesh 1971)	207	36 (17.4%)	79 (38.2%)
Swadesh-100 (Swadesh 1971)	100	19 (19%)	40 (40%)
Leipzig-Jakarta (Haspelmath & Tadmor 2009)	100	23 (23%)	43 (43%)
(Holman et al. 2008)	40	9 (22.5%)	18 (45%)
Swadesh-Yakhontov (Starostin 1991)	35	8 (22.9%)	17 (48.6%)
(Pagel et al. 2013)	23	11 (47.8%)	14 (60.9%)
Combined	**224**	**38 (16.1%)**	**85 (38%)**

This, in turn, raises the question of why sound symbolism is rather common to begin with. A number of explanations have been proposed over the years, including the hypothesis that sound-meaning associations are vestiges of macro-families or a global proto-language (Ruhlen 1994; Pagel et al. 2013;

Imai & Kita 2014), or that much of sound symbolism can be attributed to analogically motivated patterns (Haspelmath 2008). Diachronic evidence for the decay and reemergence (Johansson & Carling 2015; Flaksman 2017) and the cross-linguistic prevalence of sound symbolism, however, disprove these claims. It is, however, likely that semantically related meanings, including those featured in the present study, adhere to universal patterns of co-lexification (List et al. 2014). In addition, several related meanings also tend to have the same etymological source (Urban 2011, Urban 2012), e. g. 'small' and 'short', or 'nipple', 'breast' and 'milk'. It is also possible that only a small number of stronger sound symbolic patterns could result in the extensive array of sound-meaning associations that we discovered (Westbury et al. 2018). This could explain why some meanings have similar sound distributions, but not why the sound symbolic associations are there to begin with.

However, it should also be mentioned that a fair share of languages probably have not derived their semantically related meanings from the same source. For example, 'nipple' could be derived from 'breast' in some languages based on the meanings' functional and locational similarities, but it could be derived from 'eye' in other languages based on similarities in shape. Additionally, even if all languages used the same patterns of derivation, all individual concepts from a range of sampled languages seem to have kept the same overrepresentations of specific sounds despite inevitable sound change over time.

Thus, we turn our eyes towards the range of functional and communicative benefits of sound symbolism and iconicity (Tamariz et al. 2018). It has been shown that iconic words are easier to learn (Walker et al. 2010; Imai & Kita 2014; Massaro & Perlman 2017), which also applies to iconic nonsense words (Lupyan & Casasanto 2015). For example, English- and Dutch-speaking children are able to correctly generalize the meaning of unknown Japanese ideophones (Imai et al. 2008; Kantartzis et al. 2011; Lockwood et al. 2016a, Lockwood et al. 2016b). Iconic gestures used together with speech can enhance comprehension (Holler et al. 2009; Kelly et al. 2010). Signed languages are heavily iconic (Perniss et al. 2010), and more iconic signs in British Sign Language are recognized more quickly (Thompson et al. 2012; Vinson et al. 2015). Furthermore, people with impairments affecting language proficiency seem to have difficulties with establishing iconic patterns, as illustrated by the observation that subjects with autism spectrum disorders (ASD) correctly map sounds to shapes in the *bouba-kiki task* only around 56% of the time (Oberman & Ramachandran 2008) and dyslexic subject score at around 60% of the time (Drijvers et al. 2015), as compared to an accuracy of 90% among non-ASD subjects (Ramachandran & Hubbard 2001). Iconicity, thus, seems to have a scaffolding or bootstrapping effect on language and language learning, as well as on the grounding of language in sensory and motor systems as described by Perniss and Vigliocco (2014), albeit with some caveats. However, as also pointed out by Perniss and Vigliocco (2014) and Dingemanse et al. (2015), arbitrariness should not be completely disregarded as it has important communicative functions as well: a completely arbitrary language would be difficult to learn, a completely systematic language would limit expressive freedom, and a completely iconic language would be too constrained to cope with all our communicative needs. Hence, a mix of form-to-meaning correspondences all bring something to the table in terms of learning and communication. Furthermore, iconicity is more common early in language acquisition and gradually diminishes (Massaro & Perlman 2017; Perry et al. 2017). Thus, the share of basic vocabulary in the total vocabulary shrinks with age and language proficiency, along with the amount, and arguably the overall effect, of sound symbolism and iconicity. However, iconicity and sound symbolism in core lexicon remain prevalent and still play a crucial role in adulthood and in language as a whole.

6 Concluding remarks

We have shown that sound symbolism is an influential force in language, reaching beyond what are typically proposed as lexical universals.

(a) What is the cross-linguistic extent of sound symbolism in basic vocabulary? By amending previous shortcomings, such as a limited range of investigated concepts, inappropriately designed phonetic classifications and potential genetic and areal influences, the present study shows that even a conservative estimate provides a list of 125 associations between sounds and meanings spanning 59 concepts. While it was expected that onomatopoetic concepts, such as BLOW, and kinship concepts like MOTHER would be strongly affected by sound symbolism, a large number of other associations were found to be equally robust. We proposed that placing focus on correlations between semantic and phonetic features, rather than on specific words and phonemes, is a more appropriate way of investigating sound symbolism's universal, yet flexible structure. This further opened the path to establishing 20 macro-concepts, which were often more general in meaning than the investigated concepts, but had more explanatory power. The structure of the mappings varied considerably, and associations between different combinations of sound groups were found to play a key role for many of them. For example, rounded vowels were associated with ROUNDNESS but also with AIRFLOW when combined with labials. In addition, defining sound symbolic macro-concepts might be one way of identifying the first lexicalized semantic domains that were present at the dawn of human language. These broad lexical fields could then have expanded in different directions semantically through derivation, since there has to be a cognitive base for the saliency of the co-occurring features.

(b) Which types of sound symbolism can be distinguished? If our results are combined with previous research, three main types of sound symbolic mapping can be identified – imitative, diagrammatic, and associative – of which imitative was found to be the most common variety. These main types can be further divided into subgroups, which include previously well-described types, such as onomatopoeia and relative sound symbolism, but also two new types based on imitation. The first type, vocal gestures, mapped meaning to articulatory gestures rather than the accompanying sounds. The second type, circumstantial sound symbolism, grounded mappings through intense co-occurrence between sound and meaning under very specific circumstances such as breastfeeding.

(c) What does sound symbolism reveal about fundamental categories of human cognition? The results further made it clear that distinct types of sound symbolism are often accompanied by mappings of different types, which must be kept in mind when investigating and evaluating cognitive biases, as well as when studying strategies used for acquiring language. This means that, despite the dynamic nature of human language that spawns rich linguistic variation, sound-meaning mappings have proven to be a crucial and substantial part of our most fundamental communicative elements.

Abbreviations

[alv]	alveolar
[alv+v]	voiced alveolar
[alv–v]	voiceless alveolar
[back]	back vowel
[central]	central vowel
[cont]	continuant
[cont+v]	voiced continuant
[cont –v]	voiceless continuant
[front]	front vowel
[glot]	glottal
[glot+v]	voiced glottal
[glot–v]	voiceless glottal
[high]	high vowel
[high-back, +r]	high back rounded vowel

[high-back, –r]	high back unrounded vowel
[high-back]	high back vowel
[high-front, +r]	high front rounded vowel
[high-front, –r]	high front unrounded vowel
[high-front]	high front vowel
[lab]	labial
[lab+v]	voiced labial
[lab–v]	voiceless labial
[lat]	lateral
[lat+v]	voiced lateral
[lat–v]	voiceless lateral
[low]	low vowel
[low-back, +r]	low back rounded vowel
[low-back, –r]	low back unrounded vowel
[low-back]	low back vowel
[low-front, +r]	low front rounded vowel
[low-front, –r]	low front unrounded vowel
[low-front]	low front vowel
[mid]	mid vowel
[nas]	nasal
[nas+v]	voiced nasal
[nas–v]	voiceless nasal
[pal]	palatal
[pal+v]	voiced palatal
[pal–v]	voiceless palatal
[–round]	unrounded vowel
[+round]	rounded vowel
[stop]	stop
[stop+v]	voiced stop
[stop–v]	voiceless stop
[vel]	velar
[vel+v]	voiced velar
[vel–v]	voiceless velar
[vib+v]	voiced vibrant
[vib –v]	voiceless vibrant
[vib]	vibrant
[–voice]	voiceless consonant
[+voice]	voiced consonant
D_MS	daughter (female speaking)
D_MS	daughter (male speaking)
DD_FS	daughter’s daughter (female speaking)
DD_MS	daughter’s daughter (male speaking)
DS_FS	daughter’s son (female speaking)
DS_MS	daughter’s son (male speaking)
F_FS	father (female speaking)
F_MS	father (male speaking)
FF_FS	father’s father (female speaking)

FF_ms	father’s father (male speaking)
FM_fs	father’s mother (female speaking)
FM_ms	father’s mother (male speaking)
FoB_fs	father’s older brother
FoB_ms	father’s older brother
FoZ_fs	father’s older sister
FoZ_ms	father’s older sister
FyB_fs	father’s younger brother
FyB_ms	father’s younger brother
FyZ_fs	father’s younger sister
FyZ_ms	father’s younger sister
M_fs	mother (female speaking)
M_ms	mother (male speaking)
MF_fs	mother’s father (female speaking)
MF_ms	mother’s father (male speaking)
MM_fs	mother’s mother (female speaking)
MM_ms	mother’s mother (male speaking)
MoB_fs	mother’s older brother
MoB_ms	mother’s older brother
MoZ_fs	mother’s older sister
MoZ_ms	mother’s older sister
MyB_fs	mother’s younger brother
MyB_ms	mother’s younger brother
MyZ_fs	mother’s younger sister
MyZ_ms	mother’s younger sister
oB_fs	older brother (female speaking)
oB_ms	older brother (male speaking)
oBD_fs	older brother’s daughter (female speaking)
oBD_ms	older brother’s daughter (male speaking)
oBS_fs	older brother’s son (female speaking)
oBS_ms	older brother’s son (male speaking)
oZ_fs	older sister (female speaking)
oZ_ms	older sister (male speaking)
oZD_fs	older sister’s daughter (female speaking)
oZD_ms	older sister’s daughter (male speaking)
oZS_fs	older sister’s son (female speaking)
oZS_ms	older sister’s son (male speaking)
S_fs	son (female speaking)
S_ms	son (male speaking)
SD_fs	son’s daughter (female speaking)
SD_ms	son’s daughter (male speaking)
SS_fs	son’s son (female speaking)
SS_ms	son’s son (male speaking)
yB_fs	younger brother (female speaking)
yB_ms	younger brother (male speaking)
yBD_fs	younger brother’s daughter (female speaking)
yBD_ms	younger brother’s daughter (male speaking)

yBS_fs	younger brother's son (female speaking)
yBS_ms	younger brother's son (male speaking)
yZ_fs	younger sister (female speaking)
yZ_ms	younger sister (male speaking)
yZD_fs	younger sister's daughter (female speaking)
yZD_ms	younger sister's daughter (male speaking)
yZS_fs	younger sister's son (female speaking)
yZS_ms	younger sister's son (male speaking)

Acknowledgements: A special thanks to Šárka Erbenová for her enduring support. We would also like to thank the editor and reviewers for their comments to the manuscript at several stages.

References

Abelin, Åsa. 1999. *Analyzability and semantic associations in referring expressions: A study in comparative lexicology*. Gothenburg: University of Gothenburg dissertation.

Ahlner, Felix & Jordan. Zlatev. 2010. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign System Studies* 38(1/4). 298–348.

Akita, Kimi 2009. *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of Japanese mimetics*. Kobe: Kobe University dissertation.

Akita, Kimi. 2012. Toward a frame-semantic definition of sound-symbolic words: A collocational analysis of Japanese mimetics. *Cognitive Linguistics* 23(1). 67–90.

Alderete, John & Alexei Kochetov. 2017. Integrating sound symbolism with core grammar: The case of expressive palatalization. *Language* 93(4). 731–766.

Andersen, Elaine S. 1978. Lexical universals of body-part terminology. In Joseph H. Greenberg (ed.), *Universals of human language*, 335–368. Stanford: Stanford University Press.

Bancel, Pierre J. & Alain Matthey de l'Etang. 2013. Brave new words. In Claire Lefebvre, Bernard Comrie & Henri Cohen (eds.), *New perspectives on the origins of language, vol. 144*, 333–377. Amsterdam & Philadelphia: John Benjamins Publishing.

Beecher, Michael. D. 2017. Birdsong learning as a social process. *Animal Behaviour* 124. 233–246.

Berlin, Brent & Paul Kay. 1969. *Basic color terms: Their universality and evolution*. Berkeley & Los Angeles: University of California Press.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler & Morten H. Christiansen. 2016. Sound–meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences* 113(39). 10818–10823.

Bolinger, Dwight L. 1950. Rime, assonance and morpheme analysis. *Word* 6. 117–136.

Bruckert, Laetitia, Jean-Sylvain Liénard, André Lacroix, Michel Kreutzer & Gérard Leboucher. 2006. Women use voice parameters to assess men's characteristics. *Proceedings of the Royal Society of London B: Biological Sciences* 273(1582). 83–89.

Buerkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80. 1–28.

Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. [Linguistics Theory: Representation function of Language]. Jena: Fischer.

Burenhult, Niclas & Stephen C. Levinson. 2008. Language and landscape: A cross-linguistic perspective. *Language Sciences* 30(2). 135–150.

Campbell, Lyle & William J. Poser. 2008. *Language classification: History and method*. Cambridge: Cambridge University Press.

Carling, Gerd & Niklas Johansson. 2014. Motivated language change: Processes involved in the growth and conventionalization of onomatopoeia and sound symbolism. *Acta Linguistica Hafniensia* 46(2). 199–217.

Chastaing, M. 1966. Si les *r* étaient des *l*. *Vie Et Langage* 173. 468–472; 174. 502–507.

Cho, Taehong & Peter. Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2). 207–229.

Clark, Andy. 2006. Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences* 10(8). 370–374.

Collins, Sarah A. 2000. Men's voices and women's choices. *Animal Behaviour* 60(6). 773–780.

Comrie, Bernard. 2013. 131 Numeral bases. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world Atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, Greville, G. 2013. 30 Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Cuskley, Christine, Julia Simner & Simon. Kirby. 2015. Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research* 81(1). 119–130.

de l'Etang, Alain Matthey & Pierre J. Bancel. 2008. The age of Mama and Papa. In John D. Bengtson (ed.), *In hot pursuit of language in prehistory: Essays in the four fields of anthropology. In honor of Harold Crane Fleming*, 417–438. Amsterdam/Philadelphia: John Benjamins Publishing.

de Vignemont, Frédérique, Asifa Majid, Corinne Jola & Patrick. Haggard. 2009. Segmenting the body into parts: Evidence from biases in tactile perception. *The Quarterly Journal of Experimental Psychology* 62(3). 500–512.

de Villiers, Jill G. & Peter A. de Villiers. 1978. *Language acquisition*. Cambridge: Harvard University Press.

Diessel, Holger. 2014. Demonstratives, frames of reference, and semantic universals of space. *Language and Linguistics Compass* 8(3). 116–132.

Diffloth, Gérald. 1994. i: big, a: small. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 107–114. Cambridge: Cambridge University Press.

Dingemanse, M. 2018. Redrawing the margins of language: Lessons from research on ideophones. *Glossa: A Journal of General Linguistics* 3(1). 1–30. http://doi.org/10.5334/gjgl.444 (accessed 2 April 2018).

Dingemanse, Mark. 2011. Ezra pound among the Mawu. In Pascal Michelucci, Olga Fischer & Christina Ljungberg (eds.), *Semblance and signification. Iconicity in language and literature 10*, 39–54. Amsterdam: John Benjamins.

Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6. 654–672.

Dingemanse, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF – Language Typology and Universals* 70(2). 363–384.

Dingemanse, Mark & Kimi. Akita. 2016. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 501–532.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, iconicity and systematicity in language. *Trends in Cognitive Sciences* 19(10). 603–615.

Dingemanse, Mark, Francisco Torreira & Nick J. Enfield. 2013. Is "Huh?" a universal word? Conversational infrastructure and the convergent evolution of linguistic items. *PloS One* 8(11). 10.1371/journal.pone.0078273 (accessed 23 August 2017).

Dixon, Robert M. W. 1982. *Where have all the adjectives gone? And other essays in semantics and syntax*. Amsterdam: De Gruyter Mouton.

Dolscheid, Sara, Sabine Hunnius, Daniel Casasanto & Asifa Majid. 2012. The sound of thickness: Prelinguistic infants' associations of space and pitch. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*. 306–311.

Drijvers, Linda, Lorijn S. Zaadnoordijk & Mark Dingemanse. 2015. Sound-symbolism is disrupted in dyslexia: Implications for the role of cross-modal abstraction processes. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. 602–607.

Edmiston, Pierce, Marcus Perlman & Gary Lupyan. 2018. Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences* 285(1874). 20172709. 10.1098/rspb.2017.2709 (accessed 13 April 2018).

Enfield, Nick J., Asifa Majid & Miriam van Staden. 2006. Cross-linguistic categorisation of the body: Introduction. *Language Sciences* 28(2). 137–147.

Erickson, Robert P. 2008. A study of the science of taste: On the origins and influence of the core ideas. *Behavioral and Brain Sciences* 31(1). 59–75.

Fay, Nicolas, Michael Arbib & Simon. Garrod. 2013. How to bootstrap a human communication system. *Cognitive Science* 37. 1356–1367.

Flaksman, Maria. 2017. Iconic treadmill hypothesis. In Matthias Bauer, Angelika Zirker, Olga Fischer & Christina Ljungberg (eds.), *Dimensions of Iconicity. Iconicity in Language and Literature 15*, 15–38. Amsterdam: John Benjamins.

Fónagy, Ivan. 1963. *Die Metaphern in der Phonetik: Ein Beitrag zur Entwicklungsgeschichte des wissenschaftlichen Denkens*. [The metaphors in phonetics: a contribution to the developmental history of scientific thought]. The Hague: Mouton.

Fox, Robert Allen. 1982. Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica* 39(1). 1–22.

Fromkin, Victoria, Stephen Krashen, Susan Curtiss, David Rigler & Marilyn Rigler. 1974. The development of language in genie: A case of language acquisition beyond the "critical period". *Brain and Language* 1(1). 81–107.

Gibson, James J. 1977. The theory of affordances. In Robert E. Shaw & John Bransford (eds.), *Perceiving, acting, and knowing*, 67–82. Hillsdale NJ: Lawrence Erlbaum Associates.

Goddard, Cliff. 2001. Lexico-semantic universals. *Linguistic Typology* 5(1). 1–65.

Goddard, Cliff & Anna Wierzbicka (eds.). 2002. *Meaning and universal grammar: Theory and empirical findings* 2 volumes. Amsterdam & Philadelphia: John Benjamins.

Greenhill, Simon J. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37. 689–698.

Hamilton-Fletcher, Giles, Christoph Witzel, David Reby & Jamie Ward. 2017. Sound properties associated with equiluminant colours. *Multisensory Research* 30(3–5). 337–362.

Hammarström, Harald, Robert Forkel & Martin. Haspelmath. 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. http://glottolog.org (accessed 15 January 2017).

Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33.

Haspelmath, Martin & Uri Tadmor (eds.). 2009. *Loanwords in the world's languages: A comparative handbook*. Berlin and New York: De Gruyter Mouton.

Hinton, Leanne, Johanna Nichols & John J. Ohala. 1994. Introduction: Sound-symbolic processes. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 325–347. Cambridge: Cambridge University Press.

Holler, Judith, Heather Shovelton & Geoffrey Beattie. 2009. Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior* 33(2). 73–88.

Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller & Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica 42*(3–4. 331–354.

Humboldt, Wilhelm V. 1838. *Über die Kawi-Sprache auf der Insel Java: Nebst einer Einleitung über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts*. [On the Kawi language on the island of Java: In addition to an introduction to the diversity of human language and its influence on the spiritual development of the human race]. Berlin: Königlichen Akademie der Wissenschaften zu Berlin.

Ibarretxe-Antuñano, Iraide. 2006. Estudio lexicológico de las onomatopeyas vascas: El Euskal Onomatopeien Hiztegia: Euskara-Ingelesera-Gaztelania [A lexicological study of Basque onomatopoeia]. *Fontes Linguae Vasconum* 101. 145–159.

Ibarretxe-Antuñano, Iraide. 2017. Basque ideophones from a typological perspective. *Canadian Journal of Linguistics/Revue Canadienne De Linguistique* 62(2). 196–220.

Imai, Mutsumi & Sotaro Kita. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B* 369(1651). 10.1098/rstb.2013.0298 (accessed 23 August 2017).

Imai, Mutsumi, Sotaro Kita, Miho Nagumo & Hiroyuki. Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109(1). 54–65.

Iwasaki, Noriko, David P. Vinson & 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. 53–78.

Jack, Rachael E., Oliver G. Garrod & Philippe G. Schyns. 2014. Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology* 24(2). 187–192.

Jakobson, Roman. 1962. Why 'mama' and 'papa'? In Roman Jakobson (ed.), *Selected writings, Vol. I: Phonological studies*, 538–545. The Hague: De Gruyter Mouton.

Jakobson, Roman, C. Gunnar Fant & Morris Halle. 1951. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, Mass.: MIT Press.

Jespersen, Otto. 1922. *Language: Its nature, development and origin*. London: Allen & Unwin.

Johansson, Niklas. 2017. Tracking linguistic primitives: The phonosemantic realization of fundamental oppositional pairs. In Matthias Bauer, Angelika Zirker, Olga Fischer & Christina Ljungberg (eds.), *Dimensions of iconicity. Iconicity in language and literature 15*, 39–62. Amsterdam: John Benjamins.

Johansson, Niklas & Gerd Carling. 2015. The de-iconization and rebuilding of iconicity in spatial deixis: An Indo-European case study. *Acta Linguistica Hafniensia* 47(1). 4–32.

Johansson, Niklas & Jordan Zlatev. 2013. Motivations for sound symbolism in spatial deixis: A typological study of 101 languages. *Public Journal of Semiotics Online* 5(1). 3–20.

Jones, John Matthew, David Vinson, Nourane Clostre, Alex Lau Zhu, Julio Santiago & Gabriella Vigliocco. 2014. The bouba effect: Sound-shape iconicity in iterated and implicit learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2459–2464.

Kantartzis, Katerina, Mutsumi Imai & Sotaro Kita. 2011. Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science* 35(3). 575–586.

Kay, Paul & Luisa Maffi. 2013. 133 Number of basic colour categories. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Kelly, Spencer D., Aslı Özyürek & Eric Maris. 2010. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2). 260–267.

Kemp, Charles & Terry. Regier. 2012. Kinship categories across languages reflect general communicative principles. *Science* 336(6084). 1049–1054.

Khetarpal, Naveen, Asifa Majid, Barabara Malt, Steven Sloman & Terry Regier. 2010. Similarity judgments reflect both language and cross-language tendencies: Evidence from two semantic domains. *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*. 358–363.

Khetarpal, Naveen, Grace Neveu, Asifa Majid, Lev Michael & Terry Regier. 2013. Spatial terms across languages support near-optimal communication: Evidence from Peruvian Amazonia, and computational analyses. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 764–769.

Kibrik, Andrej. 2012. Toward a typology of verbal lexical systems: A case study in Northern Athabaskan. *Linguistics* 50(3). 495–532.

Kita, Sotaro, Katerina Kantartzis & Mutsumi Imai. 2010. Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage. In Marieke Schouwstra, Bart de Boer & Andrew D. M. Smith (eds.), *The Evolution of Language – Proceedings of the 8th International Conference (Evolang8)*, 206–213. Singapore: World Scientific.

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

Koptjevskaja-Tamm, Maria. 2008. Approaching lexical typology. In Martine Vanhove (ed.), *From polysemy to semantic change: A typology of lexical semantic associations*, 3–52. Amsterdam: John Benjamins.

Kroonen, Guus. 2010. *Etymological Dictionary of Proto-Germanic. "Grōni-"*. Leiden: Brill. http://dictionaries.brillonline.com (accessed 20 October 2017).

Kruschke, John K. & Torrin M. Liddell. 2018. The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review* 25(1). 178–206.

Ladefoged, Peter. 2001. *Vowels and consonants: An introduction to the sounds of languages*. Malden, MA: Blackwell Publishing.

Ladefoged, Peter & Ian. Maddieson. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

LaPolla, Randy. 1994. An experimental investigation into phonetic symbolism as it relates to Mandarin Chinese. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 130–147. Cambridge: Cambridge University Press.

Levinson, Stephen C. & Asifa Majid. 2014. Differential ineffability and the senses. *Mind & Language* 29(4). 407–427.

Levinson, Stephen C. & Sérgio Meira. 2003. 'Natural concepts' in the spatial topologial domain–adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 79(3). 485–516.

Lindblad, Per. 1998. *Talets akustik och perception*. [The acoustics and perception of speech]. Gothenburg: University of Gothenburg.

List, Johann-Mattis, Thomas Mayer, Anselm Terhalle & Matthias Urban. 2014. *CLICS: Database of cross-linguistic colexifications*. Marburg: Forschungszentrum Deutscher Sprachatlas (Version 1.0, online). http://CLICS.lingpy.org (accessed 3 December 2017).

Lockwood, Gwilym, Mark Dingemanse & Peter Hagoort. 2016a. Sound-symbolism boosts novel word learning. *Journal of experimental psychology. Learning, Memory, and Cognition* 42(8). 1274–1281.

Lockwood, Gwilym, Peter Hagoort & Mark Dingemanse. 2016b. How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra* 2(1). 10.1525/collabra.42 (accessed 2 April 2018).

Ludwig, Vera U. & Julia Simner. 2013. What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex* 49(4). 1089–1099.

Lupyan, Gary & Daniel Casasanto. 2015. Meaningless words promote meaningful categorization. *Language and Cognition* 7(2). 167–193.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Majid, Asifa & Stephen C. Levinson. 2008. Language does provide support for basic tastes. *Behavioral and Brain Sciences* 31. 86–87.

Massaro, Dominic W. & Marcus Perlman. 2017. Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication* 2(4). 10.3389/fcomm.2017.00004 (accessed 2 April 2018).

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

Mielke, Jeff. 2012. A phonetically based metric of sound similarity. *Lingua* 122(2). 145–163.

Moran, Steven, Daniel McCloy & Richard Wright (eds.). 2014. *PHOIBLE online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://phoible.org (accessed 29 April 2017).

Newman, Stanley S. 1933. Further experiments in phonetic symbolism. *The American Journal of Psychology* 45(1). 53–75.

Nichols, J. 1999. Why 'me' and 'thee'? In Laurel J. Brinton (ed.), *Historical linguistics 1999: Selected papers from the 14th International Conference on Historical Linguistics, Vancouver, 9–13 August 1999*, 253–276. Amsterdam & Philadelphia: John Benjamins Publishing.

Nielsen, Alan K. & Drew Rendall. 2013. Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale* 67(2). 153–163.

Oberman, Lindsay M. & Vilayanur S. Ramachandran. 2008. Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis. *Social Neuroscience* 3(3–4). 348–355.

Ohala, John J. 1994. The frequency codes underlies the sound symbolic use of voice pitch. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 325–347. Cambridge: Cambridge University Press.

Pagel, Mark, Quentin D. Atkinson, Andreea S. Calude & Andrew Meade. 2013. Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences* 110(21). 8471–8476.

Paradis, Carita, Caroline Willners & Steven Jones. 2009. Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon* 4(3). 380–429.

Pedersen, Alyssa & Michelle L. Tomaszycki. 2012. Oxytocin antagonist treatments alter the formation of pair relationships in zebra finches of both sexes. *Hormones and Behavior* 62(2). 113–119.

Penfield, Wilder & Edwin. Boldrey. 1937. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain* 60(4). 389–443.

Penfield, Wilder & Theodore Rasmussen. 1950. *The cerebral cortex of man*. New York: Maxmillan.

Perlman, Marcus & Ashley A. Cain. 2014. Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* 14. 321–351.

Perlman, Marcus, Rick Dale & Gary Lupyan. 2015. Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8). 150152. 10.1098/rsos.150152 (accessed 13 April 2018).

Perniss, Pamela, Robin L. Thompson & Gabriella Vigliocco. 2010. Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* 1(227). 1–15.

Perniss, Pamela & Gabriella. Vigliocco. 2014. The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B* 369(1651). 20130300. 10.1098/rstb.2013.0300 (accessed 21 September 2018).

Perry, Lynn K., Marcus Perlman, Bodo Winter, Dominic W. Massaro & Gary Lupyan. 2017. Iconicity in the speech of children and adults. *Developmental Science* 21(3). 10.1111/desc.12572 (accessed 13 April 2018).

Pierce, Charles Sanders. 1931–1958. *The collected papers of Charles Sanders Peirce*, 1–8. Cambridge: Cambridge University Press.

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

Roque, Lila San, Kendrick H. Kobin, Elisabeth Norcliffe, Penelope Brown, Rebecca Defina, Mark Dingemanse, Tyko Dirksmeyer, Nick J. Enfield, Simeon Floyd, Jeremy Hammond, Giovanni Rossi, Sylvia Tufvesson, Saskia van Putten & Asifa Majid. 2015. Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics* 26(1). 31–60.

Ross, Malcolm. 2005. Pronouns as a preliminary diagnostic for grouping Papuan languages. In Andrew Pawley, Robert Attenborough, Jack Golson & Robin Hide (eds.), *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15–66. Canberra: Pacific Linguistics.

Ruhlen, Merritt. 1994. *On the origin of languages: Studies in linguistic taxonomy*. Stanford: Stanford University Press.

Sander, Eric K. 1972. When are speech sounds learned? *Journal of Speech and Hearing Disorders* 37(1). 55–63.

Sapir, Edward. 1929. A study in phonetic symbolism. *Journal of Experimental Psychology* 12(3). 225–239.

Saussure, Ferdinand. 1983[1916]. *Course in general linguistics*. Duckworth: London.

Sedley, David. 2003. *Plato's Cratylus*. Cambridge: Cambridge University Press.

Sell, Aaron, Gregory A. Bryant, Leda Cosmides, John Tooby, Daniel Sznycer, Christopher Von Rueden, Andre Krauss & Michael Gurven. 2010. Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society of London B: Biological Sciences* 277(1699). 3509–3518. 10.1098/rspb.2010.0769 (accessed 09 October 2018).

Sereno, Joan A. 1994. Phonosyntactics. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 263–275. Cambridge: Cambridge University Press.

Sidhu, David M. & Penny M. Pexman. 2015. What's in a name? Sound symbolism and gender in first names. *PloS One* 10(5). e0126809.

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

Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the world, twentieth edition*. Dallas, Texas: SIL International. https://www.ethnologue.com (accessed 4 March 2016).

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

Starostin, Sergei. 1991. *Altajskaja Problema i Proisxozhdenie Japonskogo Jazyka [The Altaic problem and the origin of the Japanese language]*. Moscow: Nauka.

Stevens, Kenneth N. 1998. *Acoustic phonetics*. Cambridge, MA: MIT Press.

Swadesh, Morris. 1971. *The origin and diversification of language*. Edited post mortem by Joel Sherzer. London: Transaction Publishers.

Taitz, Alan, Assaneo, M. Florencia, Natalia Elisei, Mónica Trípodi, Laurent Cohen, Jacobo D. Sitt & Marcos A. Trevisan. 2018. The audiovisual structure of onomatopoeias: An intrusion of real-world physics in lexical creation. *PloS One* 13(3). e0193466. 10.1371/journal.pone.0193466 (accessed 14 April 2018).

Tamariz, Mónica, Seán G. Roberts, Martínez, J. Isidro & Julio Santiago. 2018. The interactive origin of iconicity. *Cognitive Science* 42(1). 334–349.

Taylor, Anna M. & David Reby. 2010. The contribution of source–filter theory to mammal vocal communication research. *Journal of Zoology* 280(3). 221–236.

Theofanopoulou, Constantina, Cedric Boeckx & Erich D. Jarvis. 2017. A hypothesis on a role of oxytocin in the social mechanisms of speech and vocal learning. *Proceedings of the Royal Society B: Biological Sciences* 284(1861). 20170988. 10.1098/rspb.2017.0988 (accessed 25 October 2017).

Thompson, Robin L., David P. Vinson, Bencie Woll & Gabriella Vigliocco. 2012. The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science* 23(12). 1443–1448.

