Experimental assessment of sound symbolism and evolutionary considerations

Léa de Carolis

N° d'ordre NNT : 2019LYSE2039

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 476 Neurosciences et Cognition

Discipline : Sciences cognitives

Soutenue publiquement le 20 juin 2019, par :

Léa De Carolis

Experimental assessment of sound symbolism and evolutionary considerations.

De la pluralité des principes de justice aux compromis.

Devant le jury composé de :

Sharon PEPERKAMP, Directrice de Recherche, C.N.R.S., Présidente

Ioana CHITORAN, Professeur des universités, Université Paris Diderot, Rapporteure

Padraic MONAGHAN, Professeur d'université, Université d'Amsterdam, Rapporteur

Luca NOBILE, Maître de conférences, Université de Bourgogne, Examinateur

François PELLEGRINO, Directeur de Recherche, C.N.R.S, Co-Directeur de thèse

Christophe COUPÉ, Assistant Professor, Université de Hong Kong, Co-Directeur de thèse

Contrat de diffusion

Université Lumière Lyon 2
École Doctorale Neurosciences et Cognition (NSCo)
Laboratoire Dynamique Du Langage (DDL), UMR5596 CNRS

THESE

Présentée en vue de l'obtention du grade de

DOCTEUR EN SCIENCES COGNITIVES DE L'UNIVERSITE DE LYON

Discipline : Sciences Cognitives
Option Linguistique

Experimental assessment of sound symbolism and evolutionary considerations

Par Léa De Carolis

Réalisée sous la direction de Christophe Coupé et François Pellegrino

Date de soutenance prévue le 20 juin 2019

Devant le jury composé de :

Ioana CHITORAN
Professeure, Université Paris 7 Denis Diderot, Rapporteur

Christophe COUPÉ
Assistant professor, The University of Hong Kong, Codirecteur

Padraic MONAGHAN
Professor, University of Lancaster, Rapporteur

Luca NOBILE
Maître de conférences, Université de Bourgogne, Examinateur

François PELLEGRINO
Directeur de recherche, CNRS, Codirecteur

Sharon PEPERKAMP
Directrice de recherche, CNRS, Présidente du jury

In memory of Matthias

Acknowledgment

To begin with, I would like to thank the members of the jury, Ioana Chitoran, Padraic Monaghan, Luca Nobile and Sharon Peperkamp for accepting to evaluate this work and to contribute to its completion.

Mes pensées se tournent ensuite vers Christophe à qui j'adresse un grand merci et mon infinie reconnaissance pour avoir cru en moi, en acceptant de m'encadrer d'abord durant mon travail de Master, puis à l'occasion de l'élaboration de cette thèse qui a marqué une période significative de ma vie. Merci pour sa rigueur et la qualité de son encadrement. Merci ensuite à Egidio pour son implication dans ce travail, ses conseils, son soutien et son aide indéfectible, surtout lors des dernières semaines consacrées à ce travail. Merci également à François de nous avoir accordé sa confiance, et pour avoir suivi cette aventure avec un regard bienveillant et des interventions toujours enrichissantes. Merci aussi aux membres du Comité de Suivi Individuel Doctoral, Sharon Peperkamp et Rémy Versace, pour m'avoir offert de leur temps, pour leurs précieux conseils et leurs encouragements.

Je voudrais également remercier l'ensemble des membres du laboratoire Dynamique Du Langage et spécialement tous ceux qui m'ont aidé de près ou de loin par maints conseils avisés et prêté main forte tant pour l'élaboration des expérimentations que pour les pré-tests et les passations.

Un merci tout particulier à ces collègues (ces mauvais-perdants !), qui sont devenus un peu plus que des collègues et avec qui j'ai passé des moments inoubliables durant ces cinq années passées au laboratoire : Christian, Egidio et Rabia.

Merci également à tous les participants aux expériences sans qui ces travaux n'auraient jamais vu le jour, ainsi qu'à Enora pour le travail qu'elle a réalisé durant son stage de Master.

Merci encore à tous mes collègues de l'université Savoie Mont-Blanc, et en particulier ceux du bureau 603, qui m'ont accompagnée et soutenue durant cette ultime année, année difficile mais enrichissante où j'ai pu découvrir le monde de l'enseignement supérieur et le plaisir de la transmission.

