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

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

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,

*Corresponding author: Niklas Erben Johansson [nɪklas æʁbɛn juʊhanːsɔn], 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 [ɛndrʲiej ɐnʲiˈɪkʲɪn], 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 kʰɑːʁlɪŋ]: E-mail: gerd.carling@ling.lu.se, **Arthur Holmer [ɑːʁtɐʁ 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.

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 uni-modal 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 *loooooong;* (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 $*\acute{g}^h reh_1\text{-}ni\text{-}$, 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	FvZ _FS	DS _MS
CLOUD	FOUR	LIVE	RED	STRAIGHT	WHAT?	FvZ _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.
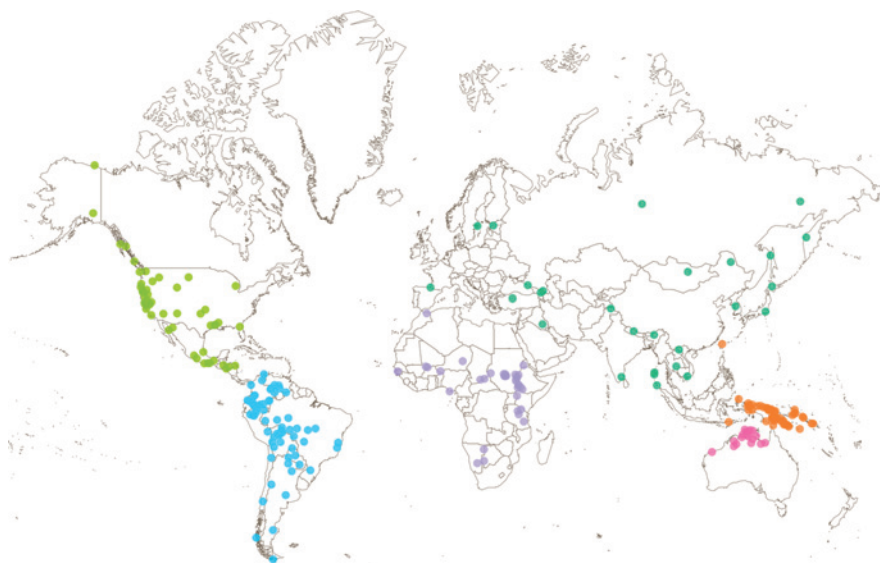

Figure 1: Featured languages divided into the same geographical macro-areas as used by Blasi et al. (2016) and shown by color (olive green: North America, blue: South America, forest green: Eurasia, purple: Africa, orange: Papunesia, pink: Australia).

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 *ti³*. 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, [kʷ], 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, [ts'], [k͡p], [pᵐ], [kʰ], [kʷ] and [kʲ] were transcribed as /t's'/, /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 [ᵍʘ]. 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/, /x'/, /q/ and /q'/. 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, [aː] 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 [aː]) as different sounds would, on the other hand, fail to record the qualitative similarity between them (for example [a] and [aː] coded as /a/ and /aː/), and coding them as the same sound (for example [a] and [aː] 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, œ, ä, ɒ̈, ɑ, ɒ, ã, œ̃, ä̃, ɒ̈̃, ɑ̃, ɒ̃
		Backness	[front]	i, y, e, ø, a, œ, 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, œ, ɒ̈, ɒ, ỹ, ʉ̃, ũ, ø̃, ɵ̃, õ, œ̃, ɒ̈̃, ɒ̃
	Aggregated	Extreme	[high-front]	i, y, e, ø, ĩ, ỹ, ẽ, ø̃
			[low-front]	a, œ, ã, œ̃
			[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]	œ, œ̃
			[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, ɳ̊, 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]	ⱱ̟, ⱱ, ɾ̥, r, ɽ̊, ɽ, ʀ̥, ʀ, ʜ, ʕ
			[lat]	ɬ, l, ʎ̥, ʎ, ʟ̥, ʟ
		Place	[lab]	m̥, m, p, b, f, v, ⱱ̟, ⱱ
			[alv]	n̥, n, t, d, s, z, ɾ̥, 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, ʕ

(continued)

Table 3: (*continued*)