Traunmüller, Hartmut. 1994. Sound symbolism in deictic words. In Hans Auli & Peter af Trampe (eds.), *In tongues and texts unlimited: Studies in honour of Tore Jansson on the occasion of his sixtieth anniversary*, 213–234. Stockholm: Department of Classical Languages, Stockholm University.

Ultan, Russel. 1978. Size-sound symbolism. In Joseph Greenberg (ed.), *Universals of human language 2, Phonology*, 525–567. Stanford: Stanford University Press.

Urban, Matthias. 2011. Conventional sound symbolism in terms for organs of speech: A cross-linguistic study. *Folia Linguistica* 45(1). 199–214.

Urban, Matthias. 2012. *Analyzability and semantic associations in referring expressions: A study in comparative lexicology*. Leiden: Leiden University dissertation.

Viberg, Åke. 1983. The verbs of perception: A typological study. *Linguistics* 21. 123–162.

Viberg, Åke. 2001. Verbs of perception. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 1294–1309. Berlin and New York: Walter de Gruyter.

Vinson, David, Robin L. Thompson, Robert Skinner & Gabriella Vigliocco. 2015. A faster path between meaning and form? Iconicity facilitates sign recognition and production in British Sign Language. *Journal of Memory and Language* 82. 56–85.

Walker, Peter. 2012. Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception & Psychophysics* 74(8). 1792–1809.

Walker, Peter, Bremner J. Gavin, Uschi Mason, Jo Spring, Karen Mattock, Alan Slater & Scott P. Johnson. 2010. Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science* 21(1). 21–25.

Ward, Jamie, Brett Huckstep & Elias Tsakanikos. 2006. Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* 42(2). 264–280.

Watanbe, Junji, Yuuka Utsunomiya, Hiroya Tsukurimichi & Maki Sakamoto. 2012. Relationship between phonemes and tactile-emotional evaluations in Japanese sound symbolic words. *Proceedings of the Annual Meeting of the Cognitive Science Society* 34(34). 2517–2522.

Westbury, Chris. 2005. Implicit sound symbolism in lexical access: Evidence from an interference task. *Brain and Language* 93(1). 10–19.

Westbury, Chris, Geoff Hollis, David M. Sidhu & Penny M. Pexman. 2018. Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language* 99. 122–150.

Wichmann, Søren, Eric W. Holman & Cecil H. Brown. 2010. Sound symbolism in basic vocabulary. *Entropy* 12(4). 844–858.

Woodworth, Nancy L. 1991. Sound symbolism in proximal and distal forms. *Linguistics* 29(2). 273–299.

Ye, Zheng, Arjen Stolk, Ivan Toni & Peter. Hagoort. 2016. Oxytocin modulates semantic integration in speech comprehension. *Journal of Cognitive Neuroscience* 29(2). 267–276.

Zhang, Hong-Feng, Yu-Chuan Dai, Jing Wu, Mei-Xiang Jia, Ji-Shui Zhang, Xiao-Jing Shou, Song-Ping Han, Rong Zhang & Ji-Sheng. Han. 2016. Plasma oxytocin and arginine-vasopressin levels in children with autism spectrum disorder in China: Associations with symptoms. *Neuroscience Bulletin* 32(5). 423–432.

Ziemke, Tom. 2016. The body of knowledge: On the role of the living body in grounding embodied cognition. *BioSystems* 148. 4–11.

Zlatev, Jordan. 2007. Embodiment, language and mimesis. *Body, Language and Mind* 1. 297–337.

Supplementary Material: All appendices, along with additional supplementary materials can be found at http://cogsci.se/publications/2019_johansson.html and https://osf.io/dh7f3/.

Study II

Cultural evolution leads to vocal iconicity in an experimental iterated learning task

Niklas Erben Johansson

Center for Language and Literature, Lund University, Sweden

Jon W. Carr

Cognitive Neuroscience, International School for Advanced Studies, Trieste, Italy

Simon Kirby

Centre for Language Evolution, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Experimental and cross-linguistic studies have shown that words that carry meanings related to SIZE and SHAPE are highly affected by *vocal iconicity*. Although these studies demonstrate the importance of vocal iconicity and reveal the cognitive biases underpinning it, there is less work demonstrating how these biases lead to the evolution of a sound symbolic lexicon in the first place. In this study, we show how words can be shaped by cognitive biases through cultural evolution. Using a simple experimental setup resembling the game *telephone*, we examined how an arbitrary word form changed as it was passed from one participant to the next by a process of *immediate iterated learning*. 1500 naïve participants were divided into five condition groups. The participants in the CONTROL-group received no information about the meaning of the

word they were about to hear, while the participants in the remaining four groups were either informed that the word meant BIG or SMALL (with the meaning being presented in text), or ROUND or POINTY (with the meaning being presented as a picture). The first participant in a transmission chain was presented with a phonetically diverse word and asked to repeat it. Thereafter, the recording of the repeated word was played for the next participant in the same chain. The sounds of the audio recordings were then transcribed and categorized according to six binary sound parameters. By modelling the proportion of vowels or consonants for each sound parameter, the SMALL-condition showed significant increases of FRONT UNROUNDED vowels and the POINTY-condition significant increases of ACUTE consonants. These effects were attributed to cognitive affordability of having only one pole of an oppositional pair iconically charged. The results show that linguistic transmission is sufficient for vocal iconicity to emerge, which demonstrates the role non-arbitrary associations play in the evolution of language.

1. Introduction

Languages have iconic structure, i.e. non-arbitrary associations between sound and meaning, woven into the very core of the lexicon (Dingemanse et al., 2015; Blasi et al., 2016). But how does such patterning enter languages and what explains its apparent universality? In this paper, we use the experimental iterated learning paradigm to show how the cultural transmission of a single artificial word converges on iconic sound-meaning correspondences that closely reflect the kinds of patterns observed in natural languages. Based on evidence from the large body of previous studies on the *bouba-kiki effect*, we predicted that:

a) The meaning SMALL, and possibly also the meaning POINTY, would result in words with a larger share of high pitch sounds than the meanings BIG and ROUND.

b) The meaning ROUND would result in words with a larger share of labial (rounded) sounds than the meaning POINTY.

1.1 Oppositional vocal iconicity

The number of studies on the genetically and areally independent, (near-)universal, non-arbitrary and flexible associations between sounds and meanings has grown considerably in recent decades. This type of association is generally referred to as *vocal iconicity* or *motivated sound symbolism* (Cuskley & Kirby, 2013). Several large cross-linguistic studies (Wichmann et al., 2010; Blasi et al., 2016; Erben Johansson et al. 2020), which in some cases incorporate data from thousands of languages, have found strong phonetic patterns across languages in basic vocabulary items, that is concepts that are supposed to be more or less universal to all speakers of all languages (e.g. *tree*, *you*, *mother*, *eat*, *black*, *small*), both culturally and historically (Swadesh, 1971; Goddard & Wierzbicka, 2002). Most experimental studies on vocal iconicity have, on the other hand, been rather restricted in scope, usually involving two meanings and two phonetic parameters. Sapir's (1929) study of size-based vocal iconicity showed that 80% of almost 500 participants preferred to associate a small table with the phonetic form /mil/

and a large table with the form /mal/. Similarly, Köhler (1929) investigated shape-based vocal iconicity by asking participants to match a round, amoeba-like shape and a pointy, star-like shape with either /takete/ or /baluma/ (later replaced by /maluma/ in his 1947 study). Most of the participants thought that the best fit for the round shape was the word containing voiced sounds and the pointy shape was accordingly paired with the word containing unvoiced sounds. Köhler's (1929) work was later built on by several scholars (e.g. Holland & Wertheimer, 1964; Rogers & Ross, 1975; Boyle & Tarte, 1980; Lindauer, 1990; Bross, 2018; for a review see Lockwood & Dingemanse, 2015), but perhaps most famously by Ramachandran & Hubbard (2001) who, using similar shape stimuli and the phonetic forms /kiki/ and /bouba/, found that more than 95% of participants agreed that /kiki/ should be paired with the pointy shape and /bouba/ with the round shape. Other studies have demonstrated iconic effects in a wide range of semantically-opposite meanings. Newman (1933) found a correspondence between both vowels and consonants in the small-large and bright-dark dimensions. Fónagy (1963) compared /i/ and /u/ in Hungarian and concluded that /i/ was considered quicker, smaller, prettier, friendlier and harder than /u/, while /u/ was perceived as thicker, hollower, darker, sadder, blunter, more bitter, and stronger than /i/ (in both children and adults). Taylor & Taylor (1962) and Taylor (1963) found iconic effects for big-small, active-passive, warm-cold and pleasant-unpleasant in four unrelated languages, and Gebels (1969) found effects in words of a sensory nature by examining 22 pairs of antonyms in five languages.

Perhaps the most widely known type of vocal iconicity is onomatopoeia (i.e. human imitations of real-world sounds with varying similarity to the source sound), which has been referred to as *imagic*, *absolute* or *imitative iconicity* (Hinton et al., 1994; Dingemanse, 2011; Dingemanse et al., 2015; Carling & Johansson, 2015). For example, the English word *cuckoo* is a direct imitation of the calls produced by the cuckoo but produced through the filter of the human vocal apparatus. However, in contrast to onomatopoeia, the type of vocal iconicity usually investigated experimentally involves referents that are based on other senses than hearing, e.g. size, shape, deixis or color, and can in most cases be classified as *relative* or *word-relational diagrammatic iconicity*. Relative iconicity is constructed by mapping semantic contrasts to phonetic contrasts which are somehow similar to each other. This usually includes binary semantic meanings that can easily be placed in opposition to each other (FAST-SLOW, BIG-SMALL, ROUND-POINTY, etc.) and phonetic attributes that can be perceived to belong to a gradable scale (e.g. voicing, quality, quantity, tone, volume, etc.). For example, if SMALL is mapped to high tone and BIG is mapped to low tone, these parallel sound-meaning associations add relations between the semantic and phonetic parameters to the internal relations within the semantic parameter SIZE (between BIG and SMALL) and the phonetic parameter tone (between high and low tone). This type of association also usually involves Ohala's (1994) so-called *frequency code* (see also Rendall et al., 2005), which states that since the size of the resonance chamber of an animal dictates the fundamental frequency of that animal's vocalizations, the sounds that the animal produces can be utilized in various ways to evoke properties such as size. This works according to the same principle as erecting feathers or fur in threatening situations to seem larger or cowering when wanting to submit. Ohala therefore argues that most animals, and maybe specifically humans, perceive low and/or falling fundamental frequencies of vocalizations such as growling as large, authoritative, confident, dominant, or distant, and high and/or rising fundamental frequencies of vocalizations such as whining as small, polite, questioning, dependent, or near.

1.2 The strengths and weaknesses of vocal iconicity experiments

Ahlner & Zlatev (2010) investigated the typical bouba-kiki task in more detail from a cognitive semiotic perspective. First, they selected vowels and consonants that had been reported to contrast iconically; for example, voiceless obstruents and front unrounded vowels (associated with 'hard', 'sharp', 'pointy', 'small') were contrasted with voiced sonorants and back rounded vowels (associated with 'soft', 'smooth', 'heavy', 'round', 'large'). They then created four sets of words by combining sounds from the vowel and consonant groups. Two of these word types were iconically congruent, that is voiceless obstruents were combined with front unrounded vowels (e.g. [titi]) and voiced sonorants with back rounded vowels (e.g. [mumu]). The other two word types were iconically incongruent and combined voiceless obstruents with back rounded vowels (e.g. [tutu]), and voiced sonorants with front unrounded vowels (e.g. [mimi]). Participants were then asked to match these words to a pointy or round shape, virtually identical to those used by Ramachandran & Hubbard (2001). The results showed that participants were significantly more likely to give the "correct" response when the vowel and consonant combinations were congruent with the standard sound-symbolic patterns. In addition, the words in which the consonantal part matched the figures showed a stronger effect compared to the vowels, which might indicate that consonants play a more important role in this iconic mapping.

D'Onofrio (2014) conducted a similar study by constructing words that combined rounded back vowels or unrounded front vowels with voiceless and voiced variants of labial, alveolar and velar stops. The words that included voiced velar consonants and rounded back vowels were the most preferred form for the round shape, with 91% of the participants answering correctly and 93% agreeing that words based on voiceless alveolar consonants and unrounded front vowels were the most fitting form for the pointy shape.

Nielsen & Rendall (2011) investigated the bouba-kiki effect in a series of studies to unravel potential secondary biases in these types of vocal iconicity experiments. The authors excluded certain letters such as *i* and *o* because of a potential orthographic confound (i.e. the letter *i* looks pointy and the letter *o* looks round) and generated both stimuli words and stimuli shapes to prevent lexical and possible visual interference. The stimuli words presented to the participants were congruent on either the consonant or vowel level, and the results showed that the consonant matching scheme yielded correct responses at around 80%, while under the vowel matching scheme there was no significant effect. They then conducted the same experiment using auditory stimuli through a text-to-speech synthesizer to exclude any orthographic bias, and they found the same general pattern regarding consonants and vowels, although the overall effect was weaker than the text stimuli.

Nielsen & Rendall (2012) examined the roles of the underlying biases of the phenomenon by dividing participants into one condition in which they were primed with words paired with congruent figures, and another in which they were primed with incongruent combinations. They were then subjected to 80 trials of random single word-figure combinations and asked to either confirm whether the combinations were "correct" or not. The participants that were taught to combine images with congruent words (pointy images with words containing voiceless plosives and round images with words containing sonorants) performed only modestly (53.3% correct) but above chance level. The participants that were taught to combine images with incongruent words performed at chance level (50.4% correct). This therefore suggests that the iconic bias might be weaker than demonstrated by previous studies and that the forced choice paradigm could inflate weak effects (Dingemanse et al., 2015).

In a second follow-up study, Nielsen & Rendall (2013) specifically investigated the role of consonants and vowels in a similar experimental setup by letting participants themselves select the best fitting word for pointy or round figures. When presented with a figure, they selected one generated syllable presented in both text and auditory form from each of two sets. In both syllables there was a preference for including plosives and unrounded vowels for the spiky figure and sonorants and rounded vowels for the round figure, but only the selection for plosives and unrounded vowels in the first syllable was significantly different from chance.

Despite the large number of studies that have found supporting evidence for the bouba-kiki effect, there are two reported cases where the effect has failed (Rogers & Ross, 1975; Styles & Gawne, 2017), both of which were conducted with participants speaking isolated and somewhat phonologically atypical languages. This raises questions about the strength of the bouba-kiki effect as well as the influence of language-specific phonological makeup and writing systems. In addition, Cuskley et al. (2017) found that orthography seems to be a major confounding factor for associations between sounds and shapes, which, they argue, had not been sufficiently controlled for in most previous studies. By both testing how well literate participants matched abstract shapes to non-words in written form along with spoken representations, and how well they matched the shapes to purely auditory non-words, Cuskley et al. showed that the curvature of letters can significantly influence the perceived roundedness of shapes in sound-shape associations. However, Hamilton-Fletcher et al. (2018) showed that these types of correspondences might be more complex. While pitch-shape correspondences required visual experience to emerge in the blind participants, pitch-size and pitch-weight were found to be unaffected by visual experience, and pitch-texture and pitch-softness even seemed to emerge or grow stronger with blindness. Thus, visual experience cannot solely explain why people with limited multisensory interactions have multimodal perception. Instead, this could be attributed to other factors such as neuroplasticity.

1.3 Vocal iconicity through iterated learning

Some general conclusions can be drawn from the different approaches that the studies we have reviewed have employed. In the bouba-kiki effect, both vowels and consonants seem to play a role, which illustrates the value of thoroughly investigating how different sounds are mapped to different meanings. Furthermore, previous studies, with some notable exceptions (e.g. Jones et al., 2014 and Tamariz et al., 2017, described below), have typically relied on experimental paradigms in which participants are asked to associate meanings with a set of words or syllables that are predefined. This means that while the bouba-kiki effect seems to be more or less universal, it is also subjective in nature, given that each individual participant is asked to combine meanings with sounds that may or may not adequately fit his or her intuition or phonology. We therefore wanted to investigate the cognitive biases that lie at the core of vocal iconicity by using a methodological approach that focuses on the transmission of vocal iconicity through the language filters of participants with a wide range of native languages, but which also excludes orthographic influence as much as possible. This approach would then allow us to get a more holistic picture of the bouba-kiki effect by revealing differences to the results of previous studies.

One way of achieving this is to use methods that are specifically designed to study how languages change over time, such as the *iterated learning paradigm* (Kirby, 2001; Kirby & Hurford, 2002; Kirby et al., 2008; Kirby et al, 2015). In iterated learning studies, some form of information, such as words, music or drawings, is transmitted from one participant to another, with the learner at generation *i* producing behavior that is input to the learner at generation *i*+1. Together, several generations of such learners form a "transmission chain". At its core, the iterated learning paradigm is reliant on the fact that information tends to be lost during the transmission process (Spike et al., 2017), causing the object of study to change in ways that reflect the learner's cognitive biases, whatever those biases happen to be, and the dynamics involved in the particular transmission channel used. For example, Canini, Griffiths, Vanpaemel & Kalish (2014) have shown how category learning biases can emerge naturally through an iterated learning study. In this way, iterated learning experiments can be used as a technique to uncover the cognitive biases of participants, acting as a complement to more targeted experimental designs which start out with specific hypotheses about what these biases might be.

However, to date, only a few studies have investigated the emergence of vocal iconicity through iterated learning. Jones et al. (2014) trained participants on miniature languages that consisted of parings between various round and pointy shapes and written labels which were rated as sound iconically neutral by English monolinguals. The participants then had to type the label learnt for each shape, including shapes they had not previously been trained on, and these labels were passed on to the next participant. Jones et al. found that iconic labels emerged to express round shapes but not pointy ones. When the participants then had to match labels that were judged as either iconically round, pointy or neutral to one of two shapes, they again only found an effect for the round shapes, which therefore suggested that the driving force behind this type of iconic mapping is the lip shape involved when producing round sounds rather than a cross-modal diagrammatic mapping.

Tamariz et al. (2017) conducted a similar study in which participants were assigned to one of two conditions. The first condition was a standard iterated learning design, as described above: participants had to learn the mapping between words and meanings (spiky and round figures) and this mapping was then taught to a new participant, and so forth. In the second condition, there were two participants in each generation who used the words to communicate with each other. The authors found that the emergent words were rated as more pointy under the communicative condition, suggesting that the process of communicating with others contributes to stronger iconicity effects. Carr et al. (2017) also found that iconic patterning can emerge through iterated learning. In their experiments, participants had to learn words for randomly generated triangles. Although the study was not designed to investigate vocal iconicity directly, the authors nevertheless noted that pointier triangles tended to be labelled by sounds listed as “pointy” by Ahlner & Zlatev (2010, p. 310) (e.g., /k/, /i/, /t/), while more equilateral triangles tended to be labelled using sounds listed as “round” (e.g., /b/, /m/, /u/). They found this effect under both a standard iterated learning design and a design in which participants had to communicate. Furthermore, Edmiston et al. (2018) showed that when environmental sounds, such as breaking glass or splashing water, are imitated, they become more stable and word-like, resembling ideophones. The final forms of the imitations could be matched to the source sounds above chance. Likewise, when people are asked to make up novel vocalizations for basic vocabulary words, naïve listeners are able to infer what they mean based on their phonetic forms (Perlman & Lupyan, 2018).

2. Method

In order to investigate how effects of vocal iconicity emerge, we used a relatively simple methodological setup. The participants were divided into five conditions (CONTROL, BIG, SMALL, ROUND and POINTY) and were presented with a recording of a single arbitrary (i.e. not iconic) seed word and asked to repeat it. These repetitions uttered by the participants were recorded and then used as stimuli for the next participant in the same transmission chain. This process was then repeated for 15 generations of participant per transmission chain. In the CONTROL-condition, the word was simply passed down 15 generations, but in the other conditions the participants were primed with a meaning connected to the word they heard as illustrated in Figure 1.

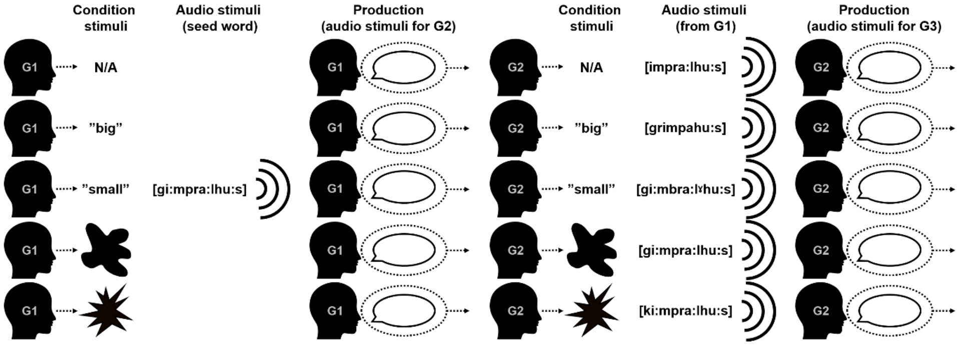


Figure 1. Illustration of the experimental procedure for the five conditions. The first-generation participants (G1) are exposed to their condition-specific visual stimuli and then to the seed word. They then repeat the word and their production was, in turn, used as the audio stimulus for the subsequent generation in the same transmission chain. This process was iterated until all chains had successfully transmitted the evolving string of sounds through 15 participants.

2.1 Participants

Participants were recruited online via the Figure Eight crowdsourcing platform which made it possible to include participants from several countries and with a range of different first languages. The participants were prevented from participating in the experiment more than once by identifying themselves with their unique worker IDs.

The aim of the study was to include 15 generations (participants) per transmission chain and 20 transmission chains for each of the five conditions, for a total of 1500 unique participants. To achieve this, we recruited 2854 participants, but 1354 were excluded for one or more of the following reasons: a) Misunderstanding the task, such as just repeating the meaning stimuli ("big", "small" etc.) asking a question about the task; b) Providing recordings of low quality (e.g. lack of sound, interfering background noise or recordings in which there were no recognizable sounds from the previous generation); or c) Providing recordings with obvious lexical interference, such as mistaking the presented audio as a word or phrase in a real language. The CONTROL-condition required 554 participants to yield 300 usable recordings, the BIG-condition required 592, the SMALL-condition required 591, the ROUND-condition required 565 and the POINTY-condition required 552. The participants were paid 50 cent USD for completing the task and the study was conducted under established ethical standards approved by the Linguistics & English Language Ethics committee at University of Edinburgh.

2.2 Stimuli

Of the five conditions, four were designed to prime the participants with a meaning by including either of the semantically oppositional poles of the SIZE-domain (*BIG-SMALL*) or the SHAPE-domain (*ROUND-POINTY*). The meanings for the BIG- and SMALL-conditions were conveyed in text since stimuli based on illustrations would require comparison in order to convey the correct meaning. The participants were either presented with the sentence "The word you are going to hear means **big**" in the BIG-condition or with "The word you are going to hear means **small**" in the SMALL-condition. The biases for the ROUND- and POINTY-condition were conveyed through shapes presented visually as shown in Figure 2. In the *CONTROL*-condition participants were not primed with a meaning.

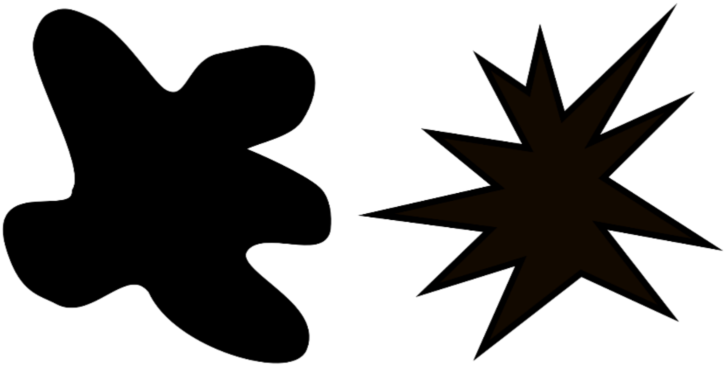


Figure 2. Visual stimuli for the ROUND- and POINTY-conditions.

All transmission chains were initialized with the same single seed word (i.e. the same audio stimulus was presented to the first participant in every transmission chain). This was to make it as easy as possible to track the development of sounds and groups of sounds over generations and for easier comparison across conditions. To allow for a variety of different potential iconic strategies to emerge, we designed the seed word to include a typologically, acoustically and articulatory varied selection of segments.

The most crucial features that had to be included in the design of the seed word was for it to be arbitrary (i.e. it does not carry iconic biases in any established semantic or phonetic direction) and to accommodate a reasonable mutation rate (i.e. to ensure that the arbitrary seed word can evolve phonetically, it should be somewhat difficult to remember). If the word were too easily learned, the participants would be able to repeat it perfectly and there would hence be no space for evolution to operate in.

The seed word was designed to consist of three syllables. The sounds were selected to be typologically common (Mielke, 2004-2020; Moran et al., 2014) since the initial seed word was assumed to adapt to the participants' phonologies quickly which would leave the use of uncommon sounds for increasing mutation rates unnecessary. Long versions of the three most extreme vowels, [iː], [aː] and [uː], were included, and the seven featured consonants were selected to be evenly distributed across manners and positions of articulation, as shown in Table 1.

Table 1. Distributions of consonants across five generalized manners of articulation and three generalized positions of articulation in the seed word (generation 0).

	labial		alveolar/palatal		velar/glottal	
	-voice	+voice	-voice	+voice	-voice	+voice
nasal		[m]				
stop	[p]					[g]
fricative			[s]		[h]	
trill				[r]		
lateral				[l]		

Approximately the same number of voiceless and voiced consonants was used in the word and consonant clusters were designed to include both voiced and voiceless sounds. In addition, the voiceless consonants were placed in the same syllables the vowels with lower F_2, [u] and [a], and the voiced, grave (Jakobson et al., 1952) consonants in the same syllable as the vowel with the lowest F_2, [i], to distribute the general spectral energy throughout the entire word. The selected parameters resulted in the word form [giːmpraːlhuːs] which was then recorded by a female native speaker of Czech with an academic background in linguistics to ensure a phonetically neutral pronunciation of the word. The selected segments of the word are present in, on average, 76% of the 2155 phonologies available in the PHOIBLE Online database (Moran et al., 2014): [g] 64%, [i] 93%, [m] 95%, [p] 87%, [r] 38%, [a] 91%, [l] 66%, [h] 65%, [u] 87%, and [s] 77%.

2.3 Procedure

The task began with the following general instructions: "In this task you will hear a word in an "alien" language. We will also tell you the meaning of the word. Your task is to listen carefully to the word and repeat it into your microphone. Make sure your speakers or headphones are switched on and the volume is turned up. First, we will tell you **the meaning of the word**. Then you will **hear the word**. There will then be a **3-second pause**. Finally, you must **repeat the word** into the microphone.". Participants in the CONTROL-condition, however, were not told that they would be presented with the meaning of the word.

Next, the participants entering the ROUND- and POINTY-conditions were presented with the round and pointy shapes. Those entering the BIG- and SMALL-conditions were presented with the text stimuli and were then required to confirm that they read the text properly by typing “big” or “small” depending on condition in order to continue with the task. This was included to make sure that the participants actually actively read the text stimuli since these could be easily overlooked as compared with the shape stimuli. This step was skipped for the participants in the CONTROL-condition who instead proceeded directly to the listening and production steps.

The first participant in each transmission chain listened once to the constructed arbitrary seed word, which was followed by a 3-second pause after which they had to repeat what they heard into their microphone. After completing the task, the participants were asked what they thought the word meant along with a few background questions (native and other languages). The utterance that the participant recorded was then uploaded to our server. All recorded stimuli were manually checked by the experimenter. Often it was also necessary to normalize the volume to a consistent level and/or trim the recording to only include the actual utterance. The recorded utterance was then used as the stimulus for the next participant in the same transmission chain.

2.4 Data analysis

After data collection was completed, the audio recordings were transcribed into the International Phonetic Alphabet (Appendix 1). Tones, stress or phonemic length were not taken into consideration for the analysis as they seldom are transmitted correctly when speakers from different languages attempted to pronounce utterances with these features. Diphthongs, triphthongs, affricates and coarticulations were divided into their components and analyzed as separate segments for comparability reasons.

The transcribed sounds were then categorized according to six binary *sound parameters*. Vowels were divided into HIGH and LOW, FRONT and BACK, and ROUNDED and UNROUNDED, while consonants were divided into GRAVE and ACUTE, VOICED and VOICELESS, and SONORANT and OBSTRUENT (see Table 2).

Table 2. The included sound parameters and sound groups, as well as examples of typologically common segments of each sound group.

Principal class	Sound parameter	Sound group	Segment examples
Consonant	Position	GRAVE	m, ŋ, p, k, b, g, f, h, v, w
		ACUTE	n, t, d, s, z, l, r, j
	Manner	SONORANT	m, n, ŋ, w, l, r, j
		OBSTRUENT	p, t, k, b, d, g, f, s, h, v, z
	Voicing	VOICELESS	p, t, k, f, s, h
		VOICED	m, n, ŋ, b, d, g, v, z, w, l, r, j
Vowel	Height	HIGH	i, e, ə, u
		LOW	a, o
	Backness	FRONT	i, e, a
		BACK	ə, o, u
	Roundedness	UNROUNDED	i, e, a, ə
		ROUNDED	o, u

The vowel sound groups correspond loosely to the first three formants, vowel height corresponds to F_1, backness to F_2 and vowel roundedness to F_3, and thus, cover most of the variation used for distinguishing vowel segments across languages (Ladefoged 2001, p. 32-36). Furthermore, energy level differences in F_1 and F_2 have been iconically linked to size, distance, dominance etc., while the roundedness of F_3 has been linked to shape. The HIGH-group included high, near-high, high-mid and true-mid vowels (including [ə]), while the remaining vowels were assigned to the LOW-group. The FRONT-group included front and near-front vowels and the BACK-group included central, including [ə], near-back and back vowels. Finally, the ROUNDED-group included all rounded vowels and UNROUNDED-group unrounded vowels.

Consonants are considerably more articulatorily diverse than vowels which is why the included sound groups feature both manner- and position-based distinctions. The VOICELESS-VOICED distinction cuts through all consonants and is used phonemically in most languages (Ladefoged & Maddieson, 1996, p. 44-46; Ladefoged, 2001, p. 63-65). In addition, it is, like F_1 and F_2, iconically associated to a number of meanings. The GRAVE-ACUTE distinction (Jakobson, Fant & Halle, 1952) was included since it differentiates between perceptually sharper and duller sounds which has also been linked to iconic associations (Lapolla, 1994). The GRAVE-group included the consonants produced by using the lips (bilabials through linguolabials) and the area from the soft palate and back (velars through glottals), while the ACUTE-group included all consonants produced using the hard palate (dentals through palatals). Lastly, the SONORANT-OBSTRUENT distinction was included since it is also one of the most fundamental ways to classify consonants (Stevens, 1998, p. 249-255). The contrast between sonorants' continuous, non-turbulent airflow and obstruents' obstructed airflow could iconically evoke e.g. noisiness vs smoothness or other related meanings.

2.5 Statistical model

We modelled the proportion of vowels or consonants of each particular sound parameter (HIGH-LOW, FRONT-BACK, ROUNDED-UNROUNDED, GRAVE-ACUTE, VOICED-VOICELESS, SONORANT-OBSTRUENT) out of the total number of vowels or consonants in the word for generation 0 (seed word) through 15. Proportions rather than absolute values were chosen in order to compensate for reduplication and word length effects. The proportions were calculated separately for vowels and consonants since it is possible that some transmission chains might utilize the former iconically, while others might utilize the latter. For example, if an association is found between a meaning and high frequency sounds, the sound could be voiceless consonants, front unrounded vowels, or both. Thus, a phonetic form such as [tuta] was analyzed as 100% [t] in terms of its consonants, and 50% [a] and 50% [u] in terms of its vowels. We then used binomial mixed models with generation and condition as predictors, with an interaction. One such model was fit for each of the six sound parameters. To account for non-independent nature of observations from the same chain, we included chain as a random intercept. This may mitigate the problem of autocorrelation of residuals from adjacent observations, although this model still represents a simplification of the iterated learning process. To minimize the risk of overfitting with 11 regression coefficients per model, we imposed their conservative shrinkage to zero with the horseshoe prior (Carvalho et al., 2009). The models were fit with R package *brms* (Buerkner, 2017). We first modeled the changes in proportion of each sound parameter and condition, including the CONTROL-condition. We then also compared the changes of proportions for each of the stimuli-conditions to the changes of proportions of the CONTROL-condition.