Enfin, je n'oublie pas toutes ces personnes qui ont fait ou qui font ma vie et à qui ce travail doit son existence. Tout d'abord, une pensée respectueuse envers Mme Vergnaud, professeur d'Allemand au lycée : elle m'a marquée par ses encouragements qu'elle m'a prodigués. Puis, il y eut Djena, que je n'oublie pas malgré les aléas de la vie ; je lui serai toujours reconnaissante d'avoir joué un rôle décisif dans mon choix d'étudier les Sciences Cognitives. Ensuite, je pense évidemment à mon binôme des bancs de la fac, Zoé, qui m'a beaucoup apporté et inspiré durant les années de Licence, et qui a représenté pour moi un « pilier », une aide des plus précieuses pendant ces déterminantes premières années d'études supérieures.

Je pense également, à mes collègues qui m'ont soutenu quotidiennement durant mes années Master et qui, je ne sais par quel miracle, ont réussi à me faire apprécier le monde des sondages téléphoniques : à vous particulièrement Adrien, Camille, Farida, « Jean-Corentin », Nico, Raphaël, Thomas, Victor et Renaud… merci !

Pour terminer, un grand merci à mes proches : à ma famille et en particulier à mon père, ma mère et mon beau-père, pour leur soutien moral pendant ces longues années d'études universitaires (qui sont passées si vite !) et pour les valeurs qu'ils m'ont inculquées. Merci à mon frère Michaël, personne si chère à mes yeux, et qui m'a toujours épaulée ; merci à Steven pour son soutien à toute épreuve et sa présence irremplaçable ; merci à Marion, mon amie d'enfance avec qui je partage les joies équestres ; merci à Thomas et Florian qui me donnent force et courage chaque jour devant les épreuves de la vie ; merci à Anthony pour son amitié mais aussi son aide précieuse et ses relectures ; et enfin ('last but not least' !) merci à Cindy, la meilleure des 'cat-sitters' et bien plus encore…

À vous tous, je dédie ce travail.

Abstract

Sound symbolism, or *motivation* as we will later refer to it, corresponds to the assumption that some words have a natural relation with their significations, instead of an arbitrary one, through their segmental composition. Some evidence stands out from the literature, from cross-linguistic investigations to psycholinguistic experimentations. For example, a closed vowel [i] is more associated to smallness, while an open vowel like [a] is more associated to largeness. This pattern appears in the lexicon of different languages (e.g. Ohala, 1997), as well as in results of associative tasks (Sapir, 1929) with participants speaking different languages and at different life stages. These commonalities (e.g. Iwasaki, Vinson, & Vigliocco, 2007) and their earliness (e.g. Ozturk, Krehm, & Vouloumanos, 2013) enable to formulate the hypothesis that *motivation* may have represented a key-driver in the emergence of language (Imai et al., 2015), by facilitating interactions and agreement between individuals.

This thesis offers several methodological contributions to the study of *motivated* associations. The first study of this thesis aimed at assessing whether animal features (e.g. dangerousness) or biological classes (birds *vs*. fish, based on Berlin, 1994) would be relevant concepts for highlighting *motivated* associations, based on the assumption that animals would have represented suitable candidates for the content of early interactions (as potential sources of food and threats). It raised issues regarding methodological settings which led to the second study consisting in comparing different protocols of association tasks that are found across experimentations. Indeed, in the literature, the settings and population vary from one study to another, and it is therefore not possible to determine which one of the two types of contrasts implied in association tasks is determinant for making associations: either the phonetic one or the conceptual one. This second study permitted to appraise the influence of different protocols by controlling for other sources of variation across the tasks. It also highlighted the need to better analyze the cognitive processes involved in *motivated* associations. This led us to complement our investigation of phonetic and conceptual contrast with a study on the influence of the graphemic shapes of letters, following Cuskley, Simmer and Kirby (2015)'s proposal of an impact of the shapes of letters in the bouba-kiki task. This task is a well-known paradigm in the study of *motivated* associations, based on associating pseudo-words with round or spiky shapes. Cuskley et al. suggested that a spiky shape would facilitate the processing of a pseudo-word that contains an angular letter such as 'k'. On our third study, we considered an implicit

version of the ‘bouba-kiki’ task, namely a lexical decision task, building on a previous experiment by Westbury (2005). In this experiment, spiky and round frames, in which the linguistic stimuli appeared, seemed to facilitate the processing of pseudo-words according to their segmental composition (e.g. spiky frames would facilitate the processing of voiceless plosives like [k]). We manipulated the shapes of letters with two different fonts for displaying linguistic stimuli – one angular and one curvy – and tried to disentangle the respective impacts of frames and of these fonts on the participants’ response times. The results highlighted the importance of taking into account low-level visual processes in the study of *motivated* associations.