Sound class		Sound group	Cardinal sounds
	Voicing	[–voice]	m̥, p, f, ɓ, n̥, t, s, ɾ̥, ɬ, ɲ̊, c, ç, ʈ̥, 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, ʕ
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]	ɓ, ɾ̥, ʈ̥, ʀ̥, ʜ
		[vib, +v]	ʙ, r, ɽ, ʀ, ʕ
		[lat, –v]	ɬ, ʎ̥, ʟ̥
		[lat, +v]	l, ʎ, ʟ
	Place-voicing	[lab, –v]	m̥, p, f, ɓ
		[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, ʈ̥, 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, ʕ

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 over-representation (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 $\star$ 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.
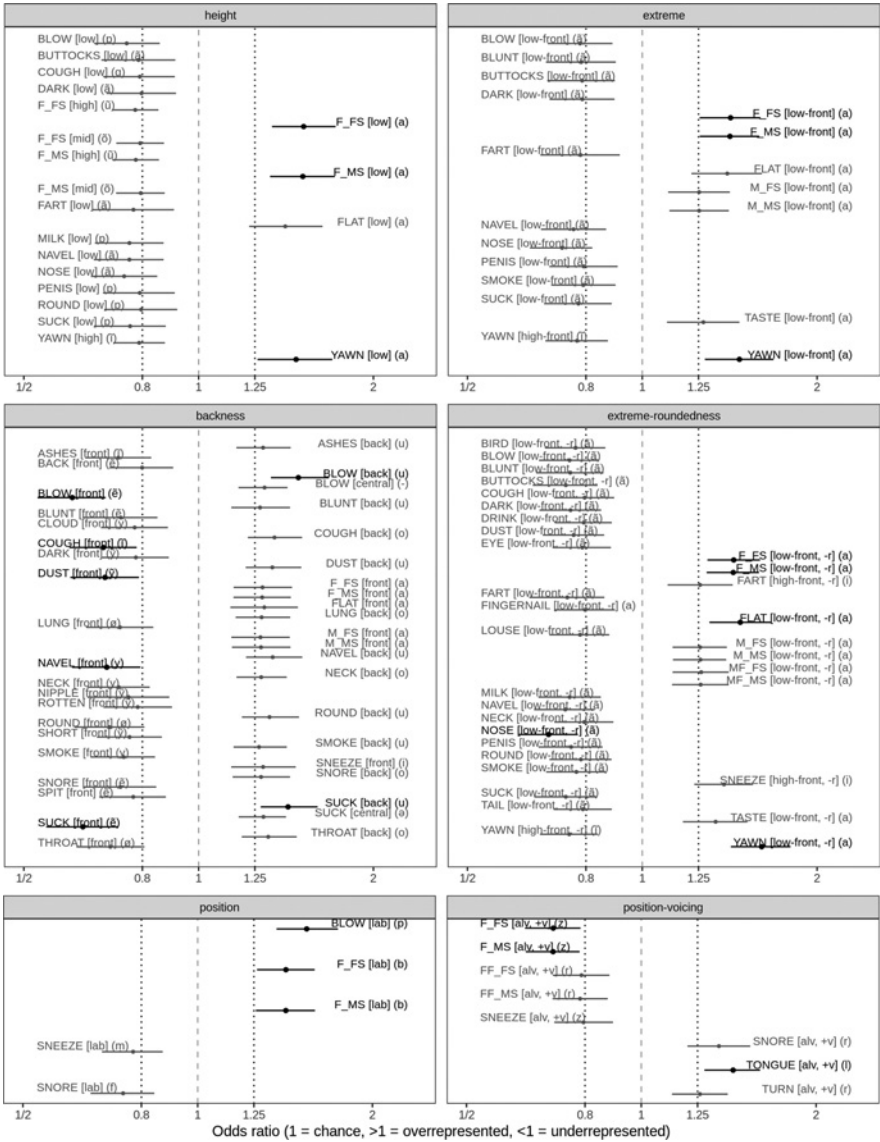
Figure 2: Over- and underrepresented sound groups in 344 concepts: strong (black) and weak (gray) associations. Each point shows the median of posterior distribution of the ratio of observed to expected odds, with 95% CI. Text labels show the concept, associated sound group, and the most strongly over- 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.
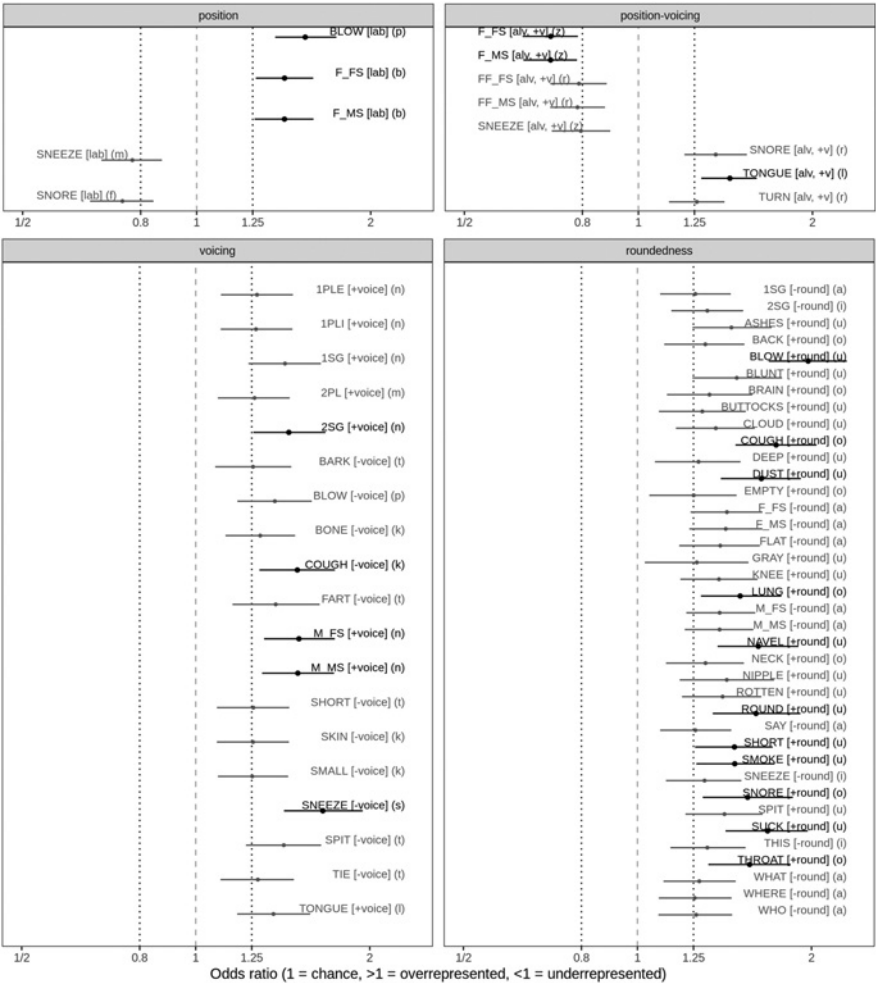
Figure 3: Over- and underrepresented sound groups in 344 concepts: strong (black) and weak (gray) associations. Each point shows the median of posterior distribution of the ratio of observed to expected odds, with 95% CI. Text labels show the concept, associated sound group, and the most strongly over- 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		–