3. General results

On average, the original 10 segments (3 vowels, 7 consonants) of the seed word were reduced by approximately 3 at generation 15. On average, words produced in the CONTROL-condition contained 6.65 segments (2.75 vowels, 3.95 consonants), words in the SMALL-condition contained 6.95 segments (2.8 vowels, 4.15 consonants), the BIG-condition contained 6.65 segments (2.7 vowels, 3.95 consonants), the ROUND-condition contained 7.05 segments (2.9 vowels, 4.15 consonants) and the POINTY-condition contained 6.65 segments (2.65 vowels, 3.85 consonants). The reduction of total word length was mainly caused by the loss of consonants, which at generation 15 were reduced from the original 7 to approximately 4. The vowels, on the other hand, were only reduced by about a quarter of a segment on average. It is quite possible that the reason for these differences between consonants and vowels could be attributed to general phonotactical effects that favor simple syllable structures such as CVCVCVC.

All conditions, except BIG, showed significant changes for at least two of the investigated sound parameters (see Figure 3 and Appendix 2). However, the HIGH-LOW and SONORANT-OBSTRUENT parameters did not produce any significant changes. The proportion of FRONT vowels decreased in the CONTROL-condition (-13.2% 95% CI [-20.7, -5.1]), the POINTY-condition (-7.1% [-16.1, -0.1]) and the ROUND-condition (-13.7% [-21.6, -6]). Correspondingly, the proportion of ROUNDED vowels increased in the ROUND-condition 12.2% [3.6, 19.8] which is to be expected, since, typologically, rounded vowels are generally back while unrounded vowels are front. Conversely, the SMALL-condition produced a notable decrease of ROUNDED vowels (-10.3% [-16.9, -2.8]). The proportion of GRAVE consonants decreased in all conditions; CONTROL-condition (-7.1% [-13.9, -0.7]), SMALL-condition (-14% [-19.3, -8.3]), BIG-condition (-15.1% [-21.6, -9.3]), POINTY-condition (-15.8% [-21.4, -10.6]), ROUND-condition (-8.1% [-14.4, -0.8]). Lastly, the proportion of VOICED consonants increased slightly in ROUND-condition (7.3% [1.8, 13]).

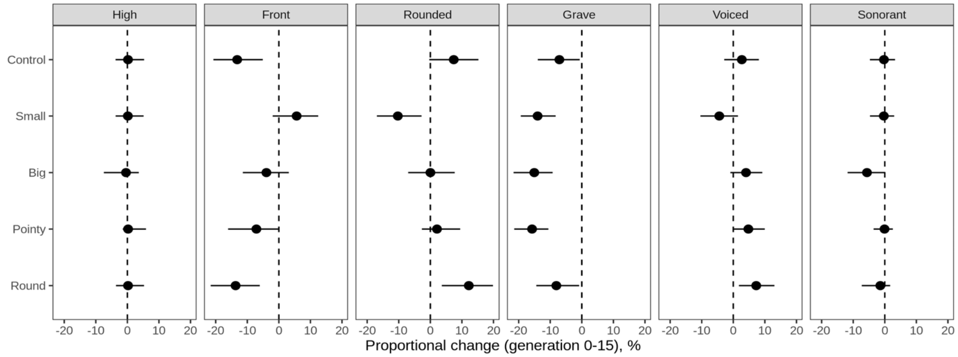


Figure 3. Change in the proportion for of the six sound parameters from generation 0 to generation 15. Shown: median of posterior distribution and 95% CI.

When comparing the stimuli-conditions to the CONTROL-condition (see Figure 4 and Appendix 3), the results crystalized and became easier to interpret. There were two cases for which the 95% CI clearly excluded zero. First, the proportion of FRONT vowels increased in the SMALL-condition by an additional 18.8% [8.3, 27.9] compared to the CONTROL-condition. Second, this was mirrored by a decrease of the proportion of ROUNDED vowels in the SMALL-condition versus the CONTROL-condition by -17.8% [-27.0, -7.4]. In addition, a weaker yet significant effect was found for the proportion of GRAVE consonants which decreased in the POINTY-condition by -8.7% [-16.6, -0.5] when compared to the CONTROL-condition.

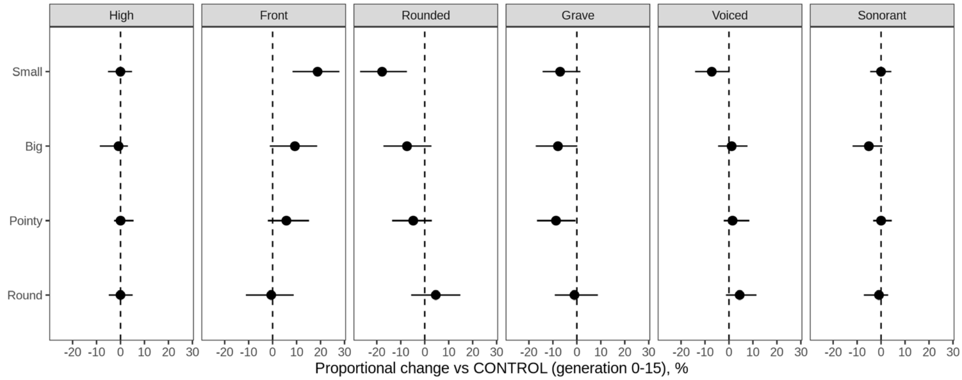


Figure 4. Contrasts between each stimuli-condition and the CONTROL-condition in the change in proportion for of the six sound parameters from generation 0 to generation 15. Shown: median of posterior distribution and 95% CI.

Furthermore, as shown in Figure 5, the significant changes compared to the CONTROL-condition started taking off around generation 5 and gradually increased, which can be seen most clearly in the rounded-unrounded parameter. This suggests that it is possible that even stronger effects might be observed over longer transmission chains (cf. Tamariz et al., 2017).

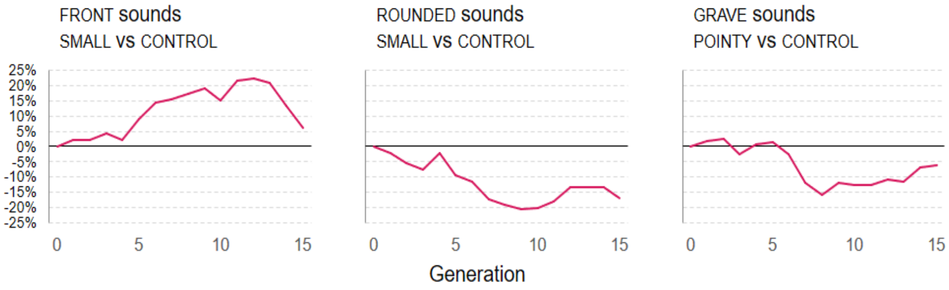


Figure 5. Showing the average proportional change of the three sound parameters which were found to be significant when compared to the CONTROL-condition from generation 0 to generation 15. Left: Proportional change of the FRONT sound group (and reversely the BACK sound group) in the SMALL-condition vs the CONTROL-condition. Center: Proportional change of the ROUNDED sound group (and reversely the UNROUNDED sound group) in the SMALL-condition vs the CONTROL-condition. Right: Proportional change of the GRAVE sound group (and reversely the ACUTE sound group) in the SMALL-condition vs the CONTROL-condition.

3.1 Secondary effects

Nasality was sometimes transmitted and spread throughout the word as a result of speakers without nasal vowels in their phonologies being influenced by the nasal vowels produced by Portuguese and French speaking participants. Several cases of palatalization likewise seem to have occurred due to the inherit phonologies of speakers of Portuguese and Slavic languages. Novel innovations of tone, stress, intonation and volume also lingered on for a couple of generations. As the participants of the BIG- and SMALL-condition were required to confirm that they correctly understood the meaning of the word, they had no problem with correctly answering the last question of the task which asked what they thought the word meant. This was also the case of the CONTROL-condition in which the participants almost exclusively answered “none” about what they thought the meaning of the word was. The ROUND- and POINTY-condition, on the other hand, involved shapes rather than text stimuli and were hence more open to interpretation. A large portion of the answers included some sort of phonetic transcription of the audio stimuli the participants were provided with or some version of “I don't know”. The clearest patterns that related to the actual stimuli included “stain”, “spot”, “spill”, “ink”, “blob” and “leaf” for the ROUND-condition and meanings related to explosions, e.g. “explosion”, “impact”, “blast”, “burst”, “star”, “bang”, “boom” and “crash”, for the POINTY-condition.

3.2 Possible increased learnability

Some of the transmission chains showed a tendency to stabilize as there was a reduced number of changes over consecutive generations (see Appendix 1). This reduced number of transmission errors could indicate that the words were changing in a way that made them easier to learn through vocal iconicity which could be tested by calculating Levenshtein distance (Gooskens & Heeringa, 2004) for each transmission. However, as a result of the participants various mother tongues, the present data is too complex for simply counting difference between two sequences of single-characters. The data featured several cases when a sound moved across syllables which would be counted as both a deletion and an insertion, for example the [m] in generation 9 [montʃabus] and generation 10 [oɹtʃambus] in CONTROL-chain 8. This is of course misleading since the word carries the same sound value in both generation 9 and 10, especially since we are interested in the proportions of sound groups in words across generations rather than their positions within the words. Furthermore, there was also cases of one sound being reinterpreted as two sounds, as illustrated by [ã] and [aw] in generation 9 [bivohã] and in generation 10 [iwohaw] in ROUND-chain 5. Similarly, there were many examples of merges in which two or more sounds were being reinterpreted as one sound, as illustrated by [mp] and [b] in generation 2 [gimprahus] and in generation 3 [hibrahu] in SMALL-chain 2. Aside from these, there were also cases of voicing being switched between two sounds in the same syllable, reduplications and, of course, substitutions to very similar sound in order to comply with the speakers’ native phonologies (e.g. [w] and [v]). Thus, such an analysis is left for a future study that can take splits, mergers, metatheses, assimilations, etc. adequately into consideration.

4. Discussion

4.1. Primary sound effects

The CONTROL-condition produced two significant changes and all experimental conditions produced decreases in GRAVE consonants which illustrates the difficulty with designing a completely non-arbitrary and typologically neutral seed word. However, as explained above, the decrease in FRONT vowels and increase in ROUNDED vowels in the CONTROL-condition are in fact two surface results from an underlying chance in the distribution of sounds being used. Furthermore, both of these effects are found in the experimental conditions as well, with the notable exception of the SMALL-condition. In addition, to minimize the risk of finding effects by chance, we controlled for multiple comparisons by imposing a conservative shrinkage prior (see Section 2.5). This suggests that these changes should be regarded as a stabilization toward a kind of typological default.

When it comes to the results yielded by comparisons between the CONTROL-condition and the experimental conditions, both vowels and consonants seemed to produce iconic effects. This is also in line with other studies that have shown that both vowels and consonants are involved in size and shape iconicity (Ahlner & Zlatev, 2010; Nielsen & Rendall, 2013; D'Onofrio, 2014). The clearest results were produced by the SMALL-condition and showed a preference for FRONT and UNROUNDED vowels and a dispreference for BACK and ROUNDED vowels. The preferred sounds were typically represented by [i], [e], [ɛ] and [a] which also have the highest average vowel frequencies for the first formant ([a] and [ɛ]) and for the second formant ([i] and [e]). Thus, the associations between sound and meaning align well with Ohala's (1994) frequency code which predicts that smallness, as well as related meanings, are evoked by high and/or rising frequencies of vocalizations. Furthermore, a plethora of cross-linguistic and experimental studies have found similar associations between size and energy level or pitch as explained in Section 1.1. For example, Erben Johansson et al. (2020) found SMALL and SHORT to be associated with voiceless consonants, which of course also involve high frequency energy (Ohala, 1994). Consequently, this association should probably be regarded as one of the most robust iconic effects found since it aligns with solid typological and experimental evidence.

The most surprising result was the decrease of GRAVE consonants, and the corresponding increase of ACUTE consonants, in the POINTY-condition, since one of the most common GRAVE consonants, [k], is often featured in pointy stimuli words, e.g. [kiki]. The results do, however, align with Blasi et al. (2016) who found that words meaning ‘star’ tend to contain alveolars and alveolars cross-linguistically and with experimental studies showing that consonants might play a somewhat larger role than vowels in shaping vocal iconicity (Nielsen & Rendall, 2011; Fort et al., 2015). This does not necessarily mean that [k] is confirmed to be disfavored when paired with pointy shapes, since the sound group also contains labial and voiced consonants. Nevertheless, this has some implications for bouba-kiki tasks since it demonstrates that using ready-made stimuli words for experiments such as this might impact the results negatively. Furthermore, the results also suggest a slightly more complex mapping between sound and meaning than pitch-to-size. Acute sounds do generally involve higher frequency energy than grave sounds, but the sound group included both voiceless and voiced sounds which is the primary consonantal distinction between high and low frequency energy. Since no effect was found for the VOICED-VOICELESS parameter, it is possible that pressing the tip or blade of the tongue against the hard palate could evoke a tactile sensation of sharpness and hardness. Thus, this mapping might be driven by the frequency of energy in the sounds, but a tactile component seems to be involved at least secondarily.

The significant changes in the ROUNDED-UNROUNDED parameter also seemed to be amplified over time. One might therefore assume that the proportions of iconic sounds would increase indefinitely until the transmitted words would consist only of front unrounded vowels and acute consonants. This is, however, unlikely for a number of reasons since linguistic material from various sources is dynamically introduced into words as languages change over time. Firstly, words, except for a very small number per language, generally adhere to phonotactic restriction that require them to include both vowels and consonants. This is because there simply are not enough unique individual phonemes in languages to be assigned to all meanings that need to be conveyed. Secondly, many languages require all words, including loans, to have affixes attached to them in order to be grammatical. Similarly, the participants included in the present study were also instructed to repeat what they heard which forced them to retain considerable parts of the syllable structure and sounds from the previous utterance. And thirdly, full-word iconicity is usually only observed in onomatopoeia, e.g. the sound-to-sound mappings in bird names that mimic bird calls (such as *cuckoo*), and not in relative diagrammatic iconicity.

4.2. Primary condition effects

The results revealed a notable correlation between associated phonetic parameters and semantic domains. The SMALL-condition, belonging to the continuous SIZE-domain, was found to be appropriately mapped to the equally continuous frequency scale, while the sounds mapped to POINTY-condition seems to be, at least partially, associated with sounds through tactile mappings. The preference for these different types of associations could be grounded in the semantic features of the stimuli as BIG and SMALL are rather abstract and require comparison in order to be defined which is a good fit for degrees of pitch. ROUND and POINTY are considerably more visually concrete and their contrasting geometrical features could also be used to tell them apart from shapes such as squares or ellipses. Accordingly, the sounds associated with POINTY portray similar concreteness.

Furthermore, when compared to the CONTROL-condition, the SMALL- and POINTY-conditions produced several iconic effects while the BIG- and ROUND-conditions did not. This could perhaps be explained by the fact that semantic poles are not equally iconically charged. Similar results have been found by Nielsen and Rendall (2011) and Tamariz et al. (2017), who showed greater iconic effects in pointy shapes. However, Jones et al. (2014) only found effects in round shapes and Fort et al. (2018) X showed that effects in round shapes are more prevalent infants while effects in pointy shapes seem to emerge with age. What joins the POINTY and SMALL-conditions together is semantic markedness which by extension could be a possible explanation for an increased iconic effect. To begin with, antonyms and semantically oppositional concepts are cognitively closely related. Several studies show that when a word fails to come to mind, antonyms replace that word (Söderpalm, 1979; Linell, 1982) and if one member of an oppositional pair is presented as stimulus, the other member is frequently uttered (Deese, 1965). Antonymous concepts also occur in the same sentence more frequently than chance (Justeson & Katz, 1991; Willners, 2001), morphological encodings usually come in oppositional pairs (Cinque, 2013) and some antonyms are processed significantly faster than the non-canonical antonyms (Paradis et al., 2009). Notwithstanding, unmarked poles of semantically opposite pairs are understood earlier by children and it takes adults longer time to make comparisons between objects when a marked pole is used as reference (de Villiers & de Villiers 1978, p. 139-141). Thus, it could be more cognitively affordable for only one pole to be iconically charged as the other pole will be sufficiently mapped by association, at least for tightly connected oppositional concepts.

In addition, studies have shown that poles with reversed position to the expected patterns can emerge in iconic conditions. By investigating words denoting spatial deixis from 12 Indo-European branches diachronically, Johansson & Carling (2015) found that languages tend to align the relative distance from the speaker with sounds of decreasing frequency, but also that most of the forms which did not adhere to this pattern were aligned in reverse order rather than randomly. While such patterns could arise from cultural specialization or marginalization which is transmitted into speech for the purpose of keeping languages secret or distinct from speakers of other languages (Carling et al., 2014, p. 89), they could also provide some functional benefits. Since the iconic association, albeit present in only one of two poles, is cognitively grounded, the reversed order of the poles is still more learnable and easier to remember than a pair of other sounds because the extremes of both mapped parameters are still present. Likewise, Westbury et al. (2018) showed that poles of the same semantic dimension differ in their iconic predictability by analyzing 18 semantic categories and thousands of randomly-generated nonwords. Additionally, the continuous SIZE-domain (SMALL) was associated with the continuous frequency scale (through the FRONT and UNROUNDED groups) while the dichotomous SHAPE-domain (POINTY) was associated with dichotomous tactile mappings (through the ACUTE group). Thus, this suggests that the alignment between sound and meaning cannot be indiscriminately selected, but that diagrammatic iconicity is grounded in mappings that incorporate felicitous and correlating semantic and phonetic features.

4.3 What is required for iconicity to emerge?

Jones et al. (2014) showed that iconicity can emerge through transmission. However, as with most previous experiments that have investigated relative iconicity, the participants were highly restricted due to the use of text-based artificial languages or forced-choice experimental design. While accompanied by the same methodological restrictions, Tamariz et al. (2017) only found that iconicity emerges through communicative interaction and not through individual reproduction. The stronger effect that interaction brings to the table was attributed to an increased number of possible innovations that could increase iconicity as well as a larger number of possible adopters of the signal, which increases the chance of labels fitting with meanings in a speech community. Therefore, Tamariz et al. argue that their results can be interpreted as evidence for random mutation and selection rather than guided variation; in other words, cultural traits acquired by a population through individual learning drive cultural evolutionary processes. This, in turn, is aligned with large-scale cross-linguistic studies on lexical iconicity which show that iconic forms are present throughout languages and language families, but that the same sound-meaning associations are not found everywhere at the same time (Wichmann et al., 2010; Blasi et al., 2016; Erben Johansson et al. 2020), which suggests that iconicity is in a perpetual process of decay and rebuilding (Johansson & Carling, 2015) and not conserved through time (Pagel et al., 2013).

Nonetheless, it would be unwise to underestimate the role of transmission and learnability in the dynamics of iconicity. Firstly, both Jones et al. (2014) and the present study showed that transmission alone is enough for iconic effects to arise. Secondly, the present study further suggests that very little is required in order for iconicity to emerge (Edmiston et al., 2018). Even without interaction between participants, constrained experimental setups, forced choice questions, premade stimuli words or using text as a proxy for spoken language, all of which could in some manner increase the likelihood of mapping sound to meanings correctly outside of the bouba-kiki effect (Cuskley et al., 2017), iconic effects seem to have emerged. Thirdly, there is overwhelming evidence that iconic forms, including language-specific ideophones, facilitate language learning and comprehension in both children and adults (Imai et al., 2008; Nygaard et al., 2009; Kantartzis et al., 2011; Imai & Kita, 2014; Lockwood et al., 2016a; Lockwood et al., 2016b; Massaro & Perlman, 2017).

However, while iconicity and synesthetic cross-modal mappings are present in the early stages of human ontogenetic development (Mondloch & Maurer, 2004; Maurer et al., 2006; Walker et al., 2010) and go at least as far back as the ancestor we share with chimpanzees (Ludwig et al., 2011; Perlman, 2017), they do not seem to disappear but rather gradually decrease with age, language competence and vocabulary size (Ludwig & Simner, 2013; Massaro & Perlman, 2017). The likely explanation for this is that iconicity does not scale well in language. In a less developed and lexically poor language, iconicity can aid in intuitively linking words to fundamental meanings, but as languages adapt to the expressive needs of their users, the number of distinctions that must be made cannot be handled by an iconic system. Thus, here is where iconicity falls short, as there simply are not enough unique iconic signals, either though sounds or gestures (Perlman & Cain, 2014), available to accommodate the diversity of meanings that language users might wish to express (Gasser, 2004; Westbury et al., 2018). Nevertheless, iconicity is still found in complex languages, though confined to specific functions where it excels in conjunction with arbitrary and systematic mappings between sound and meaning (Monaghan et al., 2011; Dingemanse et al., 2015). However, agents without advanced language competence, such as great apes, do utilize iconicity while they have very limited access to interactional language which suggests that the transmission of signals is enough to facilitate iconicity. This does not, of course, mean that interaction does not provide an even more advantageous environment for non-arbitrary associations, though it suggests that interaction might not be a prerequisite for iconicity.

5. Conclusion

We have shown that by flipping the classic bouba-kiki experiment on its head through the use of iterated learning and including a much larger number of participants than previous studies, it was possible for iconic effects to emerge. By using an explorative and simple methodological setup which included an auditorily modest linguistic environment without premade stimuli words or task training, we were able to get a deeper understanding of how vocal iconicity operates within the semantic SIZE- and SHAPE-domains. Not only were these results aligned with the sound-meaning associations found in large-scale cross-linguistic and experimental studies, but one of the effects gradually strengthened with generation as well, which indicates that stronger effects might be observed with longer transmission chains. Furthermore, semantically marked meanings (SMALL and POINTY) produced iconic effects, while unmarked meanings (BIG and ROUND) did not, probably due to the lower cognitive demand of memorizing one pole and inferring the other through contrast rather than remembering complete oppositional pairs. In addition, the different conditions showed mirrorings between the characteristics of the semantic domains and associated sounds, which indicates that diagrammatic iconicity mappings are grounded in the most suitable correlation between semantic and phonetic features available. Thus, these results indicate that linguistic transmission through disconnected language users is enough to investigate cognitive biases for vocal iconicity, which can easily be expanded to a range of iconically promising meanings, including non-oppositional semantic fields and semantically closely related meanings e.g. BIG, TALL, LONG, MANY, and so forth. Moreover, as learnability is increased through the adoption of linguistic forms that correspond to their meanings iconically, iconicity should be recognized as one of the crucial components in human language evolution and development.

Acknowledgements

A special thanks to Andrey Anikin for his help with the statistical analysis and to Šárka Erben Johansson, Gerd Carling and Arthur Holmer for their guidance and support.

References

Ahlner, F. &, Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies, 38*(1/4), 298-348.

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.

Bross, F. (2018). Cognitive Associations Between Vowel Length and Object Size: A New Feature Contributing to a Bouba-kiki Effect. *Proceedings of the Conference on Phonetic & and Phonology in German-speaking countries, 13*, 17-20.

Buerkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1-28.

Canini, K. R., Griffiths, T. L., Vanpaemel, W., &, Kalish, M. L. (2014). Revealing human inductive biases for category learning by simulating cultural transmission. *Psychonomic Bulletin & Review*, *21*(3), 785-793.

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.

Carling, G., Lindell, L. &, Ambrazaitis, G. (2014). *Scandoromani: Remnants of a Mixed Language*. Leiden: Brill.

Carr, J. W., Smith, K., Cornish, H., &, Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*, *41*, 892.923. https://doi.org/10.1111/cogs.12371

Carvalho, C. M., Nicholas G. P. &, Scott, J. G. (2009). Handling sparsity via the horseshoe. *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, 5, 73-80.

Cinque, G. (2013). Cognition, typological generalizations, and universal grammar. *Lingua*, *130*, 50-65.

Cuskley, C., &, Kirby, S. (2013). Synaesthesia, cross-modality and language evolution. *Oxford handbook of synaesthesia*, *20*, 869-907.

Cuskley, C., Simner, J. &, Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychological research, 81*(1), 119-130.

de Villiers, J. G. &, de Villiers, P. A. (1978). *Language Acquisition*. Cambridge MA: Harvard University Press.

Deese, J. (1965). *The Structure of Associations in Language and Thought*. Baltimore MD: The Johns Hopkins Press.

Dingemanse, M. (2011). Ezra pound among the Mawu. In Michelucci, P., Fischer, O. &, Ljungberg, C. (eds.), *Semblance and signification. Iconicity in Language and Literature, 10*, 39-54. Amsterdam: John Benjamins.

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.

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

Edmiston, P., Perlman, M. &, Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences, 285*(1874). 20172709. http://doi.org/10.1098/rspb.2017.2709

Erben Johansson, N., Carling, G. &, Holmer, A. (2020). *The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features*. Linguistic Typology.

Figure Eight Inc. (2018). San Francisco, United States of America. https://www.figure-eight.com/

Fónagy, I. (1963) *Die metaphern in der phonetik*. The Hague: Mouton.

Fort, M., Martin, A. &, Peperkamp, S. (2015). Consonants are more important than vowels in the bouba-kiki effect. *Language and Speech, 58*(2), 247-266.

Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., & Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Developmental science*, *21*(5), e12659.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26*(26), 434-439.

Gebels, G. (1969). An investigation of phonetic symbolism in different cultures. *Journal of Verbal Learning and Verbal Behavior, 8*, 310-312.

Goddard, C. &, Wierzbicka, A. (eds.) (2002). *Meaning and Universal Grammar: Theory and Empirical Findings* (2 volumes). Amsterdam & Philadelphia: John Benjamins.

Gooskens C., &, Heeringa W. (2004). Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language variation and change, 16*, 189-207.

Hamilton-Fletcher, G., Pisanski, K., Reby, D., Stefańczyk, M., Ward, J., & Sorokowska, A. (2018). The role of visual experience in the emergence of cross-modal correspondences. *Cognition*, *175*, 114-121

Hinton, L., Nichols, J. &, Ohala, J. J. (1994). Introduction: Sound-symbolic processes. In L. Hinton, J. Nichols &, J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Holland, M. K., &, Wertheimer, M. (1964). Some physiognomic aspects of naming, or, maluma and takete revisited. *Perceptual and Motor Skills*, *19*, 111-117.

Imai, M. &, Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B*, *369*(1651). http://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M. &, Okada, H.. (2008). Sound symbolism facilitates early verb learning. *Cognition, 109*(1), 54-65.

Jakobson, R., Fant, C. G., &, Halle, M. (1951). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, Mass.: MIT Press.

Johansson, N., &, Carling, G. (2015). The De-Iconization and Rebuilding of Iconicity in Spatial Deixis: A Indo-European Case Study. *Acta Linguistica Hafniensia, 47*(1), 4-32.

Jones, J. M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J. &, Vigliocco, G. (2014). The bouba effect: Sound-shape iconicity in iterated and implicit learning. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 2459-2464.

Justeson, J. S. &, Katz, S. (1991). Co-occurrence of antonymous adjectives and their contexts. *Computational Linguistics, 17*(1), 1-19.

Kantartzis, K., Imai, M. &, Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science, 35*(3), 575-586.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation, 5*(2), 102-110.

Kirby, S., Cornish, H. &, Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686.

Kirby, S. &, Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi &, D. Parisi (eds.), *Simulating the evolution of language*, 121-147. London: Springer.

Kirby, S., Tamariz, M., Cornish, H. &, Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87-102.

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

Ladefoged, P. (2001). *Vowels and consonants: an introduction to the sounds of languages*. Malden, MA: Blackwell Publishing.

Ladefoged, P. &, Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

LaPolla, R. J. (1994). An Experimental Investigation into Sound Symbolism as it Relates to Mandarin Chinese. In L. Hinton, J. Nichols &, J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Lindauer, M. S. (1990). The meanings of the physiognomic stimuli taketa and maluma. *Bulletin of the Psychonomic Society, 28*, 47-50.

Linell, P. (1982). *Speech errors and grammatical planning of utterances: Evidence from Swedish.* In W. Koch, C. Platzack &, G. Tottie (eds.), *Textstrategier i tal och skrift*, 134-151. Stockholm: Almqvist and Wiksell.

Lockwood, G. &, Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01246.

Lockwood, G., Dingemanse, M. &, Hagoort, P. (2016a). Sound-symbolism boosts novel word learning. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 42*(8), 1274-1281.

Lockwood, G., Hagoort, P. &, Dingemanse, M. (2016b). How iconicity helps people learn new words: Neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra, 2*(1). http://doi.org/10.1525/collabra.42

Ludwig, V. U., Adachi, I., &, Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *PNAS, 108*(51), 20661-20665.

Ludwig, V. U. &, Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex, 49*(4), 1089-1099.

Maurer, D., Pathman, T. &, Mondloch, C. J. (2006). The shape of boubas: Sound–shape correspondences in toddlers and adults. *Developmental science, 9*(3), 316-322.

Mielke, J. (2004-2020). *P-base. A database of phonological patterns.* http://pbase.phon.chass.ncsu.edu

Monaghan, P., Christiansen, M. H. &, Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347.

Mondloch, C. J., &, Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 133-136.

Moran, S., McCloy, D. &, Wright, R. (eds.) (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://phoible.org

Newman, S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology, 45*, 53-75.

Nielsen, A. K. &, Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition, 4*(2), 115-125.

Nielsen, A. K. &, Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 67*(2), 153-163.

Nielsen, A. K., &, Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 65*(2), 115-124.

Nygaard, L. C., Cook, A. E. &, Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition, 112*(1), 181-186.

O'Boyle, M. W. &, Tarte, R. D. (1980). Implications for phonetic symbolism: The relationship between pure tones and geometric figures. *Journal of Psycholinguistic Research, 9*(6), 535-544.

Ohala, John J. (1994). The frequency codes underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols &, J. J. Ohala (eds.), *Sound symbolism*, 325-347. Cambridge: Cambridge University Press.

Pagel, M., Atkinson, Q. D., Calude, A. S. &, Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences, 110*(21), 8471-8476.

Paradis, C., Willners, C. &, Jones, S. (2009). Good and bad opposites: Using textual and experimental techniques to measure antonym canonicity. *The Mental Lexicon, 4*(3), 380-429. Amsterdam: John Benjamins.

Perlman, M. (2017). Debunking two myths against vocal origins of language. *Interaction Studies, 18*(3), 376-401.

Perlman, M. &, Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture, 14*(3), 320-350.

Perlman, M. &, Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Scientific reports, 8*(1). 2634. http://doi.org/10.1038/s41598-018-20961-6

Ramachandran, V. S., &, Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Rendall, D., Kollias, S., Ney, C. &, Lloyd, P. (2005). Pitch (F 0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *The Journal of the Acoustical Society of America, 117*(2), 944-955.

Rogers, S. K. &, Ross, A. S. (1975). A cross-cultural test of the Maluma-Takete phenomenon. *Perception, 4*, 105-106.

Sapir, Edward (1929). A study in phonetic symbolism. *Journal of experimental psychology, 12*(3), 225-239.

Söderpalm, E. (1979). Speech errors in normal and pathological speech. *Travaux de l'Institut de Linguistique de Lund, 14*. Lund: CWK Gleerup.

Spike, M., Stadler, K., Kirby, S., &, Smith, K. (2017). Minimal requirements for the emergence of learned signaling. *Cognitive Science*, *41*, 623–658. doi:10.1111/cogs.12351

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Styles, S. J. &, Gawne, L. (2017). When does maluma/takete fail? Two key failures and a meta-analysis suggest that phonology and phonotactics matter. *i-Perception, 8*(4). http://doi.org/10.1177/2041669517724807

Swadesh, M. (1971). *The origin and diversification of language*. Edited post mortem by Joel Sherzer. London: Transaction Publishers.

Tamariz, M., Roberts, S. G., Martínez, J. I. &, Santiago, J. (2017). The interactive origin of iconicity. *Cognitive science, 42*(1), 334-349.

Taylor, I. K. (1963). Phonetic symbolism re-examined. *Psychological Bulletin, 60*(2). 200-209.

Taylor, I. K. &, Taylor, M. M. (1962). Phonetic symbolism in four unrelated languages. *Canadian Journal of Psychology/Revue canadienne de psychologie, 16*(4), 344-356.