Résumé

Un mot et une signification peuvent entretenir une relation naturelle, *motivée*, plutôt qu'arbitraire, via la composition segmentale dudit mot. Ce phénomène est souvent appelé symbolisme sonore, même si nous préfèrerons employer le terme de *motivation* par la suite. Dans la littérature scientifique, des éléments en faveur d'une relation *motivée* apparaissent à la fois dans des analyses translinguistiques et des expérimentations psycholinguistiques. Par exemple, une voyelle fermée telle que [i] est davantage associée à la petitesse qu'une voyelle ouverte telle que [a], davantage associée à une taille importante. Ce schéma apparait à la fois dans les lexiques de différentes langues (e.g. Ohala, 1997) et dans les résultats de tâches d'associations (e.g. Sapir, 1929), avec des participants parlant différentes langues et à différentes étapes de la vie. Du fait de ces éléments communs (e.g. Iwasaki et al., 2007) et de leur précocité (e.g. Ozturk et al., 2013), il est possible de formuler l'hypothèse que la *motivation* a pu être un élément clé dans l'émergence du langage (Imai et al., 2015), en facilitant les interactions et l'accord entre les individus.

Cette thèse offre plusieurs contributions méthodologiques à l'étude des associations *motivées* entre formes phonétiques et significations. La première étude de cette thèse avait pour objectif de déterminer si des caractéristiques associées à des animaux (e.g. la dangerosité) ou à leurs catégories biologiques (oiseaux *vs*. poissons, sur la base de l'étude conduite par Berlin en 1994) pouvaient représenter des concepts pertinents dans la mise en évidence d'associations *motivées*, en se basant sur l'hypothèse que les animaux étaient des sujets récurrents et importants des premières interactions langagières (en tant que potentielles sources de nourriture ou de menace). Cette étude a soulevé des questions méthodologiques, qui ont conduit à une seconde étude, dont le but était de comparer différents protocoles de tâches d'association que l'on peut trouver dans les études sur le symbolisme sonore. En effet, les protocoles et les populations étudiées varient d'une étude à l'autre, et il est ainsi difficile de déterminer quel est le contraste le plus déterminant pour la mise en valeur expérimentale d'associations *motivées* : le contraste phonétique, ou le contraste conceptuel. Cette deuxième étude a ainsi permis d'apprécier l'influence de différents protocoles en contrôlant d'autres sources de variations à travers les différentes tâches. Elle a aussi permis de mettre en évidence la nécessité d'étudier davantage les processus cognitifs impliqués dans les associations *motivées*. Ainsi, nous avons poursuivi notre investigation en nous tournant vers l'influence de la forme des lettres, un facteur

potentiellement déterminant dans les tâches ‘bouba-kiki’, comme l’ont proposé Cuskley et al. (2015). Bouba-kiki est un paradigme très répandu dans l’étude des associations *motivées* et consiste à associer des pseudo-mots avec des formes pointues ou arrondies. Cuskley et al. ont proposé qu’une forme pointue faciliterait le traitement d’un pseudo-mot contenant une lettre anguleuse, telle que ‘k’. Dans notre troisième étude, nous avons adopté une version implicite de la tâche bouba-kiki, plus précisément une tâche de décision lexicale, en nous basant sur une étude antérieure de Westbury (2005). Dans cette expérience précédente, des cadres pointus et arrondis, dans lesquels apparaissaient les stimuli linguistiques, facilitaient le traitement de pseudo-mots en fonction de leurs compositions segmentales, par exemple les formes pointues accéléraient le traitement d’occlusives non-voisées telle que [k]. Nous avons manipulé les formes des lettres via deux polices de caractères différentes, une anguleuse et une curviligne, et avons ainsi essayé de démêler les impacts respectifs des formes des cadres et des polices sur les temps de réponse des participants. Les résultats ont mis en lumière l’importance de prendre en considération des processus visuels de bas-niveau dans l’étude des associations *motivées*.