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 over-representation 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).
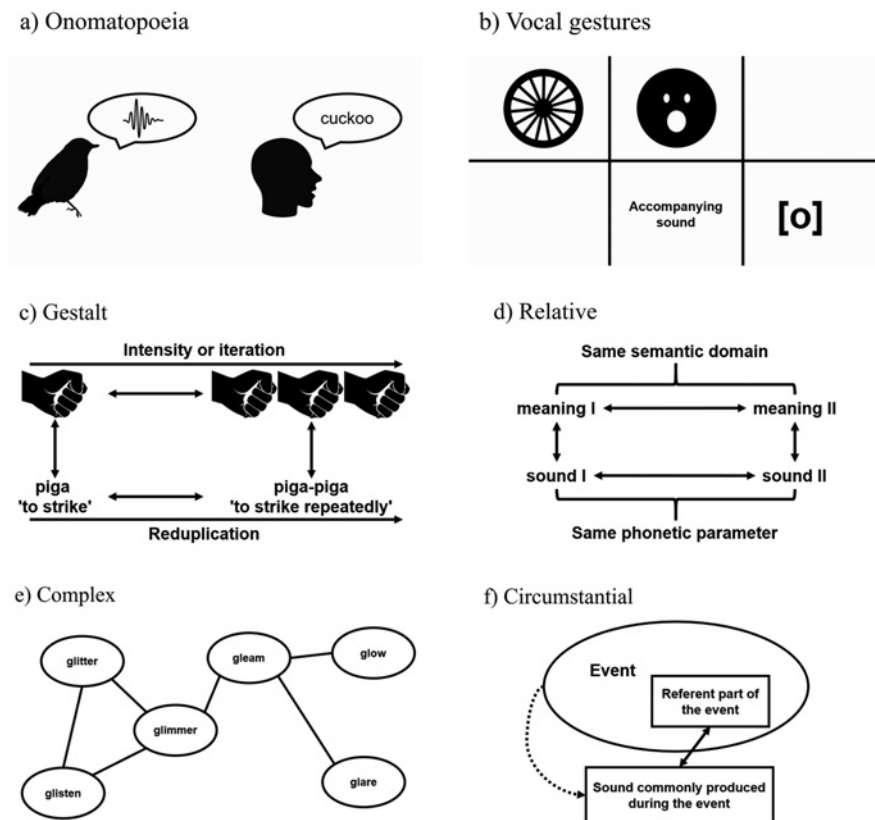