Tufvesson, S. (2011). Analogy-making in the Semai Sensory World. *The Senses and Society, 6*(1), 86-95. https://doi.org/10.2752/174589311X12893982233876

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. &, Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21-25.

Westbury, C., Hollis, G., Sidhu, D. M. &, Pexman, P. M. (2018). Weighing up the evidence for sound symbolism: Distributional properties predict cue strength. *Journal of Memory and Language, 99*, 122-150.

Wichmann, S., Holman, E. W. &, Brown, C. H. (2010). Sound Symbolism in Basic Vocabulary. *Entropy, 12*(4), 844-858.

Willners, C. (2001). *Antonyms in Context: A Corpus-based Semantic Analysis of Swedish Descriptive Adjectives*. Lund: Lund University dissertation.

Appendix 1. Phonetic transcriptions of the collected audio recordings.

Chain	Generation	CONTROL	SMALL	BIG	POINTY	ROUND
1	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
1	1	impralhus	gimbralhus	grimpahus	kimpralhus	gimpralhus
1	2	diŋpraos	gimbrawos	grilpaws	kimpralhu	mumavmuv
1	3	dinkambos	gimbambos	drinpaws	pimpabu	novamo
1	4	dinkaboʃ	gimbapos	drinpaws	pinkavu	nofano
1	5	binkabos	gimbapos	grɛdinpaws	finkawu	nofɑno
1	6	bimkabos	gimbampows	grɛdinfaws	feŋkawu	nofano
1	7	dimkabos	gimbampos	grɛdinfaws	dʒɛŋkaow	afaʁnaw
1	8	difgabos	kitbabos	ɹɛlikbawt	dɛŋgawo	gasamõ
1	9	disgabu	getpovos	ɹɛlibawʈ	gɛŋgau	gasamo
1	10	diskabu	getpovos	ɹɛlibaʈ	dindau	kasamo
1	11	diskabu	kipevos	ɹɛlibaɹ	dindadu	kazamo
1	12	diskabɹə	tipevos	ɹɛliba	dindaŋdu	hatanməɹ
1	13	diskapə	tepivols	ɹɛliba	diŋdaŋdu	hatanmarow
1	14	iskawoɹ	tɛtəfons	tʃʊlifa	dindindu	hatanmajl
1	15	iskawoɹ	kɛrfols	tʃʊdefaj	dendendu	hakanman
2	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
2	1	gimpralhus	gimprahus	impralhus	kimpaws	gimprasdu
2	2	gimpalhus	gimprahus	impralhus	kimpows	miprasdu
2	3	minpalfos	hibrahu	inpaws	dɹimpls	mirazdu
2	4	mipapots	sibrahu	inpawʃt	kimpas	miraθdu
2	5	mipafots	sbrahu	inpawʃt	gampas	jumejraθdu
2	6	mipatsfo	sbraku	nuonfawʃt	kɛlfɛʃ	jumejglaθdu
2	7	miashowm	sigolaku	undfawʃt	kalfɛs	jumejdasdu
2	8	miɛsxo	siɲolaku	andfowst	kewsrəs	jumejdasdu
2	9	viɛsxow	siɲolaku	anfowswa	kɹəsəs	tyomedastu

2	10	wihæshow	seɟolako	anfowfa	pɹəsəs	tiomedasdu
2	11	vihæsxo	seɟolako	anfowfa	ɹeses	djumidasdu
2	12	vihɛshow	seɟolako	fofofa	ɹeseses	djumidasdu
2	13	vihɛsho	senolako	hohoha	ɹesesam	piemivastu
2	14	bixɛsho	senolaŋko	hohoho	dɹeseɟɑm	kemivastu
2	15	biçɛsho	senolanko	xoxoxo	tɹɛsɛto	tjɛmibasdu
3	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
3	1	gintrahos	ginprajhus	grimpahuəs	gimprahos	gimpralhus
3	2	nintrahos	gintɹajhus	pusbansus	gimprahʊs	gikkamus
3	3	mintraxos	intrajhos	tunspanstuns	inpranhos	kamuʃ
3	4	ɛntrahos	intrajhos	daŋspastumf	durangos	ikamoʃ
3	5	ɛntraxos	endɹajdɹos	taməstʊ	durəngos	ikaməʃ
3	6	ɛntrahos	hedɹajdɹas	daɹheɹtʊ	durɛngos	ikamaʃ
3	7	ɛntrahas	hetɹadɹatɹaws	gjaʁeʁtʊ	dorɛngos	ikafaʃ
3	8	ɛntrahas	ntədʒaɹs	gjaʁeʁdu	dorɛnkos	ikafas
3	9	hɛndrahos	intendʒas	gjɒeʁdu	vorinkos	ikafas
3	10	tɛndʒoxas	insindʒɑs	jɒeʁdu	horinkas	ikafas
3	11	tɛndʒowhɑn	teŋʃiŋstaŋ	jɒerdu	haringas	ikasfas
3	12	tɛndʒowhɑn	tɛnʃistaŋ	dɒerdu	dartingas	iasas
3	13	pɛndʒowxa	tɛŋʃistaŋ	mɒɹendun	gasndas	iasas
3	14	bɛndʒuxam	taŋʃistaŋ	uintu	gasɛmdɑs	iasas
3	15	bɛndʒuxaw	daʃistaw	uintu	gasamdos	iasas
4	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
4	1	gimpralhus	gimpralhus	gimpaħus	gimprahus	gimprahus
4	2	gimpragjus	iterabʊs	ibriðragost	grirarunts	dimprahus
4	3	ginkavjʊs	entoralyns	iðbibragos	grinaruŋks	poravos
4	4	ginkafjʊs	empeɹomins	isbiŋagos	grinows	boɖvos
4	5	ginpafjʊs	tempreomins	isbiagos	grinows	porlabos
4	6	istaus	temkromvis	izbiagos	grinowʃ	wohlabo
4	7	tispaus	henkromis	izbiagos	grinowtʃ	bobobo

4	8	tispawt	henskromɛ	ibiagon	krinokʃt	bobobo
4	9	istpawt	hentskromɛ	ibiaɣon	gɹinoʃ	bobobo
4	10	itspawts	henskromɛ	bitsbiapon	ɹieɹnoʃ	bobobo
4	11	itspaks	isklamɛ	itsbiapon	ɹinoliʃ	bobobo
4	12	itspans	pispal	kwitsbiaplon	ɹiŋoliʃ	obobo
4	13	glikwɑns	pispala	plisbiao	linalis	obobo
4	14	gliksɑns	tispara	dɛsbiao	inalis	pobobo
4	15	litwɑn	tispara	bejsbiaow	nales	bowbobow
5	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
5	1	gimpralhʊs	gimprahus	gimpralħus	ginpɹahus	gintravos
5	2	impralhos	dimprahus	gimbɽawhos	gɛnpagos	hitravos
5	3	impravasa	dimpraħu	imbrawos	gejŋpagwor	kivtrawow
5	4	imprapasa	dimpraxu	imbrawos	ginbagort	kiprahow
5	5	impropesat	impraxu	imprawas	miabort	kibrahow
5	6	inopropesa	impraho	improbos	miapoɽ	ibɹahow
5	7	inopoposa	pobraho	implobosə	miapol	ibahaw
5	8	iopokosa	kobraho	impabosəm	miapal	ibahaw
5	9	iopokosa	kobraho	impabosa	ɲepaʎ	bivohã
5	10	fipokosopa	kobako	imbabosa	njepal	iwohaw
5	11	fiokoʃoba	kobako	imbaborsa	jʊfo	mohab
5	12	vibokoʃalba	howako	imbabolsa	jʊfo	mohamɛθ
5	13	bɹiʃəboninʃdʒɹ	powako	imbavolsa	jʊwfɹo	mohan
5	14	gɹiʃəboɹnisdan	kowako	imbawozba	dʒofo	moharʊ
5	15	wiʃorajasa	koako	imbuzba	iafom	mahow
6	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
6	1	gimpralhus	gindahus	gimprahus	gimpahus	gimpralhus
6	2	gimpralfus	daɹus	ɛmprahos	kikadus	kimpraðhus
6	3	gimpralfus	baɹus	ɛmbahos	pikalun	tʃympralwos
6	4	gimprawfus	baɹuʃ	ɛmbɑkos	likalow	dimbrafos
6	5	endʒawfus	taɹuʃ	ɛmbakos	mekalow	ɛdʢsʢ

6	6	endowfus	taruʃ	ɛmbokos	mɛqχalo	ɛdõsõ
6	7	dowdowfus	taruʃ	ɛmbokoʃ	mɛkalo	ædõsõ
6	8	dowgofus	karuʃt	ɛmbokoʃ	nɛdalo	ɛdrardaw
6	9	dowkofus	skweniʃ	ɛmbokoʃ	minalo	undratr
6	10	gawgawpus	kweniʃ	ɛmbokoʃ	minalo	ndratr
6	11	gawkapus	kweniʃ	ɛmbokoʃ	minaŋgo	məndratr
6	12	gawkopus	kwejniʃ	ɛmbakoʃ	minlajngwo	mundratr
6	13	gawkompus	kwejniʃ	ɛmbakoʃ	milajgwow	bondabəɹ
6	14	gawkopus	kwejniʃ	hewakoʃ	milajgwo	bodaŋgək
6	15	kowkowplus	wejniʃ	iwəgoʃ	milajwo	undadag
7	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
7	1	gimpralhus	gɹimbɹahahu	grimpralhus	imprashus	gimpɹalɹus
7	2	gimrahut	nɛŋhaahu	grinkraprus	imprasus	infraru
7	3	gimrahu	mɛjaahu	gringrasgrʊs	intraʊs	infraru
7	4	igaɹʊ	nihahahu	gringranrʊs	intrabus	infagʊw
7	5	igarʊ	hahu	wewiwagʊs	iɕabow	difaigum
7	6	igaru	jahu	wewiwaŋwʊs	izabo	ifogo
7	7	egaru	jahu	weɽiɽaŋgos	izabo	ifogo
7	8	egarʊ	jahu	weɽhihaho	bizabo	ifogo
7	9	egarʊ	jahu	weʁhihaʁho	isawo	ifogo
7	10	ejarʊ	jahʊ	ewibihaʁbiχ	isawo	ifodu
7	11	ejalʊk	bjahu	ivivibixabixa	isawok	infodo
7	12	ejalʊk	mjahu	avirbihabiha	isawok	infodo
7	13	ejalʊk	wiaɹhu	ɛrbirlihæ	pisaloko	inforðo
7	14	ejlu	hiaɹhow	ɛnəmbjurixat	susaloko	okordo
7	15	ejlo	siathaw	inembuɹiha	tusalopo	hokogo
8	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
8	1	gimpralkus	gimpɹalhus	gimpɹalhus	imbɹalhʊs	imprahus
8	2	nuŋtʃaguʃ	gejprathus	gimpajlhows	imbrohus	imprahus
8	3	nontʃaguʃ	kejbashus	gimpalows	imbraxus	imprahui

8	4	nowlʃabuʃ	gɹimashʊs	gimpalawʃ	ilbrahus	dimprahui
8	5	munʃambuʃ	gimashus	gimpahawst	ilvaʁuʃ	klowfrabõ
8	6	munʃaŋguʃ	jumasows	gimbahawst	dizahuʃ	klofabum
8	7	monʃaguʃ	jumasajs	imbahaws	dizahuʃ	kɹunambɹən
8	8	montʃabuʃ	imosajs	imbahaws	dizahuʃt	tumamboɹndɪ
8	9	montʃabus	ivosajs	ɪmbahaws	disawʃ	toroanbondɛd
8	10	oɹtʃambus	hirosvejs	embahawʃ	jusowʃ	tuandbondi
8	11	ɑɹdʒanpus	hirusves	imbahaws	jusowʃ	tuanbondi
8	12	ɑɹdʒampus	hilesves	ɛmbaws	jusowʃ	duɛnbondi
8	13	ɑɹdʒoŋgus	plisvesves	ɛnbaws	jusowʃ	duɛnhondɪ
8	14	aɹçaŋkus	plisvɛsvɛs	imbows	dʒufows	duebon
8	15	arjugus	dəsəsəs	ɛnbos	dʒʊfoɹs	duedoŋ
9	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
9	1	gimpɹalhus	gimpralħus	gimpralhus	mihabos	gimpralhus
9	2	gimpɹalhus	gimprahus	jʊtraləfʊs	pihabõs	impralhus
9	3	ɲimprawthus	klimplahus	jʊtalefeʃ	piŋkakuʃ	imprahurs
9	4	ɹuɹltamhus	klimplæmpluχ	jutalefejʃ	kuŋkakəʃ	imprabɹos
9	5	kukamkus	plimplæmplux	jukarpiʃ	komkakoʃ	χejbabu
9	6	kukamkus	plimplæmpluxta	jukalpitʃ	koŋkakoʃ	hejbabu
9	7	okago	linlænplutr	jukolpitʃ	kodʒakuʃ	hejbabun
9	8	owkaʔow	limlæmluʈa	jugoltitʃ	ozauʃ	hejbabu
9	9	owkraow	liŋlaŋloka	dʒurbarkitʃ	uʃaudz	sejbabu
9	10	howʔaow	miananoka	ubarkitʃ	iʃakɨz	sejbabu
9	11	howawow	miatnanoka	duparkitʃ	iʃakes	sejbopu
9	12	howəow	mianatnoka	dupoɽkitʃ	iʃekis	sejbupu
9	13	howawow	aploka	duborkitʃ	edʒekis	sejbupu
9	14	owawow	apnuka	duborkitʃ	ejekis	sejbʊpu
9	15	owbaow	apnoka	dubodrɪtʃ	iakis	subuku
10	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
10	1	gɹinɹaɹus	gimprahus	gimprahus	kybralgus	gimpralhus

10	2	gɹingɹagɹows	gimprahus	kimbrahus	kimpɹagows	timpalhus
10	3	gɹiwgɹawgɹows	biprasus	kimbrahus	kimprakus	timpalohuws
10	4	ɹiwɹawɹows	bibrasu	kimbahus	dibragtus	pimpalus
10	5	gɹewɹawɹows	livlasu	kiwanɛkuʃ	tepraktus	lɛpalows
10	6	dejoramorow	bliblasu	iwankytʃ	difɹatʃuf	klipalos
10	7	eoramoro	bliblasu	deonkətʃ	diprɛsəm	klipalos
10	8	eoramuo	liplasku	dʒunkutʃ	diplɛsɛm	klitaloz
10	9	euramor	nisbetsku	dʒukɛtʃ	iplɛsən	kikitapos
10	10	mewromoʃ	niʃpaskul	dʒukɛtʃ	piplesm	pikitabos
10	11	meromows	nisbasku	dʒukɛtʃ	bipeflum	pikitabos
10	12	mewuʃ	isbasku	dʒutɛtʃ	bipemflum	biɣitavos
10	13	mamuʃ	isbosku	dʒutɛtʃ	dipeflum	bihitahos
10	14	mamuʃ	isbosku	ʒutɛkʃ	distefɹum	bitovs
10	15	mamuʃ	ejsmoskju	dʒəɾɛʃ	distefron	bitows
11	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
11	1	gimpralhus	gimprahus	gimprahus	gimprawngus	gimprahus
11	2	ibrarus	gimprafus	diŋkankrus	gimprambus	gejmprahus
11	3	imbralus	dʒimprahus	viŋkaŋkus	gimprambus	gejmpɹahʊs
11	4	ibraluʃ	dimprəhuskwi	viŋkaŋkus	gimprambəs	dejpɹahows
11	5	ofalaruʃ	dinkrəbʉkwi	gɹinkalkus	gimbrombos	tejprahows
11	6	ʉbaluʃ	likrogwisbi	kalkus	imbrombros	deprahows
11	7	ʉdbalʉʃ	likrikwisprik	talkoks	imbrobro	denpawhaws
11	8	ibɹauʒ	ejpimispɹik	talkots	imbrubrump	tenpawhaws
11	9	ibals	ejbigmistɹik	tajgots	ejngoŋgoŋ	ibajhaw
11	10	ibls	bipimisri	ɹajdots	eŋkoŋkoŋ	ajgok
11	11	ivəs	pepemistɹi	dajdos	eŋkoŋkoŋ	ajkot
11	12	ivəs	pepemistɹi	dajdos	ejnskomkom	hajklo
11	13	ivows	efəɹmestri	ajdnos	dinsonklɑt	hajklow
11	14	iɹoθ	defɑɹmestri	teknas	dejnsonkrod	hajko
11	15	iɹowdz	ləfaɹməsti	ednas	diəsamkrɔnt	ajiŋgo

12	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
12	1	gimpralhus	himpagrus	gilprahus	gipɹagus	gimpralhus
12	2	dimprahus	tʃipalpus	kilbragus	gipɹagu	impralhus
12	3	libɹahʊs	sibabʉ	ibragus	dʒipɹagu	inkɑlɹus
12	4	librahos	sibabu	imbragwʊs	jepɹagu	ejtkawsgɹow
12	5	liəlvadhos	çidado	ənkrəkos	jipɹagu	ejtgatvow
12	6	wiɑɹhaɹħows	sidado	ʊkakos	jepragu	ejtkladlow
12	7	wihahaharhos	siudandon	lʊkakos	dʒepragu	ejklawsowh
12	8	wiahow	siudandon	kukakos	jevrako	ejklawdzo
12	9	wiɑɹow	iswopa	kukakotʃ	bærako	eglodzo
12	10	bihao	pliplopla	kukakotʃ	bærakɔ	eglozo
12	11	bihao	priplopla	kukakotʃ	mejrata	eglozo
12	12	bihanaw	tetʃotʃo	kokokotʃ	mejdzafa	jegotso
12	13	bihanaw	dejtsotsow	kutuklotʃ	mejapda	jolgoson
12	14	bianaw	dejtsotsow	gowtuklotʃ	lejwaka	tʃololsam
12	15	bianaw	deʈoʈow	gorklatʃ	lejwaka	tʃowlowsom
13	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
13	1	gimprahus	gimpɹahus	gimpralgus	ginpralhus	lintanħus
13	2	njytrahuʈ	gidahohi	gimpralhus	intrahuʃ	ɹintawxu
13	3	dʉtrahʉt	idaoħi	gemkrawkus	endahuʃ	lintahu
13	4	detahəts	idaoçi	kejmkɹawlus	indawuʃ	intahu
13	5	detɔhot	idahoki	kejnkɹawlus	indəwəʃ	intahu
13	6	itɑhot	itahoki	kejnkrɑwtut	indaratʃ	intaʔu
13	7	bitohoɖ	dipaħoki	kejnkrɑwtur	indaratʃ	intawu
13	8	ehoton	dibahoki	ejgɹspoʃ	ingorɔtʃ	intabu
13	9	jehotɔm	dibabowktej	ejdgɑrots	ingoratʃ	intavu
13	10	jetaŋ	dibabote	owbiqa	hiŋgonatʃ	intavu
13	11	jɛtoŋ	tibabokce	owbiat	hiŋglnatʃ	intaluf
13	12	dʒeton	ibabawtʃe	owviætʃ	iŋgəlnatʃ	intaldu
13	13	dʒetso	ibababawtʃe	owliedʒ	iŋgenjas	impalu

13	14	dʒɛtso	ibababawʃe	nowdʒejz	iŋgenjaʃ	imbalu
13	15	ɛtso	pirowbawʃəɹ	nowzdej	iŋgledaʃ	iŋbalo
14	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
14	1	ukahu	impralhust	giŋbawŋhus	gimpralhus	imprahus
14	2	ukahu	kipralhol	impagus	gimfrahus	embraws
14	3	ukakuk	kibraho	limpagus	dimfrahuʃ	albɹus
14	4	oŋkago	kibraho	mentagys	juswawus	ambɹʊs
14	5	okago	kibraho	mentəkes	jusowus	ambrus
14	6	okago	rihahow	bentʃəkes	jusowus	ambrus
14	7	ɹowkarow	ixaho	mantʃəke	dʒusowbus	ambres
14	8	bɹʉkarok	kjiaxho	najntʃəkje	dʒusowwus	ambres
14	9	blʉkaɹok	tiaxno	lantəte	jusowwul	ampres
14	10	biukaɹok	diaknaw	lantete	jusowbl	ampres
14	11	biukaɹɑk	diaknow	lanteteʃ	jusowbu	ambres
14	12	ʉkaɹɑ	diagnow	lanɟeɟetʃ	pjusowkow	ɑmbres
14	13	ʉkaɹa	nianow	lajnməməs	bisongoln	ɔmles
14	14	ukara	leanlʉ	ləwz	bisongow	ɑmplʊs
14	15	ukara	lejawlʉ	lows	imsomgɔ	okulus
15	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
15	1	dintahus	gimpralhus	ɹimbɹaho	dʒilpralbus	impralhus
15	2	inpahun	hunpɑnhus	gɹinaho	dʒirprawus	imprambus
15	3	iŋtao	heŋtʃaŋgeʃ	bimaho	dʒypraws	imprambo
15	4	indao	ɛntomnes	mimaho	dufraw	iŋkrambol
15	5	birao	ɛntomnets	mimaho	notkal	mkrambl
15	6	biʔaʔo	ɛnfʉnet	maho	matkɑt	kapu
15	7	miʔaʔow	enfane	mahowl	tskold	kapu
15	8	miʔaʔo	dinfane	maho	tskot	kapu
15	9	miʔaʔow	insoneŋ	maho	skowt	kapuçi
15	10	miʔaʔo	mensomej	mahow	skɔwt	kapuɦi
15	11	miʔaʔo	mensomen	maho	kow	kapuʃi

15	12	miʔaʔo	lɛnʃomæ	bahow	kul	kapuʃi
15	13	miʔaʔo	glenʃomæ	bahaw	kul	apuʃi
15	14	liʔaʔo	kenʃomɛɪ	bahaw	kurv	apuʃi
15	15	liʔaʔow	kentʃmɛ	ahaw	kuɪ	hafoksi
16	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
16	1	giŋpawwʉs	imprabus	itbawjuʃi	inprahus	gintraŋgus
16	2	gimpawɹʉs	grimpravus	bigmajuʃi	inprafus	mintaŋgows
16	3	impaus	grindabu	igɲuʃiʃi	implafus	nitango
16	4	impæls	etavjo	mijuʃiʃi	emplawse	ɹekotkom
16	5	impæls	edavij	wiuʃiʃi	plawse	ɹekokom
16	6	intal	edabjʊ	weɹoʃiʃi	plawse	xekokom
16	7	inkɑl	dindadjʊ	wejomʃiʃi	pawse	ekokow
16	8	iŋkɑl	diŋdaŋdu	bworiʃiʃi	awsen	ekokow
16	9	iŋkow	iŋgaŋgoŋ	goɽiʃiʃi	awsn	kokoa
16	10	iŋkomut	miŋgaŋgoŋg	goriʃiʃi	ɔlsn	boko
16	11	mejkow	miŋgaŋgoŋ	goriʃiʃi	koltsom	koko
16	12	mejkol	piŋkaŋgo	goriʃiʃi	koltsom	koko
16	13	mejkos	iŋkaŋgoŋ	boiʃiʃi	kowltiŋ	kɔkɔ
16	14	tsmejkaas	iŋkaŋgow	bojʃiki	hawedin	kokow
16	15	listɛnkaɪ	iŋkambo	utʃihi	hawedi	kowkow
17	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
17	1	ginpahus	intrahus	grimpragus	gimprahus	gimprahus
17	2	kiŋgahu	ejtrax	grimpragus	dʒiŋʔaʔus	dʒukɹahus
17	3	tʃiŋgahu	tra	grimbragos	dʒipaus	jokrahus
17	4	diŋgahu	ɹa	gimbragos	tʃiaus	jogabu
17	5	biŋgəhu	ɹaɪ	imbragus	sialgos	dodamu
17	6	biŋgau	fɹaɪ	kumbrakus	sialɣo	dawdanʉ
17	7	bindaw	rar	anrakutʃ	sialgo	dondav
17	8	bindaw	rak	anrakus	sialgo	doŋdaŋ
17	9	lindaw	rak	amɹakəs	sialgo	doŋdoŋ

17	10	lindaw	rak	anɹakos	sialgo	doŋdoŋ
17	11	iŋdaw	raks	anɹækas	sialgo	nambə
17	12	inda	raf	anwætast	sialgo	mambe
17	13	inda	rah	anwættas	tsialgo	mambeɹd
17	14	inda	ra	anwettas	dialgo	mambe
17	15	inda	ɹɑ	armis	dialgo	mambe
18	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
18	1	gimpralgus	gimbrahus	gimpɹælhus	gimpralhus	imprahus
18	2	kilfrɑʊs	vinblahus	imprævus	gimpʀahu	infraʔus
18	3	kifɹomwes	ɕinbləvuʃ	impravus	belkɑl	instraʔuʃ
18	4	dɹifɹamwes	ʃiblahʊs	impahus	belkɑ	mistraʔus
18	5	dɹifɹamɹejs	ʃinaʔus	umpakus	verʔɑ	mistraʔus
18	6	misombɹe	ʃinaʔɔs	uŋkahus	dɛpa	imistraʔus
18	7	isamwe	etʃimaʔɑ	luŋkahun	debɑm	emestrahus
18	8	nawej	zecimaʔaft	luŋkao	dejbo	imestratsus
18	9	nokwe	imaʔaft	loŋtao	dejbow	inedatʃus
18	10	notwa	imaʔaf	loŋtəɑw	tejbo	medatʃys
18	11	mopwa	deamaʔaf	lontaŋʔɑ	dɛjbol	mitatʃʉs
18	12	mopwap	deɹmaʔaf	boŋtaʔɔ	djejbo	metaʃuʃ
18	13	motɑ	denaʔas	ondaʔo	ipo	metadʒəs
18	14	bobo	benaʔas	oɲagow	ipo	metaʒəʃ
18	15	bowow	enaʔas	owlaʔow	iow	kopagjaʃ
19	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
19	1	gimpɹalhus	gimprahus	impajprus	gimpɹalhus	gimprabus
19	2	giŋfawsun	dʒimkrokus	viniŋpajkas	enjɔhus	giɹwaɹdus
19	3	njufowntʃy	dʒinkrokus	tiniŋbajpɑs	dedʒawu	dʉbadʉs
19	4	irfontu	dʒimkropus	tininwajpas	dedʒabu	dʉbadʉs
19	5	diontu	dinkrokos	miwajfas	dejabu	dɹubardus
19	6	dionto	tiŋtloŋsʉn	mijʊdefas	bejabu	ufagət
19	7	dʒohomtom	diŋgonstriəm	mijʊefas	plejabu	ofagoɹ

19	8	dʒohomtom	kitjænsɹ	mijʊɟefa	plejabu	ofagɹ
19	9	dʒohontoŋ	tʉkɑsəɹ	mindʒupdʒa	vejabu	owfagoɹ
19	10	johonto	dintɑstəɹ	meɹdeɹdudax	mejabəw	owbagloɹ
19	11	diotonkɹoŋ	dinkɑstəɹ	deɹdeɹdudaħ	bək	owbaglowɹ
19	12	biotonkɹoŋ	dikowstəɹ	dʒedʒedʒədʒa	lək	obagloɹ
19	13	fiontonkrom	dirkoster	dʒədʒədʒədʒa	ljuk	obaglor
19	14	fiontonkron	biekoste	tadʒaridʒa	ljuk	wobɹagloɹ
19	15	bijontonkɹon	bikosta	tɹadadidʒa	ljuk	wobræglor
20	0	grimpralhus	grimpralhus	grimpralhus	grimpralhus	grimpralhus
20	1	gimpralhuf	gimpralhus	grimbrahus	fwahows	gimprahus
20	2	binvahuf	gɹifajəs	grimkardhus	fraʧe	hilgasows
20	3	diŋvahuf	pipajəs	primpagus	fɹatse	hiodasus
20	4	tinvahuf	vifajeɹs	impagwə	gɹatse	biodaskuls
20	5	hʉhashʉ	vitfajləɹs	imbagwo	pɹətsik	biodasku
20	6	huhæsdu	skɹajn	imawo	heʧiks	biodasku
20	7	huhasdu	skrojn	ibawof	heʧik	iedasʔu
20	8	ħuħastu	skəjn	bigbabos	heʧi	jedashu
20	9	huhasdu	sklejn	bigbagbogbog	popʧi	jenasplus
20	10	huhɛstu	klejn	igbakokoa	bopʧi	jenasblus
20	11	huhestu	kwejn	idapitow	mokʧi	enwadslost
20	12	uesdu	pwejt	kintaŋkintoŋ	moʔʧi	dawarlos
20	13	wɛsdu	pleʈ	kintawkintow	mopʧi	imbaluaɹdos
20	14	esdu	plej	diŋdoŋg	moptsi	iŋgbaluargdos
20	15	distu	klej	diŋdoŋ	mɔpʧi	imbalwardos

Appendix 2. Model outputs for proportions of included sound parameters in the control- small, big-, pointy- and round-conditions.

Sound parameter	Condition	Fit	Lower	Upper
HIGH	CONTROL	0.230618656699477	-1.44600765240503	5.89810125366219
HIGH	SMALL	0.215307860182364	-3.59960064200368	5.27781288664448
HIGH	BIG	-0.429324801801926	-7.44853425963917	3.6145469850836
HIGH	POINTY	0.160035347019068	-3.71896312624912	5.14292958106367
HIGH	ROUND	0.205586858922572	-3.73886919920584	5.30514048194895
FRONT	CONTROL	-7.10997139459894	-16.078154127322	-0.0653874815210604
FRONT	SMALL	-13.6995009671562	-21.5792750892134	-6.04640779873769
FRONT	BIG	-3.94734923979815	-11.3975021214564	3.14690034162625
FRONT	POINTY	5.63033111236483	-1.94792308187174	12.4629525911913
FRONT	ROUND	-13.1962810641052	-20.7258959944684	-5.07764807620099
ROUNDED	CONTROL	2.10412984861119	-2.68700911516405	9.45676589246532
ROUNDED	SMALL	12.1885376922047	3.64115587393778	19.8238946356357
ROUNDED	BIG	0.0289810570395517	-7.01511685936072	7.72157972551947
ROUNDED	POINTY	-10.2964476802425	-16.9215588987712	-2.79524826579275
ROUNDED	ROUND	7.41252866818995	-0.336731530599999	15.2298722431993
GRAVE	CONTROL	-15.756488222744	-21.4013542377785	-10.5959133095439
GRAVE	SMALL	-8.07137606563015	-14.409759352628	-0.840858785745413
GRAVE	BIG	-15.0566313395363	-21.6081114724104	-9.25149273705352
GRAVE	POINTY	-14.0017662877704	-19.3259909710165	-8.2927395205946
GRAVE	ROUND	-7.10762246398289	-13.9379216570631	-0.670727152259605
VOICED	CONTROL	4.77359016450059	-0.0897539246172075	9.96759058046257
VOICED	SMALL	7.25511072671258	1.79133750356536	13.0132111813632
VOICED	BIG	4.03273103477755	-0.921432109624336	9.17931598723252
VOICED	POINTY	-4.44167880747909	-10.4039163885981	1.47397503546779
VOICED	ROUND	2.69896615471487	-2.8808916156315	8.09826945459457

SONORANT	CONTROL	-0.0835566394657477	-3.53409563787351	2.55174344512217
SONORANT	SMALL	-1.40536443840184	-7.29131826788958	1.70071455239954
SONORANT	BIG	-5.64054278646782	-11.7783349076275	0.0807908235741937
SONORANT	POINTY	-0.306888698644833	-4.6831638217908	2.9777086059265
SONORANT	ROUND	-0.266274423544893	-4.71102686489531	3.2528852889958

Appendix 3. Model outputs for proportions of included sound parameters in the small, big-, pointy- and round-conditions compared with the control-condition.