Table of contents

List of the tables

List of the figures

General introduction

In linguistics, *motivation* or Cratylism[1] is the idea according to which there is a direct, natural, fundamental link between words and concepts they refer to. *Motivation* is in opposition to *arbitrariness* which defines the *linguistic sign* – the relation between a *signifier* (words) and a *signified* (concepts) – as arising from *convention* (Saussure, 1916). For example, the difference in meaning between the French words 'pain' and 'bain' does not arise from the acoustic difference induced by voicing[2] itself: the meaning conveyed by 'bain' is not *more voiced* than the meaning conveyed by 'pain'. Thus, no meaning is associated to the sound contrast itself between the segments[3] [p] and [b] and their respective meanings are arbitrary (Nuckolls, 1999). More generally, in semiotics, *motivation* and arbitrariness are in opposition when establishing the relation between a signifier and a signified. This opposition has been expressed by Peirce's hierarchy of sign.

According to Peirce (1931) there are three types of signs: the *icon*, the *index* and the *symbol*. The *icon* is a sign that refers to the referent through a relation of similarity (in Everaert-Desmedt, 2011; Keller, 1998). The *index* does so through a causal relation, or more generally, a contextual contiguity. Finally, the *symbol* is a sign which refers to a referent through convention or out of habit.

Peirce's terminology can be applied to evidence of *motivation*. First, there is the imitation of a sound by another sound (e.g. onomatopoeias), which corresponds to an iconic relationship based on acoustic similarity. Second, with respect to an indexical relationship, just as the weather vane indicates the wind direction, a high-pitched tone can denote an object of a small size because the latter tends to make higher-pitched sounds. Hence, one dimension (pitch[4]) can represent another one (size) through causality. Third, the *symbolic relationship* can be related to the conventional relation between a linguistic signifier and a signified, and thus to arbitrariness.

---

[1] This word originates from Plato's Cratylus, in which two contrastive positions about language were juxtaposed, motivation and arbitrariness.
[2] [p] and [b] are articulated in the same way apart from a difference of voicing, [b] being articulated with a vibration of the vocal folds.
[3] Although the phoneme is the smallest linguistic unit that distinguishes words, the more neutral word *segment* will be preferred in this thesis since there is no evidence regarding the linguistic level at which motivated associations appear, i.e. whether it is phonetic or phonological.
[4] In the entire document the word *pitch* refers to the fundamental frequency of language sound (in Hz), while the word *frequency* refers to the frequency with which a sound (i.e. a segment) occurs in a language.

In the literature, two terms are used to designate the study of *motivated* relations: *iconicity* (e.g. Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015) and *sound symbolism* (e.g. Imai, Kita, Nagumo, & Okada, 2008). Iconicity corresponds to natural *motivated* relations (as onomatopoeias). The term sound symbolism is more broadly inclusive but is also ambiguous in view of Peirce's terminology. Indeed, the word *symbol* is included in the expression *sound symbolism* and it refers concomitantly to an arbitrary relation. The symbolic relation in Peirce's terminology does not correspond to the symbolic relation in *motivated* relations, but it is possible to consider that they are somehow related if we consider that sound symbolism rather refers to a *relative motivated* relation. Gasser, Sethuraman and Hockema (2011) proposed a dichotomy between *absolute* iconicity (i.e. a natural relation, e.g. onomatopoeia) and *relative* iconicity ('*related forms are associated with related meanings, as when a contrast between the vowels [i:a] depicts an analogous contrast in magnitude iconicity*', Lockwood & Dingemanse, 2015, p. 3). One particular category of words corresponds to relative iconicity: *ideophones* (also named *expressives* or *mimetics*[5]). In this category, words *evoke* their referent by their segmental composition and through analogies with linguistic sounds. For example, in Japanese, 'buruburu' describes something shaking or trembling (Gomi, 1989). According to Lockwood and Dingemanse (2015), ideophones depend on culture, to some extent, and hence are partly conventional.

Overall, the precise nature of a relation can be difficult to determine and terminologies are not very consistent across authors. Hence, in this thesis, the more generic word *motivated* will be preferred to *iconic* and *symbolic* – in order to not make a stand – mostly when the nature of a relation is ambiguous.