Figure 4: Illustrated simplifications of types of sound symbolism described by Dingemanse (2011), Carling and Johansson (2014) and the present paper. a) Onomatopoeia (imitative): acoustic approximations using the human vocal apparatus. b) Vocal gestures (imitative): cross-modal imitations in which the 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 onomato-poeia, 31 (22.1%) as vocal gestures, 16 (11.4%) as relative, 57 (40.7%) as circum-stantial 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
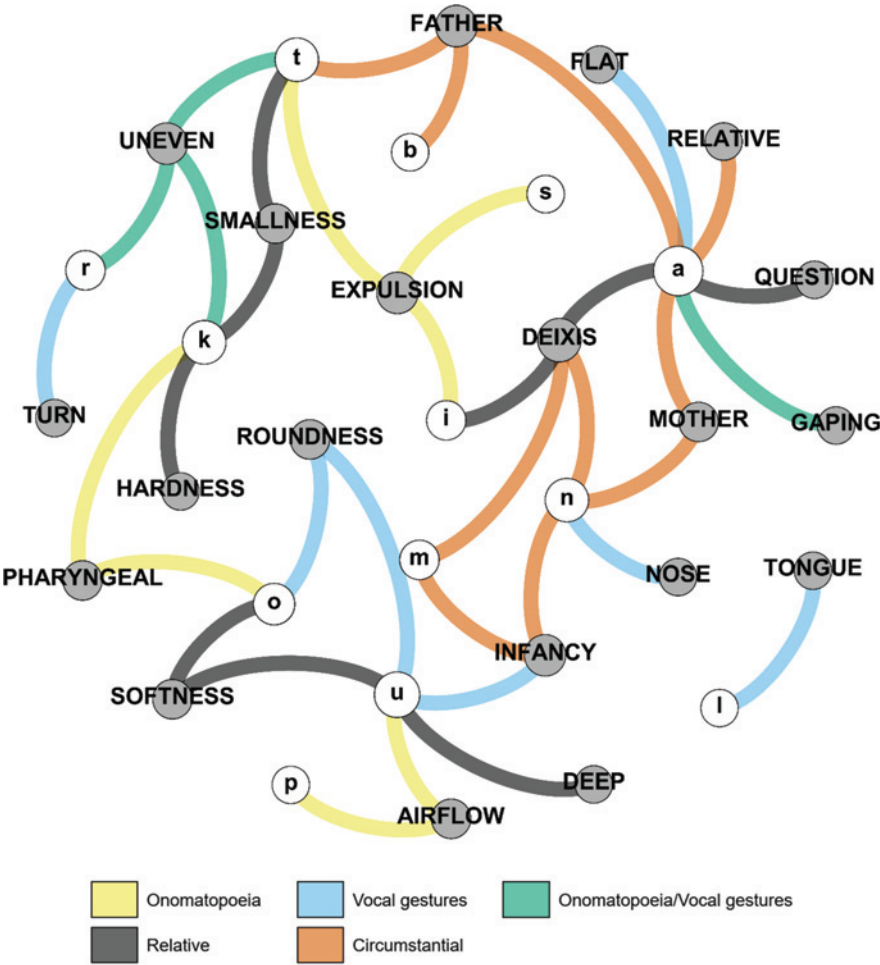
Figure 5: Macro-concepts and the most commonly occurring overrepresented cardinal sound from sound groups with the highest specificity per concept.

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)

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.