Sound parameter	Condition	Fit	Lower	Upper
HIGH	SMALL	-0.0158372121688544	-5.22494111352228	4.80672305376131
HIGH	BIG	-0.808793356097368	-8.62398201450861	3.084707554482
HIGH	POINTY	-0.00607306425445842	-4.88362484590099	5.06056663036227
HIGH	ROUND	0.0302836527710895	-2.67591384161046	5.45072363299348
FRONT	SMALL	18.7527168640107	8.30817859370361	27.8608111894988
FRONT	BIG	9.29950360253686	-1.18085739527952	18.58767658783
FRONT	POINTY	-0.578850725356183	-11.1495441929333	8.79977776628002
FRONT	ROUND	5.73914863972107	-1.99952472789698	15.1919997896965
ROUNDED	SMALL	-17.7847631647499	-26.963339736036	-7.43965971186916
ROUNDED	BIG	-7.40349066298248	-17.2297291450552	2.78594957374314
ROUNDED	POINTY	4.60534902124532	-5.66039084782107	14.8646268711208
ROUNDED	ROUND	-4.79664394458426	-13.6094126592691	2.95395728876415
GRAVE	SMALL	-6.96966978136323	-14.2891073266572	1.47463718748436
GRAVE	BIG	-7.8903248205153	-17.1401146467185	0.166798333208379
GRAVE	POINTY	-0.996415805129882	-9.21245352774291	8.75801669971933
GRAVE	ROUND	-8.69270568557871	-16.6309591029071	-0.53777084329238
VOICED	SMALL	-7.17571839521653	-14.1185889841863	0.159392159013844
VOICED	BIG	1.11855091906862	-4.53455577210316	7.74855238912453
VOICED	POINTY	4.46756899485332	-1.26613271204654	11.4818834393862
VOICED	ROUND	1.50763964798706	-2.18048088088113	8.50508504208834
SONORANT	SMALL	-0.0192545842130514	-4.5341276333049	4.27287968514856
SONORANT	BIG	-5.105602559082	-11.888138251758	0.630753049861079
SONORANT	POINTY	-0.863689646838814	-7.18329608668868	2.96510172372569
SONORANT	ROUND	0.0171630209289866	-3.30097130426665	4.45269176299176

Study III

Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-018-01639-7

Implicit associations between individual properties of color and sound

Andrey Anikin[1] · N. Johansson[2]

Abstract

We report a series of 22 experiments in which the implicit associations test (IAT) was used to investigate cross-modal correspondences between visual (luminance, hue [R-G, B-Y], saturation) and acoustic (loudness, pitch, formants [F1, F2], spectral centroid, trill) dimensions. Colors were sampled from the perceptually accurate *CIE-Lab* space, and the complex, vowel-like sounds were created with a formant synthesizer capable of separately manipulating individual acoustic properties. In line with previous reports, the loudness and pitch of acoustic stimuli were associated with both luminance and saturation of the presented colors. However, pitch was associated specifically with color lightness, whereas loudness mapped onto greater visual saliency. Manipulating the spectrum of sounds without modifying their pitch showed that an upward shift of spectral energy was associated with the same visual features (higher luminance and saturation) as higher pitch. In contrast, changing formant frequencies of synthetic vowels while minimizing the accompanying shifts in spectral centroid failed to reveal cross-modal correspondences with color. This may indicate that the commonly reported associations between vowels and colors are mediated by differences in the overall balance of low- and high-frequency energy in the spectrum rather than by vowel identity as such. Surprisingly, the hue of colors with the same luminance and saturation was not associated with any of the tested acoustic features, except for a weak preference to match higher pitch with blue (vs. yellow). We discuss these findings in the context of previous research and consider their implications for sound symbolism in world languages.

Keywords Cross-modal correspondences · Color · Synesthesia · Sound symbolism · Implicit associations test

Introduction

People have long been curious about why certain sounds and colors somehow “match.” Hearing a particular sound automatically and consistently produces a conscious experience of a particular color (Ward, 2013) in people with sound-color synesthesia. Non-synesthetes also often have strong intuitions about which sounds and colors go well together. It is a matter of ongoing debate to what extent such cross-modal correspondences share mechanisms with synesthesia (e.g., Lacey, Martinez, McCormick, & Sathian, 2016; Spence, 2011), but they certainly affect both perception and the way we talk about the world. For example, it seems natural to refer to high-frequency sounds as “bright,” although there is no *a priori* reason to associate visual brightness with auditory frequency. The pervasiveness of such metaphors emphasizes the importance of cross-modal correspondences not only for human perception but for language as well (Bankieris & Simner, 2015; Ramachandran & Hubbard, 2001; Sidhu & Pexman, 2018). Iconicity, or the motivated association between sound and meaning, has deepened our understanding of how human language and cognition evolved, as well as of how language continues to evolve culturally, by exposing several mechanisms that influence word formation and sound change. The concepts affected by lexical iconicity, or *sound symbolism*, generally have functions that relate to description or perception. Coupled with extensive perceptual evidence of cross-modal sound-color associations, this makes the names of colors good candidates both for finding evidence of sound symbolism (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Johansson, Anikin, Carling, & Holmer, 2018) and for relating it to potential psychological causes.

In the present article we address the psychological component of this problem by looking at how different color properties such as luminance, saturation, and hue are mapped onto acoustic properties such as loudness, pitch, and spectral

✉ Andrey Anikin
andrey.anikin@lucs.lu.se

[1] Division of Cognitive Science, Department of Philosophy, Lund University, Box 192, SE-221 00 Lund, Sweden

[2] Center for Language and Literature, Lund University, Lund, Sweden

Published online: 13 December 2018

characteristics. We begin by reviewing the extensive, but methodologically diverse and sometimes contradictory previous literature on sound-color associations and then report the results of our own experiments, in which we attempted to systematically test for cross-modal correspondences between linguistically meaningful acoustic features and individual perceptual dimensions of color.

It has long been known that people map auditory loudness onto visual luminance both in explicit matching tasks (Marks, 1974; Root & Ross, 1965) and in tests for implicit associations (Marks, 1987). There is some controversy surrounding the exact nature of matched dimensions that we return to in the *Discussion*, but in general, luminance-loudness associations are a straightforward example of so-called prothetic cross-modal correspondences that are based on the amount rather than the quality of sensory experience in two modalities (Spence, 2011). Loud sounds and bright colors share the property of being high on their respective prothetic dimensions and are therefore grouped together.

Pitch – the property describing how "high" or "low" a tonal sound appears to be – is reliably associated with luminance (Hubbard, 1996; Marks, 1974; Mondloch & Maurer, 2004; Ward, Huckstep, & Tsakanikos, 2006) and perhaps also with saturation (Hamilton-Fletcher, Witzel, Reby, & Ward, 2017; Ward et al., 2006). Unlike loudness, pitch is usually considered a metathetic rather than a prothetic dimension (Spence, 2011), in the sense that higher pitch is not "larger" or "greater" than low pitch, but qualitatively different. As a result, it is normally assumed that pitch is mapped onto sensory dimensions in other modalities, such as luminance, based on some qualitative correspondence between them. One complication is that some of the reported associations between pitch and color (Table 1) may have been caused by accompanying changes in loudness. The sensitivity of human hearing is frequency-dependent, and within the commonly tested range of approximately 0.2–3 kHz the subjective loudness of pure tones with the same amplitude monotonically increases with frequency (Fastl & Zwicker, 2006). It is therefore not enough to use stimuli normalized for peak or root mean square amplitude – the sound with the higher pitch may still be subjectively experienced as louder, introducing a confound. However, there is some evidence that the association of pitch with luminance (Klapetek et al., 2012), saturation, and hue (Hamilton-Fletcher et al., 2017) appears to hold even when the subjective loudness is held constant, indicating that cross-modal correspondences involving pitch are not entirely mediated by loudness.

Compared to the extensive research on color-loudness and color-pitch associations, there is less experimental evidence on how color is associated with spectral characteristics such as formants – frequency bands that are amplified by the vocal tract, creating different vowel sounds. In a large review of sound-color synesthesia spanning literally centuries of reports, Marks (1975, p. 308) concludes that certain vowels are reported to match different colors by synesthetes and non-synesthetes alike: [a] is associated with red and blue, [e] and [i] with yellow and white, [o] with red and black, and [u] with brown, blue, and black. More recent studies are largely consistent with Marks' summary (e.g., Miyahara, Koda, Sekiguchi, & Amemiya, 2012; Watanabe et al., 2014). The general rule appears to be that bright-sounding vowels, such as [i] and [e], are matched with bright colors, while dark-sounding vowels, such as [o] and [u], are matched with dark colors. The brightness of a vowel is sometimes said to be determined primarily by the second formant F2 (Marks, 1975), but in general raising the frequency of any formant tends to shift the balance of spectrum towards higher frequencies (Stevens, 2000). The center of gravity of a spectrum, also known as the spectral centroid, is a popular measure of the overall brightness or sharpness of musical timbre (Schubert, Wolfe, & Tarnopolsky, 2004), and an adjusted version of spectral centroid is used to approximate human ratings of sharpness in psychoacoustics (Fastl & Zwicker, 2006). Apparently, there is no direct evidence that the spectral centroid of complex tones with the same pitch is associated with visual luminance, but this effect is strongly predicted by the well-documented pitch-luminance associations and timbral consequences of raising the spectral centroid. There is also some experimental support for the idea that higher formants should be associated with greater luminance (Moos et al., 2014; but see Kim et al., 2017). An interesting unresolved issue is whether the association between formant frequencies and luminance is mediated by vowel quality or simply by the balance of low- and high-frequency energy in the spectrum. It seems intuitive that a vowel like [u] has an intrinsic "dark" quality that would not disappear by boosting high frequencies in the spectrum, but to the best of our knowledge, this assumption has not been tested.

There are also several reports linking formant frequencies to hue rather than luminance. Marks (1975) suggests that a high F2/F1 ratio is associated with green and a low F2/F1 ratio with red colors. Broadly consistent with this claim, Wrembel (2009) found that high front vowels, such as [i], were often matched with yellow or green hues. Furthermore, both synesthetes and non-synesthetes explicitly matched natural vowels with higher F1 to red rather than green in several experiments (Kim et al., 2017; Moos et al., 2014). Kim et al. (2017) report that yellow was associated with low F1 and high F2, although this relationship disappeared if they did not simultaneously vary the pitch of their synthetic vowels. Unfortunately, the presence of several confounds in most studies makes it difficult to determine what visual properties (hue, saturation, or luminance of the tested colors) were mapped to what acoustic properties (frequency of the first and second formants, F2/F1 ratio, or spectral centroid). In one of the most carefully controlled studies, Hamilton-Fletcher et al. (2017)

Table 1 Summary of previous reports of sound-color associations and our own data

Acoustic feature	Visual feature	Association	References	Our data	Proposed mechanism
Loudness	Luminance	Loudness ~ brightness	Bond & Stevens, 1969 Root & Ross, 1965	Loudness ~ darker gray on white background	Prothetic matching of loudness and visual salience
		Loudness ~ lightness (inconsistent, depending on background)	Marks, 1974, 1987		
	Hue	Loudness ~ orange/yellow (vs. blue)	Hamilton-Fletcher et al., 2017 Kim, Gejima, Iwamiya, & Takada, 2011 Menzel, Haufe, & Fastl, 2010	No association	Semantic matching§
		Loudness ~ red (vs. green)	Kim et al., 2011 Menzel et al., 2010		
	Saturation	Loudness ~ high saturation	Giannakis, 2001 Hamilton-Fletcher et al., 2017 Kim et al., 2011 Panek & Stevens, 1966	Loudness ~ high saturation	Prothetic matching of loudness and saturation
Pitch	Luminance	Pitch ~ luminance	Giannakis, 2001 Hubbard, 1996 Jonas, Spiller, & Hibbard, 2017 Klapetek, Ngo, & Spence, 2012 Ludwig, Adachi, & Matsuzawa, 2011 Marks, 1974, 1987 Martino & Marks, 1999 Melara, 1989 Mondloch & Maurer, 2004 Orlandatou, 2012 Ward et al., 2006 Watanabe, Greenberg, & Sagisaka, 2014	Pitch ~ lighter gray on white background	Metathetic matching of frequency and lightness
		No effect: pitch ~ visual contrast	Evans & Treisman, 2010		
	Hue	Pitch ~ yellow (vs. blue)	Hamilton-Fletcher et al., 2017 Hubbard, 1996 Orlandatou, 2012 Simpson, Quinn, & Ausubel, 1956	Pitch ~ blue (vs. yellow)	Semantic matching§
		No effect: pitch ~ blue (vs. red)	Bernstein & Edelstein, 1971		
	Saturation	Pitch ~ high saturation	Jonas et al., 2017 Ward et al., 2006	Pitch ~ high saturation	Prothetic matching of frequency and saturation§
Formants	Luminance	F1 ~ high luminance F1 ~ low luminance F2 ~ high luminance	Moos, Smith, Miller, & Simmons, 2014 Kim, Nam, & Kim, 2017 Kim et al., 2017 Moos et al., 2014	No association	Metathetic matching of frequency and lightness
		[i] [e] ~ bright colors [o] [u] ~ dark colors	Marks, 1975		
	Hue	F1 ~ red (vs. green)	Kim et al., 2017 Marks, 1975 Moos et al., 2014 Wrembel, 2009	No association	Semantic matching§
		F1 ~ blue (vs. yellow)	Kim et al., 2017		
		F2 ~ green (vs. red)	Moos et al., 2014 No effect in Kim et al., 2017		
		F2 ~ yellow (vs. blue)	Kim et al., 2017 Moos et al., 2014 Wrembel, 2009		

Table 1 (continued)

Acoustic feature	Visual feature	Association	References	Our data	Proposed mechanism
	Saturation	High F2/F1 ratio ~ green vs. red F1 ~ saturation F2 ~ saturation	Marks, 1975 Jakobson, 1962 cited in Moos et al., 2014	-	§
Other	Luminance	-	-	Spectral centroid ~ lighter gray Trill ~ darker gray	Metathetic matching of frequency and lightness Prothetic matching of visual and auditory saliency and/or metathetic matching of frequency and lightness
	Hue	Any power over 800 Hz ~ yellow (vs. blue)	Hamilton-Fletcher et al., 2017	Spectral centroid ~ blue (vs. yellow)¶	Semantic matching§
	Saturation	Spectral centroid ~ high saturation	Hamilton-Fletcher et al., 2017	Spectral centroid ~ high saturation	Prothetic matching of frequency and saturation§
		Noise vs. harmonic ~ low saturation	Orlandatou, 2012	- Trill ~ low saturation¶	§ Prothetic matching of visual and auditory saliency and/or metathetic matching of frequency and lightness

§ Uncertain mechanism

¶ Statistically marginal effect

discovered that the presence of energy above 800 Hz in the spectrum of complex synthetic tones was associated with yellow hues, even when participants were constrained to choose among equiluminant colors.

The key findings from the research on color-sound associations are presented in Table 1, with a particular emphasis on controlled experiments. Although by no means exhaustive, this summary highlights several contradictions and unresolved issues. Furthermore, many of the reported findings come from small studies with multiple potential confounds. In our opinion, the most significant progress in the field has been associated with three methodological advances:

1. *Controlling for visual confounds*. Until the last decade, researchers mainly worked with focal colors or approximations to the subjective color space, using contrasts such as light-dark or red-green. The recently pioneered use of perceptually accurate color spaces, such as *CIE-Luv* (Hamilton-Fletcher et al., 2017; Moos et al., 2014) and *CIE-Lab* (Kim et al., 2017), has the advantage of preserving subjective distances between colors while offering control over the separate dimensions of lightness, hue, and saturation. For example, there are several reports linking higher pitch to yellow (Orlandatou, 2012; Simpson et al., 1956). At the same time, focal yellow is also the brightest color (Witzel & Franklin, 2014), making it unclear whether yellow is associated with bright vowels because of its hue or because of its high luminance and saturation. By offering participants a choice among colors of the same luminance, Hamilton-Fletcher and co-workers (2017) demonstrated that yellow hues match higher frequencies in their own right, and not only because of their high luminance.
2. *Controlling for acoustic confounds*. Just as colors are defined by several perceptually distinct qualities, sounds have various acoustic properties that may contribute towards the discovered sound-color associations. The best-understood acoustic features are loudness and pitch, but speech-like harmonic sounds also vary in complex temporal and spectral characteristics such as formants, spectral noise, overall balance of low- and high-frequency energy in the spectrum, amplitude modulation, and so on. While loudness and pitch manipulations were already employed in early studies using synthetic white noise or pure tones (Marks, 1974; Root & Ross, 1965), modern techniques of formant synthesis enable researchers to create more naturalistic, speech-like sounds for testing. For example, Hamilton-Fletcher and co-workers (Hamilton-Fletcher et al., 2017) created a complex tone with several harmonics, the strength of which they could manipulate independently in order to change the spectral characteristics of their stimuli. Kim and co-authors (Kim et al., 2017)

went a step further and used articulatory synthesis to manipulate formant frequencies in vowel-like sounds. This is potentially a highly promising approach, but at present a number of challenges remain. For example, raising F1 or F2 has the effect of also boosting all frequencies above them (Stevens, 2000). In addition, manipulations of pitch and spectral characteristics can have a major effect on the perceived loudness of the stimuli. This is usually ignored (with a few exceptions, e.g., Hamilton-Fletcher et al., 2017 and Klapetek et al., 2012), but in view of the strong association between loudness and luminance it is desirable to make sure that the contrasted sounds are experienced as equally loud.

3. *Testing for implicit associations*. Until the mid-twentieth century, all evidence on color-sound associations consisted of reports by individuals, often synesthetes, who explicitly matched sounds with colors (reviewed in Marks, 1975). This method of subjective matching remains dominant in the field, but it primarily taps into what Spence (2011) calls the "decisional level," while it is also important to look for sound-color associations at a lower "perceptual level." Explicit beliefs about which color matches which sound are presumably grounded in low-level sensory correspondences, but they can also be influenced by cultural factors and personal history. Just as psychologists use implicit measures in order to study socially undesirable prejudices and biases, researchers of cross-modal correspondences have employed the speeded classification task (Ludwig et al., 2011; Marks, 1987), cross-modal Stroop interference (Ward et al., 2006), the implicit associations test (IAT; Lacey et al., 2016; Miyahara et al., 2012; Parise & Spence, 2012), the "pip-and-pop effect" (Klapetek et al., 2012), and other alternatives to explicit matching. Subjects do not have to be aware of possessing certain cross-modal correspondences for them to be detected in implicit tasks, and the results are less likely to be affected by cultural norms or idiosyncratic personal preferences.

We designed our experimental task with these three methodological considerations in mind. Like Kim et al. (2017), we sampled colors from the *CIE-Lab* space and created synthetic vowels. However, we used an adapted version of the IAT (Parise & Spence, 2012) instead of explicit matching. As argued above, implicit measures are more suitable for addressing cross-modal correspondences at a lower sensory level, which arguably holds the key to color-sound associations. In addition, with the IAT we had full control over the visual and acoustic characteristics of the contrasted pairs of stimuli, thus avoiding many confounds that arise in matching studies. Our pairs of colors differed only on one dimension at a time: luminance, saturation, or hue. In contrast, hue and saturation typically co-vary in matching studies, even if luminance is held constant (as in Hamilton-Fletcher et al., 2017). As for the acoustic stimuli, our ambition was to combine the rich spectral structure of the synthetic vowels used by Kim et al. (2017) with the careful matching of acoustic features achieved by Hamilton-Fletcher et al. (2017). We used formant synthesis to create natural-sounding vowels and manipulated one acoustic feature at a time to create six contrasted pairs; we also performed a separate pilot study to ensure that all stimuli were comparable in terms of subjective loudness.

The principal disadvantage of the chosen design was that only two pairs of colors and sounds could be compared in a single IAT experiment. A large number of participants therefore had to be tested in order to explore multiple combinations of stimuli, and even then it was impractical to determine whether the relationship between two features, such as pitch and saturation, was linear or quadratic (cf. Ward et al., 2006), based on absolute or relative values of the associated features (cf. Hamilton-Fletcher et al., 2017), etc. Because of this methodological limitation, we focused only on detecting the existence of particular cross-modal correspondences, not on their shape or robustness to variation in visual and auditory stimuli. We therefore made both visual and auditory contrasts in our stimuli pairs relatively large, well above detection thresholds. We also opted to collect data online, which allowed us to recruit a large and diverse sample of participants rapidly and at a reasonable cost (Woods, Velasco, Levitan, Wan, & Spence, 2015). Our goal was to investigate systematically, and using exactly the same experimental task, many of the previously described color-sound associations summarized in Table 1. Because in many cases the existing evidence comes from methodologically diverse studies and includes potential confounds, we did not formulate formal hypotheses to be tested, but simply looked for evidence of sound-color associations across a broad range of visual and auditory contrasts.

Methods

Stimuli

Visual stimuli were squares of 800 × 800 pixels of uniform color shown on white background. Pairs of colors were chosen so as to differ along only one dimension in the *Lab* space: luminance (*L*), hue (green-red [*a*] or yellow-blue [*b*]), or saturation (*sat*). Saturation was defined as the Euclidean distance to the central axis of the *Lab* space corresponding to shades of gray ($a = 0$, $b = 0$). The visual stimuli did not necessarily correspond to focal colors, but they were different enough to be easily distinguishable (Table 2).

The investigated acoustic features were chiefly selected based on the strongest previously reported evidence of sound-color correspondences such as loudness, pitch, and

Table 2 Contrasted pairs of visual stimuli

	L		a		b		Saturation	
Stimulus	⬛	⬜	🟩	🟥	🟨	🟦	🟩	🟩
Label	Dark gray	Light gray	Green[§]	Red	Yellow[¶]	Blue	Unsaturated green	Saturated green
Lab	25, 0, 0	75, 0, 0	50, -40, 45	50, 40, 45	70, 0, 40	70, 0, -40	70, -20, 20	70, -50, 50
RGB	59, 59, 59	185, 185, 185	66, 134, 33	193, 87, 43	194, 167, 98	117, 175, 243	147, 180, 134	98, 192, 73

§ Due to a mistake, in one experiment (F2 – green/red contrast) the colors slightly differed in saturation: green was *Lab* [60, -40, 40] and red [60, 60, 40]

¶ Bright, focal yellow is much lighter than any bluish hue, so to keep luminance constant we had to oppose blue to a bronze-like, dark yellow

spectrum. We also manipulated the frequencies of the first two formants, F1 and F2 – the two dimensions of the vowel chart – in order to connect the study more closely to natural speech sounds. In addition, the typologically most common trill, [r] (Mielke, 2004–2018; Moran, McCloy, & Wright, 2014), was also included due to its unique phonetic characteristics, such as its series of up to five pulses (Ladefoged & Maddieson, 1996, pp. 215–232), and because it has previously been found to be sound symbolically associated with the color green as well as words for movement and rotation (Johansson, Anikin, Carling, et al., 2018).

Acoustic stimuli were synthetic vowels created with *soundgen* 1.2.0*,* an open-source R package for parametric voice synthesis (Anikin, 2018). The voiced component lasted 350 ms, and the unvoiced component (aspiration) faded out over an additional 100 ms, so perceptually the duration was about 400 ms. The basic *soundgen* settings were shared by most stimuli and chosen so as to create a natural-sounding, gender-ambiguous voice pronouncing a short vowel. The fundamental frequency varied in a smooth rising-falling pattern between 160 and 200 Hz. Formant frequencies were equidistant, as in the neutral *schwa* [ə] sound (except when manipulated), and corresponded to a vocal tract length of 14 cm. Slight parallel formant transitions and aspiration were added to enhance the authenticity of stimuli. We opted to use diphthongs rather than static vowels for the contrasts that involved F1 or F2, so as to make the contrasts more salient. The manipulated formant moved up or down from a neutral *schwa* position, creating two different diphthongs.

As shown in Table 3, the spectral centroids of contrasted sounds with formant transitions were not exactly identical, but we did dynamically modify the strength of harmonics so as to achieve a relatively stable amount of high-frequency spectral energy and thereby mostly counteract the tendency for spectral centroid to shift in accordance with formant frequencies. In addition, to ensure that the subjectively experienced loudness of stimuli pairs would be as similar as possible (except when loudness was the tested contrast), the appropriate coefficients for adjusting the amplitude were estimated in a separate pilot study with five participants (Table 3, last column). All stimuli and R scripts for their generation can be downloaded from http://cogsci.se/publications.html together with the dataset and scripts for statistical analysis.

Procedure

We implemented a web-based html version of the implicit associations test (IAT) closely following the procedure described by Parise and Spence (2012). The task was to learn a rule associating the left arrow with one color and sound and the right arrow with another color and sound. Participants could examine the rule and hear the sounds for an unlimited amount of time before each block. For example, in one block of trials light gray/high pitch might be assigned to the left key and dark gray/low pitch to the right key. In the next block the rule would change, and all four possible combinations would recur in random order in multiple blocks throughout the experiment.

At the beginning of the experiment the participant was presented with instructions in the form of text and several slides followed by two blocks of 16 practice trials each. On the rare occasions when the accuracy was lower than the target level of 75%, practice blocks were repeated as many times as necessary. Once the participant had understood the procedure and achieved accuracy of 75% or better, they proceeded to complete 16 test blocks of 16 trials each.

As each trial began, a fixation cross was shown in the middle of the browser screen for a random period of 500–600 ms. After a delay of 300–400 ms the stimulus was presented. Color stimuli were shown for 400 ms in the same location as the fixation cross against a uniform white background; sounds also lasted about 400 ms. As soon as the stimulus disappeared or stopped playing, response buttons were activated and remained active until the participant had pressed the left/right arrows on the keyboard or clicked the corresponding buttons on the screen (the latter option was added for those participants who performed the experiment on a device without a physical keyboard). If the response was correct, the next trial began immediately. If it was incorrect, a red warning cross was flashed for 500 ms. Response

Table 3 Acoustic stimuli with the relevant soundgen settings

Manipulation	Contrast	Sound 1		Sound 2		Loudness equalization
		Key settings	Spectral centroid (Hz)	Key settings	Spectral centroid (Hz)	
Loudness	Two identical sounds, one 20 dB louder	Peak amplitude 0 dB	1,291	Peak amplitude -20 dB (1/10 of sound 1)	1,291	-
Pitch	Pitch difference of 1/2 octave	Low F0: 135-168-135 (-3 semitones)	1,252	High F0: 190-238-190 (+3 semitones)	1,242	-7.4 dB for low F0
F1	F1 either rises or falls 4 semitones from neutral	Rising F1: *formants = list (f1 = c(630, 790), f2 = 1900, f3 = 3160, f4 = 4430), rolloff = c(-8, -9)*[§]	1,384	Falling F1]: *formants = list (f1 = c(630, 500), f2 = 1900, f3 = 3160, f4 = 4430), rolloff = c(-8, -7)*[§]	1,463	-
F2	F2 either rises or falls 6 semitones from neutral	Rising F2: *formants = list(f1 = 630, f2 = c(1900, 2680), f3 = 3160, f4 = 4430), rolloff = c(-7.5, -9)*[§]	1,659	Falling F2: *formants = list(f1 = 630, f2 = c(1900, 1340), f3 = 3160, f4 = 4430), rolloff = c(-7.5, -6)*[§]	1,369	-1.8 dB for rising F2
Spectral centroid	Boosted vs. dampened high frequencies in source spectrum	Weak harmonics, dampened high frequencies: rolloff = -13	911	Strong harmonics, boosted high frequencies: rolloff = -3	2,170	-3.5 dB for high spectral centroid
Trill	Alveolar trill vs. no trill	~100 ms trill: [rə][¶]	1,443	No trill: [ə]	1,601	-5.8 dB for no trill

§ The "rolloff" parameter controls source spectrum, and it was dynamically adjusted to keep the amount of high-frequency in the spectrum relatively stable, since otherwise changing the frequency of F1 or F2 would have changed the overall spectral slope

¶ The trill was synthesized using amplitude modulation, F4 transitions, and rolloff modulation

See R code in the Online Electronic Supplements for implementation details

arrows remained visible on the screen throughout the trials, but they were active only during the response phase. The experiment lasted between 10 and 30 min, depending primarily on how quickly the participant mastered the procedure.

The screens and speakers used by participants were not calibrated, and in general we had no control over the devices that were used in the online experiment. However, the main variable of interest in this experiment was within-subject difference in response time and accuracy depending on sound-color pairing. As such, it was not essential for us to standardize the absolute physical characteristics of the presented colors and sounds, but only to preserve the relevant contrasts between stimuli pairs.

Participants

Participants were recruited via https://www.prolific.ac and reimbursed with £2–£2.5. They performed the study online, using a personal computer or a mobile device. All participants reported that they were fluent in English, had normal or corrected-to-normal vision, and had normal color perception. Submissions were discarded if they contained fewer than eight out of 16 complete blocks or if the average accuracy across all blocks was under 75%. A new sample of 20 participants was recruited for each of 22 experiments (N = 20 × 22 = 440 approved submissions, range 17–24 per experiment). Participants were not prevented from taking part in multiple experiments, so the total number of unique individuals across 22 experiments was 385 instead of 440. The mean number of completed test trials per participant was 253 out of 256.

Statistical analysis

All practice trials were discarded, and only test trials were analyzed (N = 111,532 trials). We worked with unaggregated, trial-level data and fit mixed models with a random intercept per target stimulus and a random intercept and slope per subject. The main predictor of interest was the rule for color-sound association in the current block. For example, in the luminance-loudness experiment light gray could be associated with the loud or quiet sound and assigned to the left or right key, for a total of four possible rules. However, there was no obvious side bias in response patterns, reducing four rules to two conditions: (1) light = loud, dark = quiet, and (2) light = quiet, dark = loud. The random intercept per target primarily captured the variance in accuracy or response time (RT) depending on the modality of the stimulus (e.g., response to visual stimuli was considerably faster than to acoustic stimuli). The random intercept per participant was included to account for individual differences in both accuracy and RT, which also accounted for possible differences in RT due to the chosen method of responding (with the keyboard, touchscreen, or mouse). Finally, we allowed the effect of condition to vary across participants by including a random slope per subject. Model comparison with information criteria suggested that the random slope improved predictive accuracy only in those experiments in which the congruence effect was weak and highly variable across participants (details not shown). Nevertheless, we included the random slope in all models, so as to keep them consistent and to be able to estimate cross-modal correspondences for each individual participant.

Two Bayesian mixed models of the same structure were fit for each experiment: a logistic model predicting accuracy and a log-normal model predicting RT in correct trials. Both models were fit in a Stan computational framework (http://mc-stan.org/) accessed from R using a *brms* package (Bürkner, 2017). We specified mildly informative regularizing priors on regression coefficients so as to reduce overfitting and improve convergence. When analyzing RT, we excluded all trials with incorrect responses (on average ~5%, no more than 25% per participant according to exclusion criteria) or with RT over 5000 ms (~0.3% of trials). To improve transparency, in Table 4 we report both observed and fitted values from regression models.

Results

The accuracy and speed of responding across all 22 experiments are summarized in Table 4. Accuracy was generally high, with the average error rate between 1% and 11% across experiments. RT in trials with a correct response was on average about 900–1,200 ms, which is slower than reported by Parise and Spence (2012). Since participants were instructed to achieve at least 75% accuracy, some may have prioritized avoiding mistakes at the cost of slowing down. In general, there is a trade-off between accuracy and speed in the IAT: some participants reveal their implicit associations by making more mistakes in the incongruent condition, while others maintain high accuracy but take longer to respond. We therefore looked for the effect of sound-color pairing on both accuracy and RT (Table 4). When both models showed significant differences in the same direction (i.e., both more errors and longer RT in condition 1 than in condition 2), that provided particularly clear evidence of non-arbitrary sound-color associations.

The findings are summarized graphically in Fig. 1, which also shows the distribution of average contrasts across participants. Higher luminance (light vs. dark gray on white background) was associated with lower loudness, higher pitch, higher spectral centroid, and the presence of a trill. The effect size for luminance was 3–4% difference in error rates and 60–120 ms difference in RT (Table 4). Congruency effects were revealed by both accuracy and RT, and were in the same direction for most participants. In contrast, there was no association between luminance and F1 or F2 frequency.

Neither green-red nor yellow-blue hue contrasts were reliably associated with any of the tested acoustic features, with one exception: high pitch was associated with blue (vs. yellow) hue (Table 4, Fig. 1). This effect was relatively small, but its confidence intervals excluded zero for both error rates (1.1% fewer errors, 95% CI 0–3.5) and response time (49 ms, 95% CI 10–96). In addition, a statistically marginal, but logically consistent congruence effect was observed between high spectral centroid and blue (vs. yellow) hue, again for both error rates (1.5%, 95% CI -0.1–4.5) and RTs (25 ms, 95% CI -3–59). The effect size for hue contrasts (0–1.5% and 0–50 ms) was thus about half of that for luminance contrasts. A few more marginal effects for hue-sound associations are shown in Fig. 1, but all of them were weak and manifested either in error rates or response times, but not both. We therefore do not consider them further.