Language is undoubtedly largely arbitrary, but there is concomitantly evidence for *motivated* relations. While arbitrariness allows efficient communication associated with compositionality, double articulation etc., *motivation* may have represented a key driver in the emergence and evolution of language. As claimed by Imai et al., (2015, p. 2) '*sound symbolic words may thus be "fossils" from earlier stages of language evolution, when sound symbolic links facilitated the rapid development of a common lexicon in human protolanguages.*' Indeed, it is unlikely that language appeared immediately fully complex, and one of the mainstream views in language evolution research is a two-step scenario in which the doubly articulated system evolved from a more basic one (Tallerman, 2011). Hence, the early stage of our

[5] In this thesis, the word *ideophone* is used as a generic term for this type of word; the word *mimetic* will refer specifically to Japanese ideophones (and also includes Japanese onomatopoeias).

communication system may have had required direct and iconic relations between signifiers and signified before convention could arise and spread among speakers. More precisely, the assumption is that producing *motivated* cries to naturally denote referents would have allowed individuals to agree on signifiers, via the consistency of productions. For example, systematically producing the same cry across individuals for a given threatening stimulus would have helped convention to emerge. After multiple repetitions, the cry only could have become enough to trigger the relevant escape response. Contrariwise, if different individuals used different cries for a given stimulus, it would have been more difficult for convention to arise. *Motivation* can underlie systematicity.

However, *motivated* relations, between cries and the objects they refer to, only exist via the *interpretant* – the communicating human being and his set of cognitive representations. In order for individuals to agree on signs, shared representations in the mind of different individuals are required. This raises the question about the nature of the cognitive correspondences underlying the relation between signifiers and signified. To this regard, numerous studies provide experimental evidences on the propensity of common *motivated* associations in people speaking different languages, coming from various cultural environment and at different stages of life (see sections 1.2 and 2.5 for references). The literature also explores the cognitive mechanisms at play in these *motivated* associations, considering them as a particular instance of a more general cognitive phenomenon, namely crossmodal correspondences, i.e. associations between different modalities (see section 3 for references).

This preamble aimed at clarifying the terminology and contextualizing sound symbolism within a theory of signs, language emergence and research in cognitive science. Consequently, this introduction will open with a first section in which different phenomena that exist in languages worldwide will be reported, as well as the experimental approaches assessing *motivated* relations through psycholinguistic studies. The second section will focus on hypotheses about the origin of *motivated* associations, namely gestural and size coding, as well as emotional approaches. In the third section, potential explanatory brain mechanisms underlying *motivated* associations will be exposed. A last section will present some methodological limitations found across the literature and will introduce the three studies conducted in this thesis and their purposes.

1. Evidences of *motivation*

1.1. Motivated *manifestations in languages*

Three categories of words which contradict sign arbitrariness can be found worldwide in languages: onomatopoeias, ideophones and phonesthemes.

Onomatopoeias are words that phonetically imitate the sounds they refer to (Tanz, 1971), like animal's sounds. For example, the rooster's crow is 'cocorico' in French, 'kokekokkō' in Japanese, 'chicchirichi' in Italian. Though these onomatopoeias look alike, their cross-language variability can be explained, although non-exclusively, by the sounds contained in the inventory of their respective language, or by the phonotactic rules of the latter.

Ideophones, the second category of words, have the particularity to evoke their referents through perceptuomotor analogies, and can be defined as '*vivid sensory words*' (Dingemanse et al., 2015, p. 605). For example, in Japanese, 'koro' is a light object rolling once while 'goro' is a heavy object rolling once (Imai et al., 2008). Hence a weight difference is expressed by a difference in voicing i.e. the vibration of the vocal folds express larger weight. As for the words 'korokoro' and 'gorogoro', they describe the repetition of the event through syllabic repetition (though reduplication is usual in Japanese mimetics). Ideophones are very common worldwide but are surprisingly absent in western European languages (Nuckolls, 1999). In Japanese, mimetics are widely used and there are several types of them, including *giseigo* (mimicking animal sounds and human voices) and *gitaigo* (mimicking manners or states) (Iwasaki et al., 2007). Since the former are imitation of voices and cries by linguistic sounds, they actually correspond to onomatopoeias, whereas the latter correspond to ideophones. According to Iwasaki et al. (2007), a continuum of *motivation* actually exists in mimetics: *giseigo* words have a direct resemblance with their referees and are thus highly *iconic* (absolute iconicity), while *gitaigo* words have more of a *symbolic* relation (sound symbolism) with their referents[6].

The third category, *phonesthemes,* are sequences of segments that are associated with meanings in a given language. For example, the cluster /gl/ appears consistently in words related to light in English (see Table 1). More interestingly, phonesthemes can behave as phonemes because they can also have a contrastive function (for example, /fl/ appears in words related to movement in contrast with words related to light). More generally, phonesthemes offer an

[6] Other types of mimetics exist in Japanese but will not be presented in this thesis.

evidence of systematicity (or redundancy) – which is the regularity of occurrence of clusters, morphemes or segments in a given semantic field. These regularities are language-specific.

Table 1. Examples of phonesthemes presented by Bolinger (1950).

		Contrastive phonesthemes	
		Light (/gl/)	Movement (/fl/)
Consistent[7] phonesthemes across words	'intermittent' (/ɪtəʳ/)	gl-itter	fl-itter
	'steady' (/əʊ/)	gl-ow	fl-ow
	'intense'(/eəʳ/)	gl-are	fl-are

According to Dingemanse et al. (2015), '*language-specific distributional regularities are likely instances of systematicity, whereas form–meaning mappings that recur across languages and rely on perceptual analogies are likely instances of iconicity*' (p. 607). Hence, phonesthemes – the distributional regularities – are specific to a semantic field without representing the referents either iconically or sound-symbolically.

On the one hand, the iconic or symbolic relation of ideophones is not always obvious – contrary to onomatopoeias (or at least most of them). On the other hand, one may believe that some phonesthemes are *motivated*, with the *motivated* origin having been obscured. We can rely on the commonalities found across languages to establish a possible *motivated* origin (for example, if worldwide languages present voiceless consonants for light objects and voiced ones for heavy objects, this would be an argument in favor of a *motivated* origin of the mimetics 'korokoro' and 'gorogoro').

In the quest for commonalities across languages and for consistencies within languages, the next section focuses on studies about *motivated* associations involving segments. The two purposes of the studies compiled in the next section are: 1) to expose consistent and/or common associations within or across languages through quantitative and linguistic analyses; 2) to assess experimentally the reality of *motivated* associations in humans using words that come from natural languages. In both cases, studies rely on a large range of concepts (e.g., size, emotions).

[7] Complements like [ɪtəʳ] (in 'glitter' or 'flitter') are not often meaningful and interchangeable (Bergen, 2004).

1.2. *Ecological approaches to the study of* motivation

1.2.1. Lexicon analyses

1.2.1.1. Size and distance

Ohala (1997) reported different evidences of segments associated to size concepts across languages. In Ewe, Yoruba, Spanish, Greek, English, Irish and French, words and morphemes expressing smallness contain more high front vowels (e.g. [i]) and voiceless consonants (e.g. [t]), and in Ewe, Yoruba, Spanish, Greek and French there are more low back vowels (e.g. [u]) and voiced consonants (e.g. [b]) in words and morphemes expressing largeness. The author explained that expressions about size exploit acoustic pitch characteristics which inversely vary with the emitter's size (i.e. the higher the pitch, the smaller is the emitter). However, there are counterexamples to this phenomenon (e.g. the words 'small' and 'big' in English, considering the vowels).

Tanz (1971)[8] studied words signifying 'here' and 'there' in several languages from different families, focusing on vocalic differences (i.e. ignoring words that differed *only* in consonants). She found three different types of contrast: 1) a difference in vowels – words for 'here' always contained [i], which was opposed to [a] or [o] in 'there'; 2) a difference of an entire syllable, the syllable in the word meaning 'here' containing a more front or high vowel; 3) an extra syllable in words meaning 'there', this syllable always containing [a]. Examples of these three types of contrast are compiled in Table 2.

Table 2. Examples of words for 'here' and 'there' in different languages (Tanz, 1971).

	Languages	'here'	'there'
Vowel contrast	Kanada	illi	alli
	Malay	sini	sana
Change in one syllable	Aztec	nika-n	onka-n
	Indonesian	disine	disitu
Addition of one syllable	Arabic	huna	hunaka
	Japanese	koko/soko	asoko

Tanz outlined parallel phenomena involving time (e.g. more [i] and [ɪ] in English verbs in the present tense than in the past tense) and social distance (e.g. 'anata' and 'kimi' both mean

[8] This study contains no statistical analyzes.

'you' in Japanese, the first one being more formal and distant). She also proposed articulatory mechanisms as a potential explanation, [i] being the most constricted[9] vowel while [a] being the least constricted one.