Finally, high (vs. low) saturation was associated with greater loudness, higher pitch, and higher spectral centroid. In addition, the sound with a trill was weakly associated with low saturation based on the response time (43 ms, 95% CI 1–92), but only marginally so based on error rates (1.2%, 95% CI -0.4–4.4).

Discussion

In a series of experiments we used the implicit associations test (IAT) to investigate cross-modal correspondences between separately manipulated visual and acoustic features. This work extends previous research in two important ways. First, the majority of earlier studies relied on explicit matching, which quickly generates large amounts of data but operates at the relatively high “decisional level” (Spence, 2011) of consciously available beliefs. In contrast, implicit tasks like the IAT require more data, but they offer an insight into lower-level processing of perceptual input and thus provide a useful complementary perspective on sound-color associations. Second, we aimed to further refine the control over both visual and acoustic features, building upon several recent studies that employed perceptually accurate color spaces and sophisticated methods of sound synthesis (Hamilton-Fletcher et al., 2017; Kim et al., 2017). We created complex, vowel-like acoustic stimuli with a formant synthesizer, combining natural-sounding voice quality with precise control over formants, spectral envelope, intonation, loudness, and amplitude modulation. This enabled us to explore novel acoustic features in synthetic vowels, notably formant frequencies and spectral centroid, while avoiding several potential acoustic confounds. Visual stimuli were created using the *Lab* color space and

Table 4 Error rates and response times in 22 separate experiments

Acoustic contrast	Visual contrast	Rule	Error rate, %			Response time, ms		
			Observed (mean)	Fitted	Difference [95% CI]	Observed (mean)	Fitted	Difference [95% CI]
Loudness	L	Loud = light gray	6.2	4.8	3.9 [1.1–13.6]	1,451	1,268	128 [63–211]
		Loud = dark gray	1.2	0.9		1,190	1,140	
	a	Loud = red	4.2	3.1	0.4 [-0.6–2.2]	1,196	1,129	34 [3–71]
		Loud = green	3.5	2.6		1,157	1,094	
	b	Loud = yellow	4.4	3.9	1.6 [0.2–3.9]	1,105	1,042	21 [-6–54]
		Loud = blue	3.4	2.2		1,056	1,020	
	Sat	Loud = unsaturated	7.4	6.5	4.1 [1.9–8.5]	1,223	1,145	84 [39–137]
		Loud = saturated	3	2.4		1,113	1,061	
Pitch	L	High pitch = dark gray	8.3	6	3.2 [0.4–14.3]	1,201	1,137	64 [16–121]
		High pitch = light gray	4.4	2.8		1,153	1,075	
	a	High pitch = green	3.4	2.4	-0.3 [-2.3–0.6]	1,196	1,118	-10 [-37–17]
		High pitch = red	3.9	2.8		1,211	1,127	
	b	High pitch = yellow	5.2	3.9	1.1 [0.0–3.5]	1,358	1,212	49 [10–96]
		High pitch = blue	4	2.8		1,268	1,161	
	Sat	High pitch = unsaturated	9.9	7	4.9 [1.6–13.2]	1,416	1,296	108 [59–177]
		High pitch = saturated	4.7	2.1		1,259	1,188	
F1	L	High F1 = dark gray	11.6	9.3	0.5 [-2.6–4.6]	1,200	1,118	6 [-18–33]
		High F1 = light gray	11.5	8.6		1,200	1,112	
	a	High F1 = green	6.1	4.2	-0.1 [-2.4–1.8]	1,203	1,134	-22 [-59–9]
		High F1 = red	6.5	4.3		1,221	1,157	
	b	High F1 = blue	5.3	4.4	-0.7 [-3.6–0.8]	1,219	1,128	-8 [-37–19]
		High F1 = yellow	6.3	5.2		1,221	1,137	
F2	L	High F2 = dark gray	5.8	4.2	0.2 [-1.5–2.0]	1,164	1,103	16 [-16–48]
		High F2 = light gray	5.3	4		1,159	1,087	
	a	High F2 = green	4.1	2.5	-0.8 [-3.0–0.6]	1,291	1,092	-21 [-49–5]
		High F2 = red	4.7	3.4		1,168	1,112	
	b	High F1 = blue	3.6	2.3	-0.2 [-1.3–0.8]	1,151	1,071	-29 [-70–9]
		High F1 = yellow	3.8	2.5		1,160	1,100	
Spectrum	L	High freq = dark gray	7.6	5.9	4.0 [1./0 12.5]	1,287	1,203	83 [30–148]
		High freq = light gray	3	1.8		1,159	1,119	
	a	High freq = green	6.6	3.6	0.4 [-1.5–3.0]	1,017	959	-18 [-49–8]
		High freq = red	6.5	3.1		1,036	977	
	b	High freq = yellow	7.7	5.7	1.5 [-0.1–4.5]	1,169	1,114	25 [-3–59]
		High freq = blue	5.6	4.1		1,163	1,088	
	Sat	High freq = unsaturated	9.1	6.8	3.5 [0.5–10]	1,342	1,217	55 [9–109]
		High freq = saturated	6.1	3.1		1,261	1,163	
Trill	L	Trill = light gray	7	4.5	2.5 [0.6–8.2]	1,389	1,258	82 [34–146]
		Trill = dark gray	4.4	1.9		1,266	1,175	
	a	Trill = green	4.6	3.1	1.0 [-0.4–6.3]	1,111	1,052	9 [-15–32]
		Trill = red	3.3	1.9		1,100	1,043	
	b	Trill = blue	3.4	2.3	1.1 [0.2–4.6]	1,167	1,090	22 [-8–61]
		Trill = yellow	2.2	1.1		1,143	1,067	
	Sat	Trill = saturated	5.7	3.3	1.2 [-0.4–4.4]	1,244	1,183	43 [1–92]

varied along one dimension at a time (luminance, hue, or saturation). This experimental technique has the potential to pinpoint the individual visual and acoustic features driving cross-modal correspondences at a perceptual level. At the

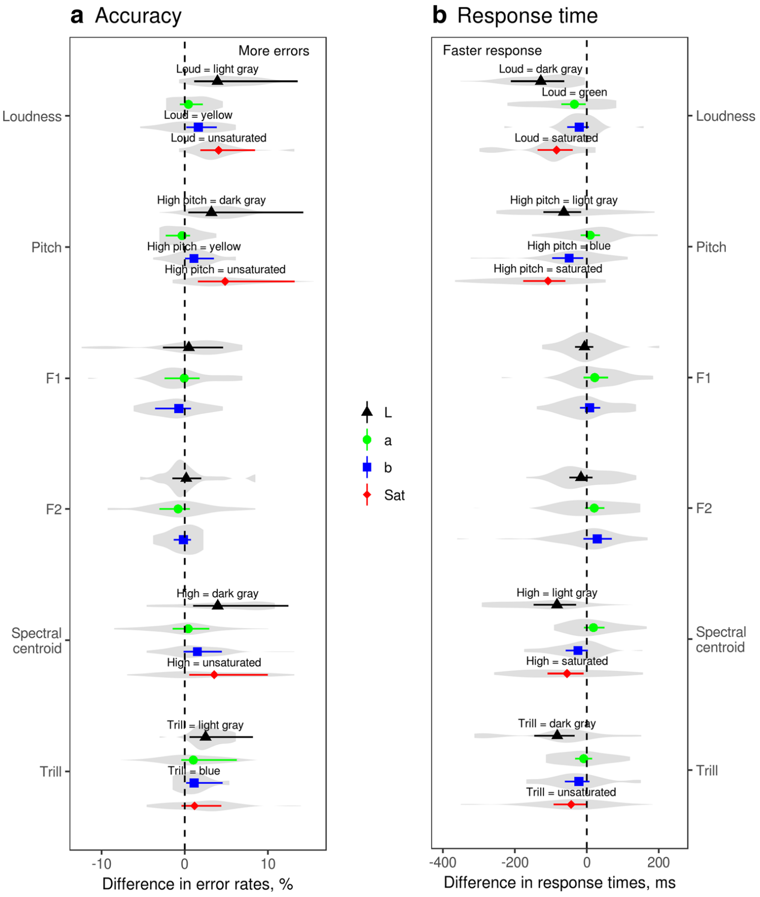


Fig. 1 Predicted difference in error rates (**A**) and response times (**B**) depending on the rule for pairing sounds and colors in 22 separate experiments. Solid points and error bars show the median of the posterior distribution and 95% CI. Labeled points have confidence intervals that do not overlap with zero. Violin plots show the distribution of observed values of the contrasts across participants (~20 per experiment, $N = 440$). L = luminance, a = green-red, b = yellow-blue, *Sat* = saturation

same time, the methodological differences between the current project and most previous research, particularly the use of an implicit outcome measure and complex, vowel-like sounds instead of pure tones, call for caution when directly comparing the results. In many cases our data confirm or nuance previous observations, but there are also several important differences, as discussed below.

In this study light gray was associated with low loudness and dark gray with high loudness, which seemingly contradicts the often reported association of visual luminance with auditory loudness (Table 1). However, the context in which stimuli varying in luminance are presented may strongly affect the result. The brightness of a physical source of light, such as a light bulb, seems to be unequivocally mapped onto the loudness of an accompanying sound (Bond & Stevens, 1969; Root & Ross, 1965). When the visual stimuli are patches of color, however, the way their lightness is mapped onto loudness depends on the background (Hubbard, 1996; Marks, 1974, 1987). When the background is darker than both stimuli, lighter colors are associated with louder sounds. When the background is intermediate in luminance between that of the stimuli, the association becomes inconsistent (Marks, 1974, 1987), unless the background is more similar in luminance to one stimulus than to the other (e.g., in Martino & Marks, 1999). The likely explanation is that luminance-loudness associations are driven by the amount of contrast between the stimulus and the background – more generally, by visual saliency (Itti & Koch, 2000) – rather than by lightness or luminance as such. In our experiment, visual stimuli (dark gray and light gray squares) were presented against a white background, making the dark stimulus more salient and therefore causing it to be associated with the louder of two sounds. It is also worth pointing out that the same effect was observed consistently for practically all participants (Fig. 1).

Interestingly, higher pitch was associated with light as opposed to dark gray, even though the association of dark gray with loudness indicates that the dark stimulus had higher visual saliency. This dissociation between pitch and loudness suggests that two different mechanisms are responsible for cross-modal correspondences between luminance and loudness, on the one hand, and luminance and pitch, on the other. We suggest that the luminance-loudness associations are prothetic (quantitative) in nature and driven by congruence in visual and auditory saliency, making them sensitive to contextual effects such as background color. In contrast, luminance-pitch appears to be a metathetic (qualitative) cross-modal correspondence. The same pattern was observed when both sounds had the same pitch and differed only in their spectral centroid: the sound with stronger upper harmonics and thus higher spectral centroid was associated with light versus dark gray. This is a novel finding in the context of research on sound-color associations, but it is fully in accord with the well-established fact that human ratings of timbral brightness or sharpness correlate closely with spectral centroid (Fastl & Zwicker, 2006; Schubert et al., 2004). We can thus conclude that lighter colors are mapped not only onto a higher pitch, but also onto an upward shift in spectral energy, even without a change in the fundamental frequency. This has important consequences for the likely interpretation of associations between formant frequencies and colors (see below). It is also worth reiterating that in our study the association between auditory frequency and luminance was not mediated by differences in perceived loudness since we normalized the stimuli for subjective loudness (as also reported by Hamilton-Fletcher et al., 2017).

Unlike luminance, saturation displayed the same pattern of association with loudness (loud = saturated) and with auditory frequency (high pitch or high spectral centroid = saturated). Hamilton-Fletcher and co-authors (Hamilton-Fletcher et al., 2017) suggest that the association between saturation and several acoustic characteristics – such as loudness, pitch, and spectral centroid – is based on ranking stimuli along each dimension from low to high, and therefore in essence these are prothetic cross-modal correspondences. This explanation is consistent with our results for saturation, since it was indeed associated with higher loudness, pitch, and spectral centroid, but this logic breaks down when applied to luminance. Since we established that the dark gray stimulus was the marked, more salient visual stimulus, we would expect dark gray to be paired with higher pitch if this association was prothetic. In actual fact, however, higher pitch was associated with a lighter (in this case less salient) color, as was also reported in numerous earlier studies (Table 1). One explanation is that auditory frequency can be compared to other modalities either qualitatively (higher frequency = lighter color) or quantitatively ("more" frequency = "more" saturation), perhaps depending on the existence and strength of pre-existing cross-modal correspondences. For example, if there is a powerful metathetic association of high frequency with lighter colors, it might override the weaker prothetic alignment of low-to-high visual saliency (which in this case was the reverse of lightness) with low-to-high frequency. Other explanations are certainly possible, and the exact cognitive mechanisms responsible for the observed cross-modal correspondences are yet to be elucidated.

Moving on to other findings, we did not observe any association between changes in the frequencies of the first two formants and either luminance or hue of the presented colors. We did not test for an association between formants and saturation, but it appears unlikely that there would be any. This null result contradicts a rich research tradition (Marks, 1975), according to which most informants agree which vowels best match which colors. However, natural focal colors differ not only in hue, but also in luminance and saturation. In more recent experimental research there have been attempts to use multiple regression (Moos et al., 2014) or palettes of equiluminant colors (Hamilton-Fletcher et al., 2017) to tease apart the contributions of these color dimensions, but even these better controlled studies did not distinguish between formant frequencies and the overall distribution of spectral energy. An increase in formant frequency not only modifies vowel quality, but also strongly shifts the spectral centroid upwards, which is in itself sufficient to make a sound "brighter" (Fastl & Zwicker, 2006; Stevens, 2000). We dynamically adjusted the spectrum of our synthetic vowels, largely – but not completely – eliminating the effect of formant transitions on the overall distribution of energy in the spectrum. The resulting diphthongs were easily distinguishable by listeners, as evidenced by the high accuracy in the IAT, but the relatively stable spectral centroid prevented the sounds with higher formants from sounding "brighter," canceling out an otherwise expected association between higher formants and higher luminance. Since we also demonstrated a clear association between spectral centroid and luminance, the logical conclusion seems to be that the often reported associations between formants and luminance are driven by the spectral consequences of formant transitions in natural vowels, not by formant frequencies per se. In other words, perceptually "bright" vowels, such as [i] and [a] (Johansson, Anikin, & Aseyev, 2018), probably owe their brightness to the fact that raising the frequency of individual formants (F2 for [i], F1 for [a]) shifts the balance of low- and high-frequency energy in the spectrum. If that is true, it should be possible to manipulate the perceived "brightness" of any vowel without changing its nature, simply by boosting or dampening higher frequencies in the spectrum, which can be verified in future studies.

One of the most surprising findings was the nearly complete lack of association between hue and any of the tested acoustic contrasts, with the possible exception of the relatively weak tendency to match higher pitch and higher spectral centroid with blue (vs. yellow) hue. It is possible that the effect size for hue was too small, falling below the sensitivity threshold of the experimental method. Alternatively, the previously reported hue-sound associations may only manifest themselves in the context of explicit matching. There is a considerable body of evidence, including a few studies that controlled for luminance (Hamilton-Fletcher et al., 2017; Moos et al., 2014; Kim et al., 2017), proving that informants consistently match hue to pitch, loudness, and formant frequencies. On the other hand, the weak IAT results suggest that hue may be associated with sound on a higher conceptual level through a mechanism that we tentatively labeled "semantic matching" in Table 1. For example, participants faced with a range of equiluminant colors might match high-frequency sounds with yellowish hues (Hamilton-Fletcher et al., 2017) by means of re-categorizing the available hues in terms of lexically labeled focal colors, so that the presented "yellowish" hue is treated as an approximation to the focal yellow, which would indeed be the best match due to its superior brightness. In an implicit task, however, this association disappears or can even be reversed, so that high pitch matches blue instead of yellow, as in the present study. Likewise, listeners may have relatively stable internal representations of different vowel sounds, so that [u] might be perceived and explicitly classified as "dark" and [i] as "bright" even if the stimuli are acoustically filtered, giving the [u] more high-frequency spectral energy. Although post-perceptual cross-modal correspondences have been observed with the IAT (Lacey et al., 2016), high-level, non-automatic, and relatively slow effects of this kind may manifest themselves more readily in explicit as opposed to implicit tests. This explanation is highly speculative, and our results will need to be replicated. But even with these provisos, the present findings clearly show that prothetic, low-to-high dimensions of color – luminance and saturation – dominate over hue in the context of implicit cross-modal matching.

The most acoustically complicated manipulation in the present study was to add rapid, trill-like amplitude modulation at the beginning of a syllable, leaving the other stimulus in the pair without a trill. While interesting from a linguistic point of view, this manipulation is difficult to interpret because it introduces two acoustic contrasts instead of one. The syllable with a trill is marked by virtue of containing an additional phoneme, but it also has a noticeably lower spectral centroid (Table 3). Listeners associated the trill with dark (vs. light) gray and, marginally, with low (vs. high) saturation. The association with luminance may be a case of prothetic matching of visual saliency (higher for dark gray) and acoustic saliency (higher for the marked syllable with a trill). Alternatively, this effect may be mediated by an association between spectral centroid (higher without a trill) and color lightness, which would also explain why the trill was associated with low rather than high saturation. Both of these effects may also be present simultaneously; in fact, summation of cross-modal correspondences has been shown experimentally (Jonas et al., 2017), and it may be a common occurrence in the real world, where objects have more than two sensory dimensions. This ambiguity showcases one of the problems facing cross-modal research, namely the inevitable tradeoff between the control over experimental stimuli and their ecological validity. It is also worth pointing out that, in contrast to some previous results (Johansson, Anikin, Carling, et al., 2018), we found no direct association between the trill and green-red contrast. On the other hand, linguistic studies of sound symbolism concern focal colors, which were not featured in the present study. Assuming that cool colors, such as blues and greens, are lower than warm colors in luminance and saturation, the presence of trills in words for the color green might still be sound symbolically charged, but this will have to be verified in future studies.

The study presented here has a number of other limitations. First of all, the chosen method of implicit associations required such a large sample that only a single pair of visual and acoustic stimuli could be tested within each condition. For example, "luminance" in the discussion above corresponds to the contrast between two shades of gray on the same white background, "pitch" represents a single, rather arbitrarily chosen contrast of six semitones, and so on. It remains to be seen how our conclusions will hold once a more diverse range of stimuli has been tested. Furthermore, online data collection entails certain methodological complications. For example, response times were on average about 1 s, which is slightly slower than in the study whose design we closely reproduced (Parise & Spence, 2012). One likely reason is that participants responded more slowly to the acoustic stimuli, which lasted 400 ms and in some conditions contained dynamic cues such as moving formants, making it necessary to hear the entire stimulus before even beginning to classify it. It is also possible that some participants were slowed down by using the mouse to click the response buttons instead of pressing keys on a physical keyboard or touching the buttons directly on the screen. An inability to standardize the equipment used by participants is one of the shortcomings of the present study, even though we could largely account for such variation by using a within-subject design and mixed models with a participant-specific intercept. A within-subject design is in general recommended in the context of online research, particularly when the outcome measure is device-dependent, as in the case of response time (Woods et al., 2015). Nevertheless, assuming that fast responses are relatively automatic, while slower responses are indicative of more extensive cognitive processing (Parise & Spence, 2012), it would be useful to replicate our results in a more controlled setting, ensuring that all participants pressed physical buttons and had less time for deliberation. This should make the estimates more precise and possibly reveal weaker cross-modal correspondences, for example, between loudness and hue or pitch and hue.

Taking a step back, the present method allowed us to study the interaction between perception, language, and cognition by isolating relevant visual and acoustic parameters without disconnecting them too much from natural speech sounds and the colors we perceive in the surrounding world. An important avenue for further research is to investigate how the discovered perceptual sound-color associations relate to sound symbolism in names of colors in natural languages. The mapping of high pitch and high spectral centroid on lighter colors is largely in line with previous cross-linguistic studies that have shown associations between [u] and concepts denoting darkness (Blasi et al., 2016). In a follow-up study (Johansson, Anikin, et al., 2018) we confirmed that both sonorous and bright vowels are strongly over-represented in the names of bright colors across world languages, while sonorous consonants are over-represented in the names of saturated colors. Interestingly, in the present study we observed implicit cross-modal correspondences for spectral centroid, but not formant frequencies (which define vowel quality), confirming that sound symbolism operates at the level of individual acoustic features rather than phonemes (Sidhu & Pexman, 2018). Together with other evidence of cross-modal correspondences on a basic perceptual level (Hamilton-Fletcher et al., 2017; Kim et al., 2017), the present findings also indicate that sound-meaning associations do not have to be mediated by orthography (cf. Nielsen & Rendall, 2011). A similar experimental approach can be useful for research on other audiovisual correspondences beyond the domain of color (Walker, 2012) as well as for research on other sensory modalities. Likewise, the differences between prothetic and metathetic mappings, as well as the fact that luminance and saturation were found to be the driving factors in sound-color mappings, add a further dimension to our understanding of how iconic associations are grounded and operate on semantic, phonetic, semiotic, and cognitive levels. Crucially, luminance, followed by saturation and the possible association of cool colors and trills, emerges as the primary visual component in color-sound symbolism, although its role should be further investigated in words of natural languages in order to connect cross-modal correspondences on a perceptual level with the development and change of lexicalization patterns and semantic boundaries across languages.

Conclusions

Using the implicit associations test, we confirmed the following previously reported cross-modal correspondences between visual and acoustic features:

- high loudness with high saturation,
- high pitch with high luminance,
- high pitch with high saturation,
- high spectral centroid with high saturation.

We propose to reinterpret the following associations:

- loudness with luminance: driven by visual saliency rather than color lightness, therefore reversed when more luminant stimuli are less salient,
- high formants with high luminance and saturation: driven by spectral shape rather than vowel quality, therefore no effect when controlling for spectral centroid.

We also report two purportedly novel associations:

- high spectral centroid with high luminance,
- alveolar trill with low luminance and low saturation.

Finally, none of the previously reported associations between hue and acoustic features were observed in the IAT, with the possible exception of a marginal and previously unreported tendency to match high pitch with blue (vs. yellow) hue.

References

Anikin, A. (2018). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavoir Research Methods*, 1-15. https://doi.org/10.3758/s13428-018-1095-7

Bankieris, K., & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition, 136*, 186-195.

Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241-247.

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.

Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*(6), 337-339.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1-28.

Evans, K. K, & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1):6, 1-12.

Fastl, H., & Zwicker, E. (2006). *Psychoacoustics: Facts and models, 2nd ed*. (Vol. 22). Springer Science & Business Media, Berlin.

Giannakis, K. (2001). Sound mosaics: A graphical user interface for sound synthesis based on audio-visual associations. Doctoral dissertation, Middlesex University, UK. Retrieved from http://eprints.mdx.ac.uk/6634/1/Konstantinos-sound_mosaics.phd.pdf

Hamilton-Fletcher, G., Witzel, C., Reby, D., & Ward, J. (2017). Sound properties associated with equiluminant colours. *Multisensory Research*, *30*(3-5), 337-362.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American Journal of Psychology, 109*(2), 219-238.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489-1506.

Jakobson, R. (1962). *Selected writings I. Phonological studies*. Gravenhage: Mouton & Co.

Johansson, N., Anikin, A., & Aseyev, N. (2018). *Color-sound symbolism in natural languages.* Manuscript in preparation.

Johansson, N., Anikin, A., Carling, G., & Holmer, A. (2018). *The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features*. Manuscript submitted for publication.

Jonas, C., Spiller, M. J., & Hibbard, P. (2017). Summation of visual attributes in auditory–visual crossmodal correspondences. *Psychonomic Bulletin & Review, 24*(4), 1104-1112.

Kim, H. W., Nam, H., & Kim, C. Y. (2017). [i] is lighter and more greenish than [o]: Intrinsic association between vowel sounds and colors. *Multisensory Research, 31*(5), 419-437.

Kim, K. H., Gejima, A., Iwamiya, S. I., & Takada, M. (2011). The effect of chroma of color on perceived loudness caused by noise. In *40th International Congress and Exposition on Noise Control Engineering 2011, 4* (pp. 3151–3156).

Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics, 74*(6), 1154-1167.

Lacey, S., Martinez, M., McCormick, K., & Sathian, K. (2016). Synesthesia strengthens sound-symbolic cross modal correspondences. *European Journal of Neuroscience, 44*(9), 2716-2721.

Ladefoged, P. & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *PNAS, 108*(51), 20661-20665.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American Journal of Psychology*, *87*(1-2), 173-188.

Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, *82*(3), 303-331.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384-394.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*, *28*(7), 903-923.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 69-79.

Menzel, D., Haufe, N., & Fastl, H. (2010). Colour-influences on loudness judgements. In *Proc. 20th Intern. Congress on Acoustics, ICA (2010), Sydney, Australia*.

Mielke, J. (2004–2018). *P-base. A database of phonological patterns*. http://pbase.phon.chass.ncsu.edu.

Miyahara, T., Koda, A., Sekiguchi, R., & Amemiya, T. (2012). A psychological experiment on the correspondence between colors and voiced vowels in non-synesthetes. *Kansei Engineering International Journal, 11*(1), 27-34.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *i-Perception*, *5*(2), 132-142.

Moran, S., McCloy, D. & Wright, R. (eds.) (2014). *PHOIBLE Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://phoible.org.

Nielsen, A., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale*, *65*(2), 115-124.

Orlandatou, K. (2012). The role of pitch and timbre in the synaesthetic experience. In *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki, Greece* (pp. 751-758).

Panek, W., & Stevens, S. S. (1966). Saturation of red: A prothetic continuum. *Perception & Psychophysics*, *1*(1), 59-66.

Parise, C. V., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research*, *220*(3-4), 319-333.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies, 8*(12), 3-34.

Root, R. T., & Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American Journal of Psychology, 78*(2), 285-289.

Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proceedings of the international conference on music perception and cognition, North Western University, Illinois* (pp. 112-116).

Sidhu, D. M., & Pexman, P. M. (2018). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review, 25*(5), 1619-1643.

Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies. *The Journal of Genetic Psychology, 89*(1), 95-103.

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

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). Cambridge: MIT press.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics*, *74*(8), 1792-1809.

Ward, J. (2013). Synesthesia. *Annual Review of Psychology*, *64*, 49-75.

Ward, J., Huckstep, B., & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex*, *42*(2), 264-280.

Watanabe, K., Greenberg, Y., & Sagisaka, Y. (2014). Sentiment analysis of color attributes derived from vowel sound impression for multimodal expression. In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific* (pp. 1-5).

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful"? *JOSA A*, *31*(4), A365-A374.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X., & Spence, C. (2015). Conducting perception research over the internet: a tutorial review. *PeerJ, 3*, e1058.

Wrembel, M. (2009). On hearing colours—cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studies in Contemporary Linguistics*, *45*(4), 595-612.

Study IV

Language and Cognition **12** (2020), 56–83. doi:10.1017/langcog.2019.35

Color sound symbolism in natural languages

NIKLAS JOHANSSON

Division of General Linguistics, Center for Language and Literature, Lund University

ANDREY ANIKIN

Division of Cognitive Science, Department of Philosophy, Lund University

AND

NIKOLAY ASEYEV

Institute of Higher Nervous Activity and Neurophysiology of RAS

(*Received* 15 *January* 2019 – *Revised* 21 *August* 2019 – *Accepted* 21 *August* 2019 – *First published online* 18 *October* 2019)

ABSTRACT

This paper investigates the underlying cognitive processes of sound–color associations by connecting perceptual evidence from research on cross-modal correspondences to sound symbolic patterns in the words for colors in natural languages. Building upon earlier perceptual experiments, we hypothesized that sonorous and bright phonemes would be over-represented in the words for bright and saturated colors. This hypothesis was tested on eleven color words and related concepts (RED–GREEN, YELLOW–BLUE, BLACK–WHITE, GRAY, NIGHT–DAY, DARK–LIGHT) from 245 language families. Textual data was transcribed into the International Phonetic Alphabet (IPA), and each phoneme was described acoustically using high-quality IPA recordings. These acoustic measurements were then correlated with the luminance and saturation of each color obtained from cross-linguistic color-naming data in the World Color Survey. As expected, vowels with high brightness and sonority ratings were over-represented in the words for colors with high luminance, while sonorous consonants were more common in the words for saturated colors. We discuss these results in relation to lexicalization patterns and the links between iconicity and perceptual cross-modal associations.

KEYWORDS: color sound symbolism, cross-modal correspondences, color words, lexical typology, lexicalization.

1. Introduction

Associations between sounds and meanings have been studied independently by linguists, psychologists, and cognitive scientists through a range of different methodologies. Yet, the results and conclusions drawn from this research show considerable overlap and a shared desire to further investigate underlying cognitive processes which cause these associations. Aiming to bridge the gap between the cognitive and linguistic levels of analysis, we focused on one particular domain – color – and investigated both perceptual cross-modal associations and sound symbolism in color words. Building upon previous psychological research, the experimental part of this project provided an insight into the cognitive mechanisms responsible for sound–color associations on a perceptual level, as reported elsewhere (Anikin & Johansson, 2019). This experimental work enabled us to formulate concrete hypotheses regarding the expected patterns of sound symbolism in color words. In the present study we report the results of testing these hypotheses using linguistic evidence from a large corpus of basic vocabulary (Johansson, Anikin, Carling, & Holmer, in press). We begin by introducing sound symbolism and sound–meaning associations in general, followed by a discussion of the domain of color and its cross-modal and sound symbolic associations across and within languages. We then investigate the link between perceived loudness (sonority) and brightness of phonemes with their relative frequencies in color word data from natural languages.

1.1. SOUND–MEANING ASSOCIATIONS

Languages include different sounds in their phonologies, and these sounds can be combined into an almost endless number of strings that make up words. The words, in turn, also change over time and are frequently replaced by other words due to areal contact. It has therefore been commonly held that the connection between what a word means and how it is pronounced is on the whole arbitrary (Saussure, 1916). However, an increasing number of studies have demonstrated that a motivated association between sound and meaning, known as ICONICITY in general and SOUND SYMBOLISM in regard to spoken words, is far from being a fringe phenomenon and plays a crucial role for our ability to understand human language. In particular, the growing availability of written language description has made it possible to conduct large-scale cross-linguistic comparisons, which have revealed numerous over-representations of sounds, primarily in what is considered basic vocabulary (Blasi, Wichmann, Hammarström, Stadler, & Christiansen, 2016; Johansson et al., in press; Wichmann, Holman, & Brown, 2010). Furthermore, for almost a century sound symbolism has been investigated experimentally in a range of semantic fields, which have often included semantically oppositional adjectival pairs (e.g., Diffloth, 1994; Newman, 1933; Sapir, 1929). Among the most notable studies, Köhler (1929), as well as a range of follow-up studies (e.g., Ramachandran & Hubbard, 2001), found that an overwhelming majority of participants prefer to pair words with voiced sonorants and rounded vowels (e.g., *bouba*) to rounded shapes, and words with unvoiced obstruents and unrounded vowels (e.g., *kiki*) to pointy shapes.

It has been suggested that iconicity may be associated with several functional and communicative benefits. Iconic forms, including nonsense words, may increase learnability, especially for children (Imai, Kita, Nagumo, & Okada, 2008; Imai & Kita, 2014; Lupyan & Casasanto, 2015; Massaro & Perlman, 2017; Walker et al., 2010). In addition, iconicity may emerge from arbitrary word forms and environmental sounds through transmission and interaction between language users (Edmiston, Perlman, & Lupyan, 2018; Jones et al., 2014; Tamariz et al., 2018). As a result, iconicity seems to play a functionally scaffolding role in language learning, and the cultural evolution of language helps to explain the biases that cause people's intuitive expectations about specific artificial language material such as *bouba* and *kiki*. However, these patterns generally seem to be much less consistent in natural languages (Nielsen & Rendall, 2012: 116–117), at least regarding whole-word iconic forms, which suggests that iconicity operates on a deeper, feature level rather than on a phoneme level. At the same time, linguistic evidence of sound symbolism is seldom sufficiently related to psychological research which could trace the phenomenon back to cognitive mechanisms (Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015, p. 611).

1.2. COLOR AS THE INTERFACE BETWEEN CROSS-MODALITY AND SOUND SYMBOLISM

The domain of color, which belongs to basic vocabulary, is estimated to have a relatively high frequency of use both currently and prehistorically (Haspelmath & Tadmor, 2009; Swadesh, 1971). It is also perceptually salient and confirmed to be sound symbolically affected (Blasi et al., 2016; Johansson et al., in press; Wichmann et al., 2010), which makes it a good candidate for bridging the gulf between cross-modal associations on a perceptual level and sound symbolism in natural languages. A considerable amount of psychological research has focused on color synesthesia – a type of cross-modal sensory integration in which stimuli in one sensory modality involuntarily and automatically cause experiences in another sensory modality (Ramachandran & Hubbard, 2001). For example, for some synesthetes, sequential concepts such as graphemes, days of the week, or numbers can appear to have specific colors. However, associating sounds with colors also seems to be widespread among non-synesthetes.

Color can be broken down into three main properties. LIGHTNESS or LUMINANCE is a measure of a color's reflection of light, SATURATION corresponds to a color's colorfulness, and HUE corresponds to a color's dominant reflected wavelengths, which in *CIELAB* color space can be further divided into the green-to-red *a** axis and the blue-to-yellow *b** axis. Cross-modally, several acoustic parameters have reliably been associated with visual luminance. Specifically, high auditory loudness and pitch have been shown to map onto luminance, and possibly also to saturation, using a range of methodological set-ups (Hubbard, 1996; Marks, 1974, 1987; Mok, Li, Li, Ng, & Cheung, 2019; Mondloch & Maurer, 2004; Ward, Huckstep, & Tsakanikos, 2006). High pitch has also been reported to be linked to specific hues such as YELLOW (Orlandatou, 2012), although this could be explained by the fact that YELLOW is the brightest color (Hamilton-Fletcher, Witzel, Reby, & Ward, 2017). Likewise, both synesthetes and non-synesthetes have been reported to associate specific vowels and vowel formant ratios with specific hues (Kim, Nam, & Kim, 2017; Marks, 1975; Miyahara, Koda, Sekiguchi, & Amemiya, 2012; Moos, Smith, Miller, & Simmons, 2014; Wrembel, 2009). Arguably, many of these associations stem from a more fundamental tendency to match bright-sounding vowels like [i] with bright colors, and dark-sounding vowels like [u] with dark colors (Mok et al., 2019). However, Cuskley, Dingemanse, Kirby, & van Leeuwen (2019) found that, while acoustic features of vowels predict sound–color mappings in Dutch-speaking synesthetes and non-synesthetes, phoneme categories (Dutch monophthongs) and grapheme categories (orthographical representations of Dutch vowels) were even more consistently associated with particular colors. This could suggest that categorical perception can shape how cross-modal associations are structured.

To clarify which acoustic parameters are associated with which visual parameters, we performed a series of Implicit Associations Test experiments (Anikin & Johansson, 2019) sampling colors from the CIELAB space to create contrasts on only a single visual dimension (luminance, saturation, or hue) and generating natural-sounding speech sounds with a formant synthesizer. The strongest perceptual associations were as follows: (1) high auditory salience (loudness, markedness) with high visual salience (contrast or saturation), and (2) high auditory frequency with visual lightness. In other words, color–sound associations appear to be dominated by quantitative (prothetic) cross-modal associations between sensory properties that vary along a single dimension with a natural low-to-high direction, such as loudness and luminance (Spence 2011). In contrast, we found less evidence of qualitative (metathetic) associations between qualitatively different or dichotomous aspects of color and sound. In particular, no associations were found between acoustic characteristics and hue when luminance and saturation were held constant.

Translating these findings and other psychological evidence (reviewed in Anikin & Johansson, 2019) into properties relevant for natural phonemes, we hypothesized that sound symbolism in color words would be manifested in the tendency for sonorant and high-frequency phonemes to be over-represented in the words for bright and saturated colors. Previous data on the occurrence of color sound symbolism in natural languages comes from a few large cross-linguistic studies. Despite the fact that these studies combined have investigated hundreds of lexical items in thousands of languages, color words have not been featured to any large extent. In an attempt to link the Eurasian language families genetically, Pagel, Atkinson, Calude, and Meade (2013) investigated basic vocabulary and estimated their lexical replacement rate and sound similarity based on their frequency in everyday speech. They found that, among the investigated colors terms, words for BLACK had similar phonetic forms across the featured families. Wichmann et al. (2010) also found that words for NIGHT were phonetically similar when comparing 40 basic vocabulary items in almost half of the world's languages. More recently, by comparing the sound patterns of larger samples of basic vocabulary across the majority of the world's language families, Blasi et al. (2016) found over-representations of rhotics in words for RED.

Color words also show similarities with the sound symbolic class of words referred to as IDEOPHONES, i.e., language-specific words which evoke and describe sounds, shapes, actions, movements, and other perception concepts. Ideophones are rather scarce, at least in Indo-European languages, but having a least some words for colors is considered to be a linguistic universal (Berlin & Kay, 1969; Kay & Maffi, 1999). Interestingly, in Korean there does not seem to be a clear functional boundary between ideophones and colors since they follow the same set of sound symbolic rules as described by Rhee (2019). Firstly, the visual dimensions of the Korean color words can be systematically manipulated by phonotactic processes and by changing phonological features. Furthermore, Korean color sound symbolism is highly productive since the processes can create new color-related words through derivation and coinage. This means that the five base color words, *hayah*- (하얗) 'be white', *kkamah*- (까맣) 'be black', *ppalkah*- (빨갛) 'be red', *phalah*- (파랗) 'be grue', and *nolah*- (노랗) 'be yellow', can be expanded to hundreds of color words through alterations between vowel harmony, consonant tensing, and morphological processes. Luminance can be altered by replacing a so-called positive/yang vowel with a negative/yin vowel. For example, *ppalkah*- (빨갛) means 'be red' but *ppelkeh*- (뻘겋) means 'be dark red'. Likewise, saturation can be reduced by replacing a tensed consonant (spelled with a double consonant) with a de-tensed version such as in *ppalkah-* (빨갛) ‘be red’ and *palkah-* (발갛) ‘be reddish’. Reduplication or extension of a color word via suffixation does not change visual dimensions of colors as such but alter the distribution of the color over a surface. For example, the reduplicated from of *pwulk-* (붉) ‘be red’, *pwulkuspwulkus* (불긋불긋), changes meaning to ‘be reddish here and there’ or ‘spotty red’. These examples illustrate just how elaborately sound–color mappings can operate within a linguistic system. Not only are several visual parameters and overall coverage coded but, crucially, they are mapped separately through vowel quality, consonant tensing or morphological processes. Albeit less explicit than the Korean structures, Semai also utilizes modifiable templates for ideophones that relate to the sensory experiences of color, odor, and sound (Tufvesson, 2011). By changing the vowel of Semai color words their luminance level is altered, e.g., *blʔik* ‘gray’ vs. *blʔak* ‘black’. In addition, the replacement of a vowel can also change the meaning of color words to lighter–darker version of specific hues, e.g., *blʔɛk* ‘rust-brown’ vs. *blʔɨk* ‘darker rust-brown’, and *blʔuk* ‘dark purple’ vs. *blʔɔk* ‘darker purple’. Furthermore, Westermann (1927) showed that several West African languages also display contrasts between ‘light’ and ‘dark’, by altering vowels (unrounded front vs. rounded back) and tone (high vs. low). A similar tonal distinction between ‘black’, ‘green’, and ‘blue’ (high tone) vs. ‘red’, ‘yellow’, and ‘brown’ (low tone) is also found in Bini (Wescott, 1975).

Based on these cross-linguistic and language-internal findings, degrees of lightness and saturation do not only produce cross-modal associations in experimental set-ups but may also be sound symbolically charged in natural languages. It is also possible that specific hues, such as RED and GREEN, produce independent sound symbolic associations. The findings from Korean also suggest that color sound symbolism might be carried by both vowels and consonants, although vowels seem to be primarily tied to luminance and consonants to saturation. It is less clear, however, whether particular phonemes or vowel formants and dimensions such as spectral energy and manner and place of articulation of consonants are driving these effects. Based on the observed cross-modal correspondences of color luminance and saturation with acoustic loudness and frequency (Anikin & Johansson, 2019), we hypothesized that sound symbolism in color words would be manifested in a tendency to find sonorous and bright phonemes in the words for bright and saturated colors. To test this hypothesis, we obtained color words from a large sample of unrelated languages, which represented a large portion of the world’s language families (Johansson et al., in press), and analyzed the sonority and brightness of phonemes in these words. We obtained sonority ranks from an earlier study (Parker, 2002) and investigated the acoustic correlates of perceived phonemic brightness in a pilot study. Having these measures of the perceived sonority and brightness of individual phonemes, we then proceeded

https://doi.org/10.1017/langcog.2019.35

to test whether these acoustic properties varied systematically in the words for different colors as a function of their luminance and saturation.

2. Method

2.1. PERCEIVED BRIGHTNESS AND LOUDNESS OF PHONEMES

The acoustic properties that seem to have the greatest effect in sound–color correspondences are loudness and frequency (Anikin & Johansson, 2019). However, while loudness is often used pragmatically throughout languages, it is not used phonemically, i.e., there are no minimal pairs which are distinguished only by their level of loudness. Thus, we needed to find a suitable proxy for loudness which is also generally utilized by most languages.

Sonority, or perceived loudness, is among the most salient properties of phonemes cross-linguistically. One way of estimating subjective loudness is provided by the relative sonority ranks of different phonemes. Physically, sonority relates to sound intensity, which in turn depends on obstruction in the vocal tract. For example, the open vowel [a] is among the most sonorous sounds since it creates the least obstruction for the air to pass through the vocal tract. Stops, on the other hand, produce a great amount of obstruction and are thus found at the bottom of the sonority hierarchy. Phonologically, the sonority hierarchically determines the syllable structure in languages: the most sonorous sounds are placed in the nucleus and the less sonorous sounds at the end of the syllables. There are various suggestions for relative sonority ranks, but Parker (2002) provides one of the most thoroughly investigated classifications. Parker's sonority hierarchy is based on universal sonority patterns, language-specific effects, and acoustic, aerodynamic, and psycholinguistic factors. By measuring several acoustic and aerodynamic correlates of sonority in English and Spanish, Parker found a strong correlation between intensity and sonority indices which confirmed the physical reality of the sonority hierarchy. He then performed a psycholinguistic experiment which involved hundreds of native speakers evaluating 99 constructed rhyming pairs, e.g., *roshy–toshy*. The results showed that pairs which obeyed the sonority hierarchy were generally preferred by the participants and thus, despite minor phonological variations, confirmed the importance of sonority in language processing. The 16 sonority levels can be treated as equidistant since more precise indices are too variable for practical use (Figure 1).

Similarly to perceived loudness, we were also interested in testing the hypothesis that words for more luminant colors would contain phonemes that are perceived to be relatively 'brighter' since luminance and brightness have been shown to have a number of cross-modal correspondences (Ludwig & Simner, 2013; Walker, 2012; Walker et al., 2010). However, this task faces a

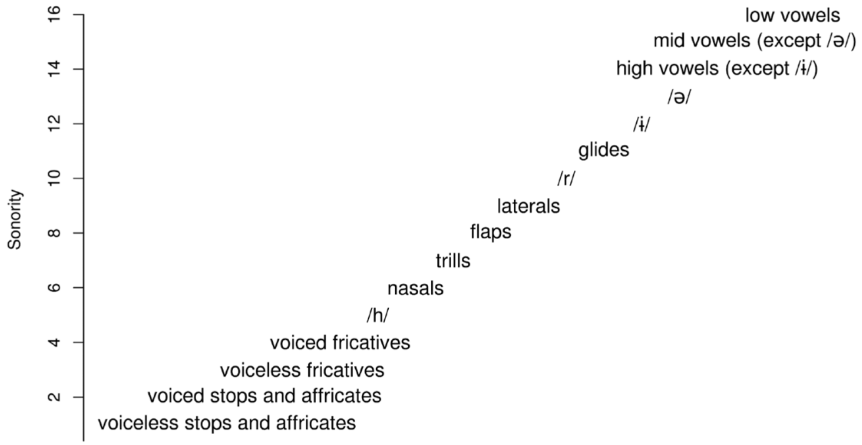


Fig. 1. Adapted version of Parker's (2002) sonority hierarchy.

methodological difficulty: it is not obvious which acoustic measure could serve as a proxy for the perceived 'brightness'. A related concept in psychoacoustics is 'sharpness', which is calculated as a weighted centroid of the cochleogram – that is, the spectrogram transformed into physiologically appropriate frequency and amplitude scales – followed by one of several empirically derived adjustments (Fastl & Zwicker, 2006). There is also some evidence that the ordinary, unadjusted spectral centroid is a reasonable predictor of the perceived brightness of the timber of musical instruments (Schubert, Wolfe, & Tarnopolsky, 2004), but it is not clear whether this is also the case for phonemes, including both vowels and consonants. While most measures of spectral central tendency are positively correlated, they are not identical, and the choice of one or the other may have a major effect on how phonemes will be ranked by relative brightness. Given this uncertainty about the most appropriate spectral descriptors that would capture the perceived brightness of individual phonemes, we performed a separate pilot study to check which acoustic features would best predict human ratings of brightness. Recordings of isolated phonemes – vowels, synthetic vowels, and consonants – were rated on brightness in three small experiments, after which the audio files were analyzed acoustically, and various spectral measures were correlated with the empirically obtained brightness ratings.

2.1.1. *Pilot: materials and methods*

We obtained audio recordings of 99 IPA phonemes from *Seeing Speech* (Lawson et al., 2015), a database which provides accurate recordings of a range of IPA symbols by an adult female speaker along with the movements of vocal organs visualized through ultrasound, Magnetic Resonance Imaging, and animation. The recordings were truncated to individual phonemes (e.g., by removing the extra vowels after plosives), normalized for peak amplitude in Audacity (https://www.audacityteam.org/), and analyzed acoustically. In particular, several measures of spectral shape were calculated on a linear or Mel-transformed scale: spectral centroid, or the spectral center of gravity; spectral median, or the 50th percentile of spectral energy distribution; and peak frequency, or the frequency of maximum power. For vowels, formants F1–F6 were estimated manually from a spectrogram.

Three samples of 19–22 phonemes were then rated on brightness in three perceptual tests: (1) 22 vowels; (2) 22 synthetic versions of the same vowels, which were created with the formant synthesizer *soundgen* (Anikin, 2019) using frequencies of formants F1–F6 manually measured in the original recordings, and standard values of other acoustic parameters, to remove unwanted accidental variation in the voice quality of recorded vowels (duration, intonation contour, amplitude envelope, the strength of harmonics, etc.); and (3) 19 consonants chosen as representative of the overall variety of consonants in the IPA system. Participants were shown badges with IPA symbols in a scrambled order and had to arrange them along a horizontal Visual Analogue Scale (VAS). The instructions were to arrange the sounds "from 'low' to 'high' – darkest-sounding on the left, brightest-sounding on the right". The sounds could be heard repeatedly by clicking the badges, which were dragged and dropped onto the scale and could be rearranged until the participant was satisfied. The experiment was written in HTML/JavaScript and could be performed in any modern web browser. Participants were predominantly native speakers of Swedish recruited via personal contacts; most completed all three tasks. Each experiment resulted in relative brightness ratings of the tested phonemes. These ratings were normalized to range from 0 to 100 for each participant, and then the average brightness of each phoneme was calculated as the median value across all participants. In a few cases individual brightness ratings correlated poorly with the median values (Pearson's $r < .2$), resulting in the exclusion of four out of 37 submissions and leaving 11 submissions for vowels, 11 for synthetic vowels, and 12 for consonants.

2.1.2. *Pilot: results*

There was moderate agreement between raters about the relative brightness of phonemes: intra-class correlation coefficient (ICC) was .55 for vowels, .66 for synthetic vowels, and .41 for consonants. Most common measures summarizing the shape of spectrum positively correlated with median brightness ratings (Table 1), but the highest overall correlation across all

TABLE 1. *Pearson's correlation of median brightness ratings with various measures of spectral shape and sonority, from highest to lowest*

Acoustic predictor*	Vowels	Synthetic vowels	Consonants	Mean**
Spectral centroid	0.76	0.90	0.92	0.86
Mel-centroid	0.77	0.78	0.94	0.83
F3	0.76	0.67	–	0.72
Spectral median	0.55	0.51	0.91	0.65
Mel-median	0.44	0.36	0.87	0.56
F2	0.24	0.65	–	0.45
F1	0.47	0.37	–	0.42
Peak frequency	–0.03	0.33	0.81	0.37
F4	0.24	0.27	–	0.25
Loudness, sone	0.34	0.36	–0.56	0.22
Sonority	0.38	0.06	–0.55	–0.04
F5	0.00	0.00	–	0.00
F6	–0.08	–0.13	–	–0.10
Root mean square amplitude	–0.06	–0.22	–0.66	–0.31

NOTES: * All acoustic predictors except sonority were log-transformed prior to correlating them with brightness ratings. *Spectral centroid* refers to the center of gravity of spectrum, while *spectral median* refers to the 50th percentile of the distribution of spectral energy. Formants were measured manually; all other descriptors were extracted with the *soundgen* R package (Anikin, 2019) per frame and summarized by their median value over the entire sound; ** the mean of three correlations: for vowels, synthetic vowels, and consonants.

three groups of phonemes was achieved by the spectral centroid calculated on a linear frequency scale: r = .76 for vowels, .90 for synthetic vowels, and .92 for consonants (Figure 2). Converting the spectrum to a physiologically more appropriate Mel frequency scale did not noticeably improve the correlation with brightness. Spectral medians performed considerably worse than spectral centroids. Interestingly, peak frequency was associated with perceived brightness in consonants, but not in vowels, presumably because in voiced sounds peak frequency traced the fundamental or one of the lower harmonics to the exclusion of the perceptually relevant higher parts of the spectrum.

Considering that vowels are distinguished above all by the frequencies of the first two formants, it was important to check the effect of formant frequencies on perceived brightness. Formants F1–F4, and particularly F2–F3, strongly correlated with brightness ratings of both real and synthetic vowels (Table 1). F1–F3, but not F4, were also positive predictors of the perceived brightness of vowels in multiple regression (Figure 3), suggesting that each of the first few formants makes an independent contribution towards making the sound 'bright'. As a result, both open vowels (high F1) and front vowels (high F2) were on average rated as brighter than closed and back vowels (Figure 3A, 3B). Interestingly, although F3 varied within a

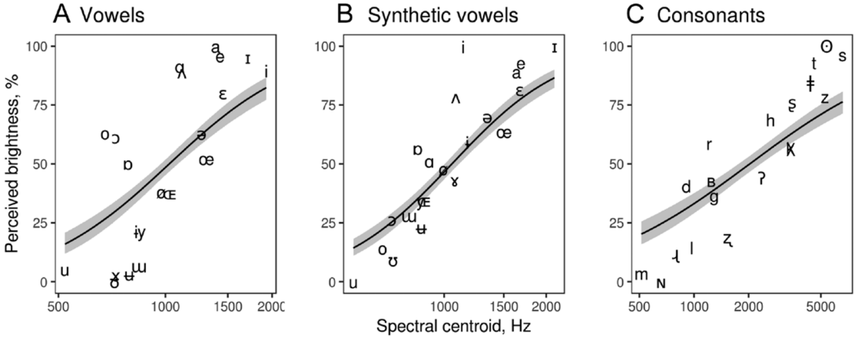


Fig. 2. Perceived brightness of isolated vowels (A), synthetic vowels (B), and consonants (C) as a function of their spectral centroid. Medians of the observed ratings are plotted as IPA symbols. The solid line and shaded area show fitted values from beta-regression with 95% CI.

narrower range than F1 and F2, it had a disproportionately large effect on perceived brightness, probably because F3 is located in the range of frequencies to which humans are particularly sensitive (for this speaker, F3 varied from 2.5 to 4 kHz).

An even stronger association between formants and brightness was observed for synthetic vowels (Table 1, Figure 3), which differed only in formant frequencies and were free from other phonetic confounds such as variation in amplitude envelope, pitch, and vocal effort. An upward shift in formants, with no other changes in pitch or voice quality, is thus sufficient to make a vowel sound brighter. Higher formants (F4–F6) did not contribute much to perceived brightness, presumably because harmonics above F3 were relatively weak. For sounds produced at higher intensity, there would be more energy in high frequencies, possibly making formants above F3 more salient and giving them a greater role in determining the perceived brightness. But in any case, spectral centroid appears to be a more robust measure than the frequencies of individual formants, and it is equally applicable to both vowel and consonant sounds.

To summarize, the pilot study demonstrated that perceived brightness is, for vowels, dependent on upward shifts of the first three vowel formants, and that spectral centroid seems to be the best proxy of perceived brightness that can be used for both vowels and consonants. Consequently, we predicted that perceived sonority and brightness (operationalized as the actual ratings from the pilot study or acoustic proxies like spectral centroid) of phonemes in color names would correlate with color luminance, and possibly also with saturation. Furthermore, we also wanted to know if these effects were tied specifically to vowels or consonants or both.

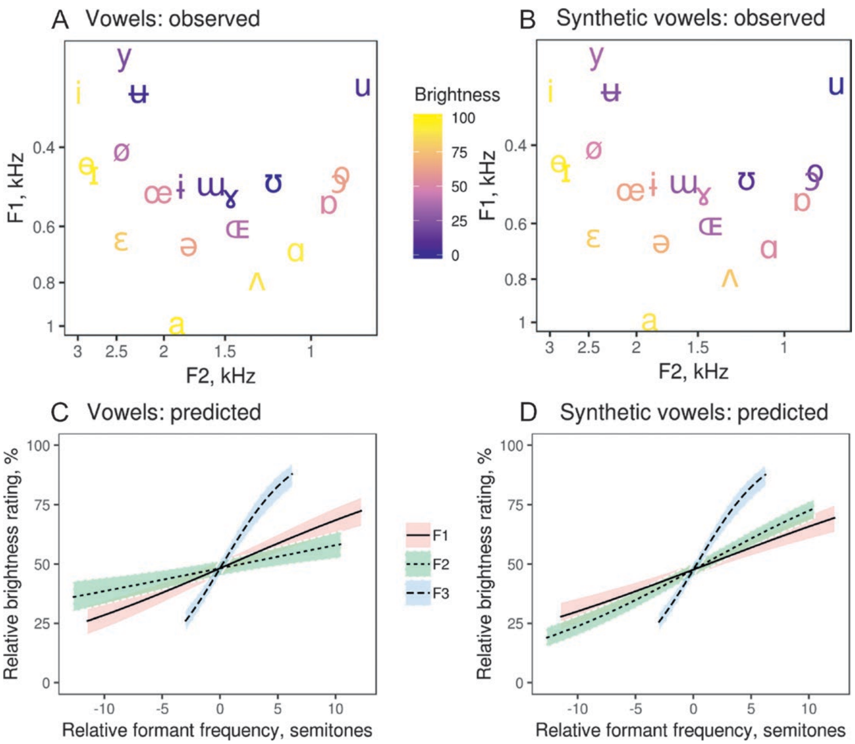


Fig. 3. The observed brightness ratings from highest (marked with yellow) to lowest (marked with blue) of real (A) and synthetic (B) vowels in F1–F2 space and the predicted effect of varying the frequency of one formant on the perceived brightness of real (C) and synthetic (D) vowels while keeping other formants constant at their average values. Fitted values from a beta regression model and 95% CI.

2.2. DATA SOURCES

The language data in text form was gathered from the corpus compiled by Johansson et al. (in press) for a cross-linguistic examination of 344 basic vocabulary items in 245 language families. Description concepts, including color words, constituted a large proportion of the items and appeared to be among the domains most affected by sound symbolism. Studies on the semantic typology of color words have generally only considered mono-lexemic words, as in Berlin and Kay's (1969) famous study. However, since many mono-lexemic terms can be traced back to natural referents, e.g., GREEN from 'to grow' or 'leaf', the color concepts selection was based on color opponency and included RED–GREEN, YELLOW–BLUE and BLACK–WHITE, as well as the combination of the most basic colors, GRAY. These concepts were also used for the present study, along with four other oppositional concepts that are semantically related to light, namely NIGHT–DAY and DARK–LIGHT.

The language sampling for the investigated concepts was very restrictive. Firstly, the language family database *Glottolog* (Hammarström, Forkel, & Haspelmath, 2017) was used due to its cautious approach to grouping languages into families. Secondly, only a single language per language family was selected in order to exclude any possible genetic bias. The sample included approximately 58.5% of the world's documented living and extinct language families without considering artificial, sign, unattested, and unclassifiable languages, as well as creoles, mixed languages, pidgins, and speech registers. The data was collected from various sources, such as databases, dictionaries, grammar descriptions, and grammar sketches. Since several language families consist only of poorly documented languages, the collection process was also influenced by data availability. Detected loans from culturally influential languages, such as Arabic, English, French, Malay, Mandarin Chinese, Portuguese, and Spanish, were removed, but using such a large dataset from a large number of typologically distinct languages also tends to impose issues with semantic boundaries. However, among all the extracted color words, there were only seven cases with alternative or dialectal forms, which mostly only involved small vowel differences, and therefore the form first cited in the source was selected for the present study.

The gathered data was transcribed into the International Phonetic Alphabet (IPA) with some differences. Oral and nasal vowels, pulmonic, co-articulated, and non-pulmonic consonants, as well as nasalized consonants, breathy (murmured) vowels, and creaky voiced sounds were coded as separate phonemes. This also applied to plain, voiced, nasalized, and voiceless nasalized clicks. Diphthongs, triphthongs, and affricates were divided into their components and counted as separate phonemes, e.g., [t͡s] resulted in [t] and [s]. Likewise, consonantal release types, aspiration, and co-articulations were also split and followed their respective place of articulation. Tones and stress were not recorded, but phonetic length was coded as a double occurrence of the same phoneme, e.g. [a:] resulted in [aa]. For more details on the collection and coding of textual data, refer to Johansson et al. (in press).

In order to make the data in text form comparable with acoustic measurements, such as spectral centroid and sonority, we utilized the same recordings of individual phonemes as in the pilot study, namely Seeing Speech (Lawson et al., 2015). As recordings of all possible speech sounds with all possible articulation types were not available, several sounds were grouped with their phonetically closest recorded proxy. This did not, however, have any large effects on the data as a whole since the available recordings covered the most typologically common and frequently occurring phonemes. Nasal, breathy, and creaky phonemes were assigned to their plain counterparts.

True mid vowels [e̞, o̞] were replaced with open-mid vowels [ɛ, ɔ], near-close [ɪ̈, ʊ̈, ʏ] except [ɪ] with close vowels [ɨ, ʉ, y], most central vowels with [ə], [æ, ä, ɐ] with [a], and [ʊ̈, ɞ̞] with [œ]. Lacking dentals, labiodentals, and palato-alveolars were assigned to alveolars, linguolabials to labials, uvulars to velars, pharyngeal trills to uvular trills, and pharyngeal plosives to glottal plosives. Lacking approximants were assigned to the corresponding fricatives, ejectives to their closest plain analogue, and missing taps/flaps were replaced with corresponding trills. Voiceless versions of nasal, laterals, clicks, and vibrants were replaced with voiced analogs. Recordings of the speech sounds in isolation were used, and in the few cases when this was not available (plosives [p, b, t, d, ʈ, ɖ, c, ɟ, k, g, q, ɢ, ʔ], taps/flaps [ⱱ, ɾ, ɽ], and implosives [ɓ, ɗ, ʄ, ɠ, ʛ]), recordings of the sounds in medial position between two neutral vowels (to prevent labialization, palatalization, velarization, pharyngealization, etc.) were used instead, with vowel segments deleted. Audio files and other supplementary materials can be downloaded from <https://osf.io/cu3bk/download>.

2.3. ANALYSIS

When analyzing the frequency of phonemes in different color words as a function of their spectral characteristics, it is preferable to treat vowels and consonants separately. Vowel quality is primarily determined by the frequency of the first two or three formants, which is not a meaningful acoustic feature for many consonants. Furthermore, based on the accounts of sound–color associations in Korean ideophones, it is plausible to assume that vowels and consonants could produce different sound symbolic mappings and strength of effects. In addition, vowels generally have less high-frequency energy than most consonants, particularly voiceless consonants like sibilants or clicks. [≠] While acoustically the most effective way to achieve a 'dark' sound would be to dispense with consonants altogether, words like *ouou* are seldom phonotactically tolerated because lexemes are phonotactically constrained to contain a mixture of vowels and consonants in most natural languages. We therefore analyzed the relation between color properties and the spectral characteristics of phonemes within the words for these colors separately for vowels and consonants. In both cases the models we built attempted to predict the brightness ratings (as obtained in the pilot study) or acoustic characteristics of a phoneme based on a particular feature of color (luminance or saturation) designated by the word in which this phoneme occurred.

The acoustic analysis of the IPA recordings is described in the pilot study. Luminance and saturation of the color words were calculated based on previously published CIELAB coordinates of collected cross-linguistic color-naming data from multiple speakers of 110 languages in the World Color Survey (Regier, Kay, & Cook, 2005). The speakers selected the chip(s) that represented the best example of each color in their respective language from an array of 330 color chips. The best-example choices for each color were then pooled to form cross-linguistic focal-color coordinates. In other words, because the World Color Survey did not include most of the languages in our sample, we made a crucial simplifying assumption that the CIELAB coordinates (and therefore also luminance and saturation) of the 11 sampled colors were roughly the same in all languages. We arbitrarily assigned a luminance of 25 to the words *dark* and *night* and 75 to the words *light* and *day*. Saturation was calculated as the distance from the achromatic central axis in CIELAB color space. The concept of saturation is arguably not applicable to the words *dark*, *night*, *day*, and *light*, so we dropped them from all analyses involving saturation.

We did not consider possible effects of hue as such for the following reasons: (1) its contribution would be impossible to distinguish from that of luminance and saturation with only four chromatic colors (red, blue, yellow, and green), and (2) psychological research on cross-modal associations between color and sound has produced much stronger evidence for cross-modal associations between sound and luminance or saturation than between sound and hue. Instead, to account for possible idiosyncratic sound symbolic patterns for each individual color, we analyzed its residual from the regression line. For a summary of the color concepts' visual parameter values, see Table 2.

Statistical analyses were performed on the unaggregated dataset using Bayesian mixed models, in which the unit of analysis was a single phoneme from the word for a particular color in one of the sampled languages. The task was to predict the acoustic characteristics of each phoneme (e.g., its sonority or formant frequencies) from the luminance or saturation of the color. Model selection with information criteria indicated that predictive power improved after including a random intercept per color and a random slope of the visual predictor (luminance or saturation) per language. In other words, for each acoustic characteristic, we estimated the trend driven by a visual predictor like luminance, while allowing individual colors to be associated with various acoustic properties and allowing the effect of the visual predictor to be language-specific. The random intercept per color provided an inferential measure of how much each color deviated from the main pattern (its 'residual'). All frequency measures (formant frequencies, spectral centroid) were log-transformed prior to modeling. Mixed models were fit in the Stan computational framework (http://mc-stan.org/) accessed from R using the *brms* package (Bürkner, 2017).[≠Refs] We specified mildly informative regularizing priors on regression coefficients so as to reduce overfitting and improve convergence.

TABLE 2. *The eleven investigated color concepts*

Color concept	WCS chip[1]	RGB	CIELAB	Luminance	Saturation[2]
RED	G1	186.9, 28.7, 71.3	41.2, 61.4, 17.9	41.2	64.0
GREEN	F17	0, 146.2, 69.9	51.6, –63.3, 29.0	51.6	69.6
YELLOW	C9	253.7, 193.4, 0	81.4, 7.3, 109.1	81.4	109.4
BLUE	F29	0, 129, 205	51.6, –3.4, –48.1	51.6	48.2
BLACK	J0	0, 0, 0	0, 0, 0	0	0
WHITE	A0	255, 255, 255	100, 0, 0	100	0
GRAY	–	119, 119, 119	50, 0, 0	50	0
NIGHT	–	59, 59, 59	25, 0, 0	25	–
DAY	–	185, 185, 185	75, 0, 0	75	–
DARK	–	59, 59, 59	25, 0, 0	25	–
LIGHT	–	185, 185, 185	75, 0, 0	75	–

NOTES: [1] World Color Survey chip corresponding to focal color (Regier et al., 2005); [2] saturation was calculated as Euclidean distance from the achromatic central spindle of the CIELAB space: $\surd(a^2 + b^2)$.

3. Results

3.1. VOWELS

We discovered a significant association between the luminance of a color and the sonority of vowels in the word for this color (Figure 4A). The average sonority of vowels was predicted to be 0.4 points (95% CI [0.2, 0.6]) higher on a scale of 12 to 16 in words for WHITE (luminance = 100) than in words for BLACK (luminance = 0). Luminance also predicted the subjective brightness of vowels in a color's name (Figure 4B): the average brightness rating of vowels in words for WHITE was predicted to be 12% (95% CI [4, 20]) higher than in words for BLACK. The association between spectral centroid and luminance was not statistically significant, but it was in the predicted direction (83 Hz, 95% CI [–34, 199]). In addition, luminance predicted an increase in the frequency of F1, but not F2 or F3 (Figure 4D, 4E, 4F). It is worth pointing out that all these vowel characteristics tend to be positively correlated. For example, according to the acoustic analysis of IPA recordings in the pilot study, F1 correlates with both vowel sonority (r = .41) and vowel brightness (r = .47), and sonority is also positively associated with brightness (r = .37). In other words, the large picture is that there is a tendency for both bright and sonorous vowels (which are largely the same) to occur in the words for light colors, while dark and less sonorous vowels are more common in the words for darker colors.

In order to ascertain that the observed association between color luminance and the 'brightness' of phonemes in the corresponding words is not caused by another color characteristic, it would be desirable to perform multiple regression controlling for saturation and hue. Unfortunately, this is not possible with only a few color words. Looking at univariate effects of

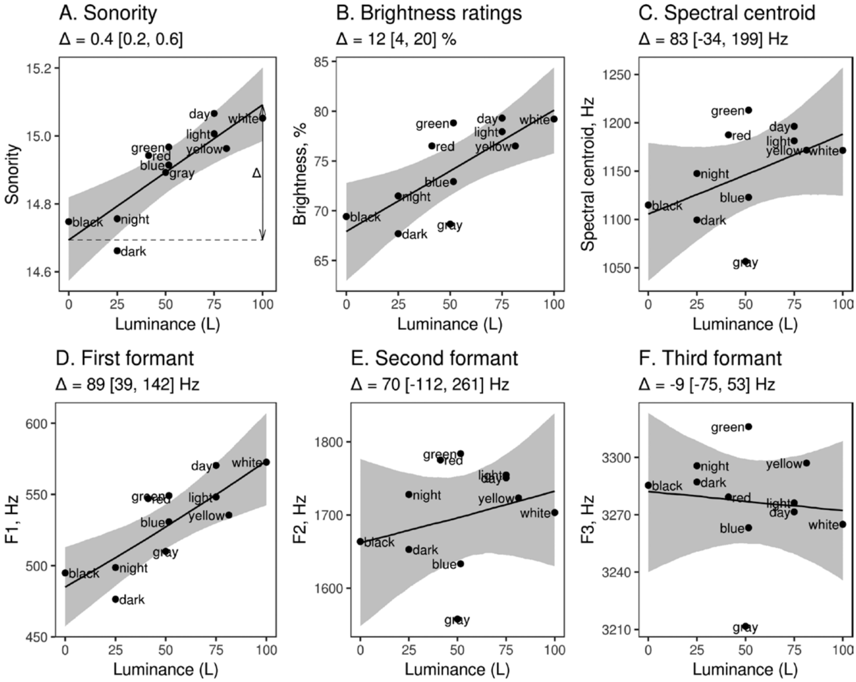


Fig. 4. Vowels: univariate associations between acoustic characteristics of vowels found in different color words and the luminance of the designated color. Brightness ratings are taken from the pilot study (recorded rather than synthetic vowels). The line and shaded area show fitted values from linear mixed models as the median of posterior distribution and 95% CI. Text labels mark the means of observed values. The deltas show the predicted difference between WHITE (L = 100) and BLACK (L = 0).

saturation, we failed to detect any association with measures of vowel sonority or brightness (Figure 5). On the other hand, a few outliers in Figure 4 hint that other processes might be involved. For sonority and luminance (Figure 4A), all colors lie close to the regression line except DARK with a lower-than-expected sonority of vowels for its luminance. More formally, the 95% CI for the random intercept overlaps with zero for all colors except DARK, for which the sonority of vowels was slightly lower (–0.1 [–0.21, –0.01]) than expected for its luminance. On the other hand, the luminance of DARK was arbitrarily set to 25, so this result should not be over-interpreted. More tellingly, the brightness ratings, spectral centroid, and F2 of vowels were all unexpectedly high in GREEN and low in GRAY (random intercepts exclude or nearly exclude zero; details not shown). RED also had vowels with marginally higher-than-expected brightness and spectral centroid and a significantly elevated F2. Some other aspects of these colors, apart from their

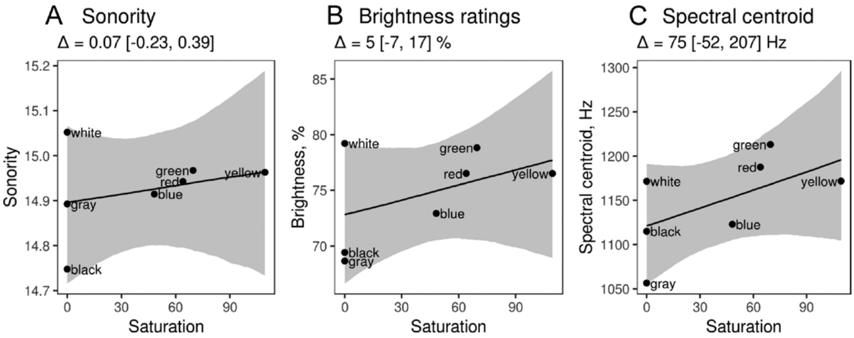


Fig. 5. Vowels: univariate associations between acoustic characteristics of vowels found in different color words and the saturation of the designated color. See Figure 4 for detailed explanations.

luminance and saturation, may thus be affecting the brightness of vowels in their names. Overall, however, luminance remains the most relevant color property that is mapped onto the acoustic characteristics of vowels.

3.2. CONSONANTS

Moving on to consonants, we could not analyze the association between luminance and formants or subjective brightness because (1) formants are not meaningful for many of the consonants and (2) in the pilot study we only obtained ratings of subjective brightness for a subset of 19 out of 78 consonants. The analysis in consonants was therefore limited to the sonority and spectral centroid. We found no relation between the spectral centroid of consonants and color luminance (Figure 6B) or saturation (Figure 6D), suggesting that the 'brightness' of consonants in color words does not depend on the perceptual characteristics of the designated color. In contrast, there was some evidence that sonorous consonants were over-represented in words for both luminant and saturated colors. For luminance, there was a marginal positive effect: the sonority of consonants was predicted to be 0.43 (95% CI [–0.07, 0.93]) higher in WHITE vs. BLACK (Figure 6A). For saturation, the positive effect of sonority was slightly stronger: 0.51, 95% CI [0.02, 0.84] (Figure 6C). In both cases, the effect size was comparable to that found in vowels (Figure 4A), corresponding to a difference of about 0.5 points in sonority rank.

To summarize, the evidence for consonants was less consistent than for vowels. We found no indication that the distribution of spectral energy in consonants in color words was aligned with visual characteristics of the designated color. However, there appears to be a tendency for sonorous consonants to appear more often in words for more luminant and saturated colors.

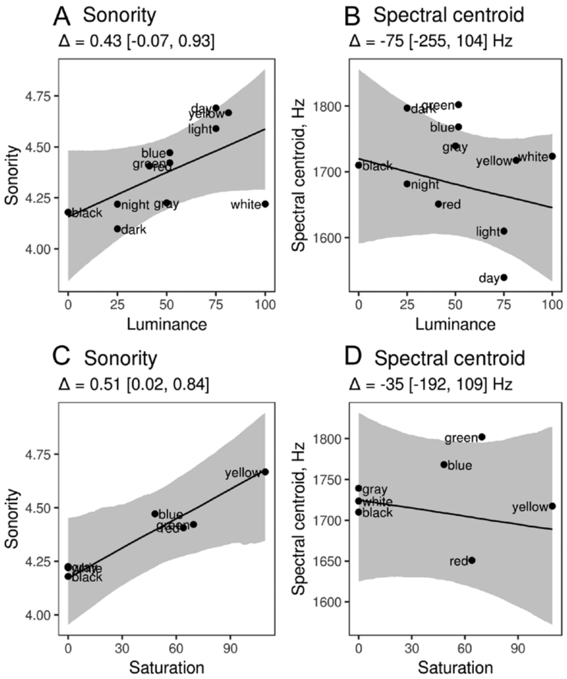


Fig. 6. Consonants: univariate associations of acoustic characteristics of consonants with the luminance and saturation of the designated color. See Figure 4 for detailed explanations.

4. General discussion

We investigated associations between sound and color in color words from 245 areally spread language families by testing specific predictions regarding phoneme distributions based on experimental evidence from psychological research on sound–color cross-modal correspondences. In particular, we looked at two visual parameters, luminance and saturation, which were derived from typological data of color coordinates, and a series of acoustic parameters: the sonority and spectral centroid of vowels and consonants, the perceived brightness of vowels, and the first three vowel formants. We first confirmed that spectral centroid can be used as a convenient proxy for perceived brightness of both vowels and consonants when direct brightness ratings are not available. As expected, based on previous descriptions of sound symbolism and experimental evidence, the main analysis then demonstrated that brighter and more sonorous vowels, as well as vowels with high F1, are more common in the words for more luminant colors (cf. Parise & Pavani, 2011). More sonorous consonants, on the other hand, are common in the words for colors with high saturation, and possibly high luminance as well.

Among the investigated acoustic and visual dimensions, the strongest and most consistent associations were found between acoustic characteristics of vowels and color luminance. The sound symbolic strength of vowels could be explained by the fact that different vowels are more gradient than consonants. All that is needed to change the acoustic signal of a vowel is to modify the tongue's height or backness. Lip-rounding also plays a part, but it is generally confined to back vowels. While the place of articulation of consonants is somewhat comparable to how vowels are produced, their manner of articulation is qualitatively different since it involves both active and passive articulators, which allows for greater variance in speech sound quality. Consequently, it could be easier to map continuous or gradient dimensions, such as color luminance or saturation, to vowels than to consonants (Tufvesson, 2011). It is still possible to create similar mappings using consonants, but it might be less obvious how to arrange the distinct combinations of features that result from both place and manner of articulation.

4.1. SOUND SYMBOLISM AS A BASIS FOR LEXICALIZATION PATTERNS

The results also revealed similarities with lexicalization patterns (the process of adding lexemes to the lexicon) of color words globally. The study of typological color word semantics has its origins in Berlin and Kay (1969), who proposed that, cross-linguistically, color words are added in a particular order: BLACK and WHITE > RED > GREEN and YELLOW > BLUE > BROWN > PURPLE, PINK, ORANGE, and GRAY. Kay and Maffi (1999) developed and nuanced the hierarchy by investigating six basic colors (RED, GREEN, YELLOW, BLUE, WHITE, BLACK) in 110 unwritten languages of non-industrialized societies in order to approximate the use of colors throughout human history. They showed that all languages seem to make at least one color distinction that cuts right through the three CIELAB color space parameters by separating light/warm colors (WHITE, YELLOW, RED) from dark/cool colors (BLACK, BLUE, GREEN). Languages that distinguish between at least three color words also seemed to keep the dark/cool colors coded as a single color word, but separate the light color WHITE from the warm colors, YELLOW and RED.

The results of the present paper showed that luminance produced the strongest sound symbolic results, and it is also the most fundamental parameter for distinguishing colors based on lexicalization patterns. The second split (WHITE from RED and YELLOW) separates the most luminant color from the warm colors, which can also be perceived as the most saturated colors (Witzel & Franklin, 2014). Although uncertain, we did find a tendency for vowels and consonants to have different sound symbolic functions, which also seems to correlate with Korean color sound symbolism, in which color luminance can be manipulated by changing vowel height, which correlates with F1, and saturation by changing consonant tensing (Rhee, 2019). If this possible relationship holds, it suggests that primary acoustic and articulatory affordances provide an efficient vehicle for communicating perceptual contrasts and could therefore influence lexicalization processes.

A common pattern found across languages is that color words that are lexicalized late are derived from referents that are located in the surrounding world. In many languages, the word meaning 'orange' is derived from the fruit *Citrus sinensis*, the word meaning 'pink' is derived from roses, salmon, or peaches, and the word meaning 'gray' is often derived from the word for ashes. Likewise, it is quite possible that monolexemic color terms are historically derived from concrete referents as well, but the etymological distance could be too great or the historical development of a language too poorly understood to link the referents to the color word. For example, many basic color words in Indo-European (perhaps the most historically well-documented language family in the world) and several other language families can be traced back to concrete referents, such as the meaning 'red' from rust or worms and the meaning 'green' from plants (Derksen, 2008, 2010; Kroonen, 2010; Matisoff, 2011; Metsmägi, Sedrik, & Soosaar, 2012). Thus, color terms generally regarded as monolexemic can still carry phonetic similarities to their original referents.

Interestingly, our findings showed that the color GRAY seemed to behave somewhat differently from other colors of comparable luminance as it contained vowels with rather low spectral centroid, F2 and F3. These features are typical for close back rounded vowels, such as [u], and might also be sound symbolically motivated. Similarly, ASHES has been confirmed to be rather strongly sound symbolic (Blasi et al., 2016; Johansson et al., in press). ASHES, along with other concepts that relate to air and airflow, such as BLOW, BREATHE, WING, etc., tend to contain rounded vowels (which are generally back vowels) and voiceless labial and fricative consonants. The most plausible explanation for these associations is that sounds that involve air leaving the body, which is further intensified by the extra friction created by lip rounding, onomatopoeically evoke a general meaning of air moving or something moving through air. This means that, cross-linguistically, GRAY rests on a sound symbolic foundation as well, but not on the same foundation as the six basic colors, making it an apparent outlier in terms of color sound symbolism.

We also found that GREEN had phonemes with more high-frequency energy than expected for a color of this luminance. A possible reason is that these words are, just like GRAY, also derived from a natural referent iconically associated with high-frequency energy. However, as opposed to GRAY, previous cross-linguistic studies have not found any good candidates for such an association. Nevertheless, since words meaning 'green' are often derived from words relating to 'growth' and 'movement', it is possible that this also gives this color an 'energetic' connotation. If so, it could easily be mapped to high energy sounds similar to the cross-modal effects found for GREEN when in contrast with RED or darker colors (Cuskley et al., 2019; Kim et al., 2017; Marks, 1975; Miyahara et al., 2012; Moos et al., 2014; Wrembel, 2009). Alternatively, hue as such – that is, red–green and yellow–blue oppositions as distinct from the effects of luminance and saturation – could be contributing to sound symbolism in color words. Unfortunately, we could not test this intriguing possibility directly in the present study since the effects of luminance, saturation, and hue could not be separated based on only four chromatic colors.

Another important limitation of the present study is the assumed universality of focal colors (Regier et al., 2005): for example, we assumed that GREEN has the same CIELAB coordinates in all sampled languages, whereas in fact it represents somewhat different foci and regions of the color space in different languages. A more stringent analysis can become possible in future, when color concepts have been mapped in a sufficient number of minimally related languages. As another direction for future research, the names of less fundamental colors, such as BROWN, ORANGE, PINK, and PURPLE, could potentially also be affected by sound symbolism and can therefore also be investigated, providing a more nuanced picture and helping to disambiguate sound symbolic effects of luminance, saturation, and hue.

A potential limitation of analyzing additional colors is that their names are generally derived from natural referents in the surrounding world, diluting sound symbolic effects of color per se, as we saw in this study with GRAY. Going beyond color words, auditory frequency has been mapped onto several modalities (Walker, 2012; Westermann, 1927); for example, higher pitch is associated with both brightness and angular shapes, while lower pitch is associated with darkness and smoother shapes (Walker et al, 2012). Because of this, just as luminance and saturation affect phoneme sonority and brightness in color words, similar sound symbolic effects could potentially be discovered in a range of semantically related fundamental descriptive concepts which denote shape, size, weight, height, density, etc.

4.2. THE ORIGINS OF SOUND–COLOR ASSOCIATIONS

Aside from lexicalization patterns, there is strong evidence that associations between luminance and phonetic dimensions, such as pitch, are among the most fundamental types of cross-modal mappings. Various types of experiments have shown that both synesthetic and non-synesthetic people (Moos et al., 2014;

Ward et al., 2006), toddlers (Mondloch & Maurer, 2004), and chimpanzees (Ludwig, Adachi, & Matsuzawa, 2011) map WHITE/BRIGHT to high-pitched sounds and BLACK/DARK to low-pitched sounds. This suggests that these cross-modal correspondences are present early in human ontogenesis and must have evolved before the human and our closest living relatives' lineages split apart. Furthermore, Bankieris and Simner (2015) argue that synesthesia and cross-modal correspondences are qualitatively the same phenomenon and link them to the origin of sound symbolism. This linkage, along with other possible underlying mechanisms of sound symbolism, is discussed in detail by Sidhu and Pexman (2018).

Although the sound symbolic effects related to saturation were more uncertain than those for luminance, the perceptual saliency of saturation makes it likely that these sound symbolic effects are credible. Evidence from prelinguistic infants suggests that color categorization is not purely shaped by communication and culture, but also by underlying biological mechanisms (Skelton, Catchpole, Abbott, Bosten, & Franklin, 2017). Furthermore, one of the primary color distinctions found in infants aged three months or younger is a distinction between long wavelength colors and short wavelength colors (Adams, 1987), which, in turn, correspond to colors that are perceived as more saturated, such as RED and YELLOW, and less saturated, such as GREEN and BLUE (Witzel & Franklin, 2014). Likewise, evidence from spatial clustering of neurons in the macaque primary visual cortex indicates that contrasts between the L and M cone cell type could form a biological foundation for this distinction between high and low saturation (Xiao, Kavanau, Bertin, & Kaplan, 2011). Furthermore, Sugita (2004) showed that exposing infant female Japanese macaques to only monochromatic lighting for one month impoverished their ability to distinguish colors compared to monkeys who had had access to the full spectrum of colors.

Both luminant and saturated colors seem to play particularly important roles in color perception as well as in the surrounding LIFEWORLD which we experience and interact with (Gibson, 1977), since they contrast sharply with the green–brown backdrop that nature generally provides. For example, the yellow-to-red colors of ripe fruits attract fruit-eating animals which, in exchange for food, distribute the plants' seeds, and insects use bright colors or patterns to prevent predators from eating them (Valenta et al., 2018). These marked colors also translate into cultural color associations, such as the connection between red and blood, life, death, danger, excitement, etc., which means that, conceptually, color is one of the most salient descriptive features available for humans. Unsurprisingly, dimensions of color are not only distinguished in language, but the most central dimensions seem to follow the same pattern of lexicalization. In contrast to most of our sensory perception, language has to be learned, and important distinctions between features of objects have to be acquired quickly and easily. In color words, sound–color mappings offer a bridge between language and sensory experiences, which enables language users to efficiently organize sensory features (Tufvesson, 2011).

Consequently, luminance and saturation serve as stepping-stones for carving up the gradient color spectrum into a number of manageable segments. These can then be used for description and tend to be iconically named as a result of increased learnability. Indeed, several studies have shown that iconicity makes words easier to learn (Imai & Kita, 2014; Massaro & Perlman, 2017; Nygaard, Cook, & Namy, 2009) and has a number of functional and communicative benefits (Tamariz, Roberts, Martínez, & Santiago, 2018). It is therefore reasonable to assume that the prevalence of sound symbolism in color words across languages has been perpetuated because it aids lexical acquisition, leading to a cultural transmission bias. In addition, iconic patterns, just like cross-linguistic lexicalization patterns of color words, seem to be universal tendencies with some exceptions rather than absolute universals (Levinson, 2000). However, while these malleable patterns are not necessarily present in the same words, in all languages, and at the same time, they seem to decay and reform within languages over time (Flaksman, 2017; Johansson & Carling, 2015).

5. Concluding remarks

This study aimed to ground color sound symbolism in natural languages in low-level perceptual processes such as cross-modal associations. We investigated a range of acoustic measurements (the first three vowel formants, spectral centroid, sonority, and brightness ratings obtained in a perceptual experiment) in eleven words for basic colors or semantically related concepts from 245 language families. The results showed that luminance was associated with the sonority, brightness, and the first formant of vowels, while saturation and possibly luminance were less robustly associated with the sonority of consonants. An important implication is that sound–meaning associations might have great significance for our understanding of how linguistic categories have developed, since high luminance and high saturation are the two visual traits that guide the lexicalization of color words across languages. These associations can be linked to the increased learnability provided by iconicity, and they can be considered from both ontogenetic and phylogenetic perspectives, considering that cross-modal mappings between pitch and luminance can be traced back at least to our last common ancestor with chimpanzees. In sum, color sound symbolism seems to be grounded in evolutionary, environmental, biological, and developmental constraints.

However, in order to thoroughly understand how these sound symbolic associations are formed, it is necessary to map more fully the roles that vowels, consonants, and other, more fine-grained phonetic distinctions play within these associations.

REFERENCES

Adams, R. J. (1987). An evaluation of color preference in early infancy. *Infant Behavior and Development* **10**(2), 143–150.

Anikin, A. (2019). Soundgen: an open-source tool for synthesizing nonverbal vocalizations. *Behavior Research Methods* **51**(2), 778–792.

Anikin, A. & Johansson, N. (2019). Implicit associations between individual properties of color and sound. *Attention, Perception, & Psychophysics* **81**(3), 764–777.

Bankieris, K. & Simner, J. (2015). What is the link between synaesthesia and sound symbolism? *Cognition* **136**, 186–195.

Berlin, B. & Kay, P. (1969). *Basic color terms: their universality and evolution*. Berkeley & Los Angeles: University of California Press.

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.

Bürkner, P. C. (2017). brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**(1), 1–28.

Cuskley, C., Dingemanse, M., Kirby, S. & van Leeuwen, T. M. (2019). Cross-modal associations and synesthesia: categorical perception and structure in vowel–color mappings in a large online sample. *Behavior Research Methods* **51**(4), 1651–1675.

Derksen, R. (2008). *Etymological dictionary of the Slavic inherited lexicon* (Leiden Indo-European Etymological Dictionary Series). Leiden, Boston: Brill.

Derksen, R. (2010). *Etymological dictionary of the Baltic inherited lexicon*. “Žalias”. Leiden: Brill. Online <http://dictionaries.brillonline.com> (last accessed 12 December 2018).

Diffloth, G. (1994). i: big, a: small. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound symbolism* (pp. 107–114). Cambridge: Cambridge University Press.

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.

Edmiston, P., Perlman, M. & Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proceedings of the Royal Society B: Biological Sciences* **285**(1874), https://doi.org/10.1098/rspb.2017.2709

Fastl, H. & Zwicker, E. (2006). *Psychoacoustics: facts and models* (Vol. 22). Berlin, Heidelberg: Springer-Verlag.

Flaksman, M. (2017). Iconic treadmill hypothesis. In M. Bauer, A. Zirker, O. Fischer & C. Ljungberg (eds.), *Dimensions of iconicity* (Iconicity in Language and Literature 15) (pp. 15–38). Amsterdam, Philadelphia: John Benjamins.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (eds.), *Perceiving, acting, and knowing* (pp. 67–82). Hillsdale, NJ: Erlbaum.

Hamilton-Fletcher, G., Witzel, C., Reby, D. & Ward, J. (2017). Sound properties associated with equiluminant colours. *Multisensory Research* **30**(3–5), 337–362.

Hammarström, H., Forkel, R. & Haspelmath, M. (2017). Glottolog 3.0. Jena: Max Planck Institute for the Science of Human History. Online <http://glottolog.org>.

Haspelmath, M. & Tadmor, U. (eds.) (2009). *Loanwords in the world’s languages: a comparative handbook*. Berlin, New York: De Gruyter Mouton.

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology* **109**(2), 219–238.

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), https://doi.org/10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M. & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition* **109**(1), 54–65.

Johansson, N., Anikin, A., Carling, G. & Holmer, A. (in press). *The typology of sound symbolism: defining macro-concepts via their semantic and phonetic features. Linguistic Typology*.

Johansson, N. & Carling, G. (2015). The de-iconization and rebuilding of iconicity in spatial deixis: an Indo-European case study. *Acta Linguistica Hafniensia* **47**(1), 4–32.

Jones, J. M., Vinson, D., Clostre, N., Zhu, A. L., Santiago, J. & Vigliocco, G. (2014). The bouba effect: sound-shape iconicity in iterated and implicit learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* **36**(36), 2459–2464.

Kay, P. & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist* **101**(4), 743–760.

Kim, H. W., Nam, H. & Kim, C. Y. (2017). [i] is lighter and more greenish than [o]: intrinsic association between vowel sounds and colors. *Multisensory Research* **31**(5), 419–437.

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

Kroonen, G. (2010). *Etymological dictionary of Proto-Germanic. "Grōni-"*. Leiden: Brill. Online <http://dictionaries.brillonline.com> (last accessed 12 December 2018).

Lawson, E., Stuart-Smith, J., Scobbie, J. M., Nakai, S., Beavan, D., Edmonds, F., Edmonds, I., Turk, A., Timmins, C., Beck, J., Esling, J., Leplatre, G., Cowen, S., Barras, W. & Durham, M. (2015). Seeing speech: an articulatory web resource for the study of Phonetics. University of Glasgow. Online <http://www.seeingspeech.ac.uk/>.

Levinson, S. C. (2000). Yélî Dnye and the theory of basic color terms. *Journal of Linguistic Anthropology* **10**(1), 3–55.

Ludwig, V. U., Adachi, I. & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences* **108**(51), 20661–20665.

Ludwig, V. U. & Simner, J. (2013). What colour does that feel? Tactile–visual mapping and the development of cross-modality. *Cortex* **49**(4), 1089–1099.

Lupyan, G. & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition* **7**(2), 167–193.

Marks, L. E. (1974). On associations of light and sound: the mediation of brightness, pitch, and loudness. *American Journal of Psychology* **87**(1/2), 173–188.

Marks, L. E. (1975). On colored-hearing synesthesia: cross-modal translations of sensory dimensions. *Psychological Bulletin* **82**(3), 303–311.

Marks, L. E. (1987). On cross-modal similarity: auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance* **13**(3), 384–394.

Massaro, D. W. & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: implications for vocabulary learning. *Frontiers in Communication* **2**(4). https://doi.org/10.3389/fcomm.2017.00004

Matisoff, J. A. (ed.) (2011). The Sino-Tibetan etymological dictionary and thesaurus. University of California, Berkeley. Online <http://stedt.berkeley.edu/> (last accessed 12 December 2018).

Metsmägi, I., Sedrik, M. & Soosaar, S.-E. (2012). *ETY – Eesti etümoloogiasõnaraamat* [Estonian etymology dictionary]. Eesti Keele Instituut. Online <http://www.eki.ee/dict/ety/> (last accessed 12 December 2018).

Miyahara, T., Koda, A., Sekiguchi, R. & Amemiya, T. (2012). A psychological experiment on the correspondence between colors and voiced vowels in non-synesthetes. *Kansei Engineering International Journal* **11**(1), 27–34.

Mok, P. P. K., Li, G., Li, J. J., Ng, H. T. Y. & Cheung, H. (2019). Cross-modal association between vowels and colours: a cross-linguistic perspective. *Journal of the Acoustical Society of America* **145**(4), 2265–2276.

Mondloch, C. J. & Maurer, D. (2004). Do small white balls squeak? Pitch–object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience* **4**(2), 133–136.

Moos, A., Smith, R., Miller, S. R. & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder. *i-Perception* **5**(2), 132–142.

Newman, S. S. (1933). Further experiments in phonetic symbolism. *American Journal of Psychology* **45**(1), 53–75.

Nielsen, A. & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition* **4**(2), 115–125.

Nygaard, L. C., Cook, A. E. & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition* **112**(1), 181–186.

Orlandatou, K. (2012). The role of pitch and timbre in the synaesthetic experience. In E. Cambouropoulos, C. Tsougras, P. Mavromatis & K. Pastiadis, *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 751–758). Online <https://www.semanticscholar.org/paper/The-Role-of-Pitch-and-Timbre-in-the-Synaesthetic-Orlandatou/c948f85bca6bce956e5deb96533d821aecb5b9de>.

Pagel, M., Atkinson, Q. D., Calude, A. S. & Meade, A. (2013). Ultraconserved words point to deep language ancestry across Eurasia. *Proceedings of the National Academy of Sciences* **110**(21), 8471–8476.

Parise, C. V. & Pavani, F. (2011). Evidence of sound symbolism in simple vocalizations. *Experimental Brain Research* **214**(3), 373–380.

Parker, S. G. (2002). *Quantifying the sonority hierarchy*. Unpublished doctoral dissertation, University of Massachusetts at Amherst.

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

Regier, T., Kay, P. & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences* **102**(23), 8386–8391.

Rhee, S. (2019). Lexicalization patterns in color naming in Korean. In I. Raffaelli, D. Katunar & B. Kerovec (eds), *Lexicalization patterns in color naming: a cross-linguistic perspective* (pp. 109–128). Amsterdam: John Benjamins.

Sapir, E. (1929). A study in phonetic symbolism. *Journal of Experimental Psychology* **12**(3), 225–239.

Saussure, F. (1916). *Course in general linguistics*. Duckworth: London.

Schubert, E., Wolfe, J. & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen & P. Webster (eds), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 112–116). Online <https://phys.unsw.edu.au/jw/reprints/SchWolTarICMPC8.pdf>.

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

Skelton, A. E., Catchpole, G., Abbott, J. T., Bosten, J. M. & Franklin, A. (2017). Biological origins of color categorization. *Proceedings of the National Academy of Sciences* **114**(21), 5545–5550.

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

Sugita, Y. (2004). Experience in early infancy is indispensable for color perception. *Current Biology* **14**(14), 1267–1271.

Swadesh, M. (1971). *The origin and diversification of language* (ed. post mortem by J. Sherzer). London: Transaction Publishers.

Tamariz, M., Roberts, S. G., Martínez, J. I. & Santiago, J. (2018). The interactive origin of iconicity. *Cognitive Science* **42**(1), 334–349.

Tufvesson, S. (2011). Analogy-making in the Semai sensory world. *The Senses and Society* **6**(1), 86–95.

Valenta, K., Kalbitzer, U., Razafimandimby, D., Omeja, P., Ayasse, M., Chapman, C. A. & Nevo, O. (2018). The evolution of fruit colour: phylogeny, abiotic factors and the role of mutualists. *Scientific Reports* **8**(1). doi:10.1038/s41598-018-32604-x

Walker, L., Walker, P., & Francis, B. (2012). A common scheme for cross-sensory correspondences across stimulus domains. *Perception* **41**(10), 1186–1192.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: visual angularity is hard, high-pitched, and bright. *Attention, Perception, & Psychophysics* **74**(8), 1792–1809.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A. & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science* **21**(1), 21–25.

Ward, J., Huckstep, B. & Tsakanikos, E. (2006). Sound–colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* **42**(2), 264–280.

Wescott, R. W. (1975). Tonal iconicity in Bini colour terms. *African Studies* **34**(3), 185–192.

Westermann, D. H. (1927). Laut, Ton und Sinn in westafrikanischen Sudansprachen. In F. Boas (ed.), *Festschrift Meinhof* (pp. 315–328). Glückstadt, Hamburg: Gedruckt bei J. J. Augustin.

Wichmann, S., Holman, E. W. & Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy* **12**(4), 844–858.

Witzel, C. & Franklin, A. (2014). Do focal colors look particularly 'colorful'? *Journal of the Optical Society of America A* **31**(4), A365–A374.

Wrembel, M. (2009). On hearing colours – cross-modal associations in vowel perception in a non-synaesthetic population. *Poznań Studies in Contemporary Linguistics* **45**(4), 595–612.

Xiao, Y., Kavanau, C., Bertin, L. & Kaplan, E. (2011). The biological basis of a universal constraint on color naming: cone contrasts and the two-way categorization of colors. PloS one **6**(9). https://doi.org/10.1371/journal.pone.0024994

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

LUND
UNIVERSITY

The Faculties of Humanities and Theology
Centre for Languages and Literature
ISBN 978-91-89213-02-9

9 789189 213029