Haynie, Bowern and LaPalombara (2014) investigated the issue of *motivated* words denoting size and distance in 120 Australian languages. The authors gathered words related to distance (e.g., here, there, near, far) and size (e.g., small, large, skinny, fat) that were expected to exhibit specific segments, according to their meaning. They found significant differences in the proportion of some linguistic features between the basic vocabulary and the vocabulary related to size and distance. More precisely, words related to smallness and nearness contain more palatal consonants and front vowels, while, although less strongly, largeness and farness are associated to back vowels and velar consonants. These effects are relative to segment positions (i.e. generally, these segments appear more in initial or final position of the words) and are more consistent for consonants than for vowels. There are also differences across languages in the way they make distinctions about size and distance (e.g. some languages exhibit specific patterns for large distances and for small sizes) and in the segments that carry these distinctions (e.g. the segments indicating nearness or smallness are palatal consonants in 12 languages and front vowels in 17 languages). Overall, however, the *motivated* distinctions are consistent across languages when they appear (in 54% of the tested languages).

This approach can be questioned on several aspects. First, the authors looked for associations based on previous studies and hence with respect to specific segments or features which were grouped in categories, depending on their expectations. This clustering may not be relevant for some of the targeted Australian languages. For example, they found that words for smallness and nearness also contained segments denoting largeness and farness, which may seem contradictory. Further analyzes revealed that one type of vowel was underlying this effect: high back vowels were more present in words denoting smallness and nearness than in basic vocabulary, even though they were rather expected in words denoting largeness and farness because of their back articulation. This may reveal specificities depending on languages. Moreover, the authors did not include [a] because of its central position, while it might have been relevant based on previous studies, such as Sapir (1929)'s. Second, they included a large amount of languages that belong to the Pama-Nyungan family – 104 languages out of 120 in the total sample. The authors conducted post-hoc analyzes which indicated that the distribution

[9] More constricted means leaving less space between the tongue and the palate.

of segments for size and distance was not consistent across the languages of this family. Hence, if there were a phylogenetic impact, it would only have a weak influence.

1.2.1.2. Emotions

Fónagy (1961) analyzed the segmental content of poetries that were previously evaluated by participants as aggressive or tender. He found specific frequencies of segments: aggressive poems contained more voiceless consonants like [k] and [t], while tender ones contained more sonorants like [l] and [m]. He also conducted a later study with Hungarian participants, asking them to evaluate some segments on several dimensions (Fónagy, 1983). According to their judgements, [i] is small, agile, gentle, nice etc. whereas [u] is corpulent, obtuse, sad, bitter, strong and dark.

Adelman, Estes and Cossu (2018) analyzed thousands of words in five languages: English, Spanish, Dutch, German and Polish. They examined the possibility that phonemic composition could predict the emotional valency ratings of words. They found significant effects for each language ($p < .001$) with effect sizes (R^2) varying from 1.40 to 4.32%, depending on the languages. A large proportion of segments significantly bears a valency, either positive or negative (from 21 to 45% of segments). The authors also tested whether these effects could be better explained by both sub-phonemic features and segments rather than segments alone. They found significant but weaker effects for linguistic features (from 0.35 to 1.79%), while segments were still significant with higher effect sizes (from 0.81 to 2.54%). Hence, segments predict valency more strongly than phonetic features. For example, in English, [f] appears more in positive words and [s] in negative ones. Thus, the feature *fricative* cannot predict alone valency ratings in this case. In addition, the emotional values of segments differ between languages: while [d] and [n] appear to be '*positive*' segments in Spanish, they are '*negative*' in German; similarly to German, [d] is negative in English and [j] is positive in both, while [n] is neutral in English. Segments thus seem to bear emotional values depending on the languages they belong to, and more precisely on the phonological systems, phonotactic rules and lexicons of these languages.

The *motivated* relation between meanings and segments thus seems to vary depending on the languages. However, a relative consistency also exists across languages, especially on words expressing size and distance, as revealed by Haynie et al. (2014), Ohala (1997) and Tanz (1971)'s studies. Other semantic fields have been explored, like basic vocabulary and animal names, which will be the subjects of the following sub-sections. These studies will help us learn more about such consistency across languages and speakers.

1.2.1.3. Diverse concepts

Blasi, Hammarström, Stadler and Christiansen (2016) published a study conducted on 62% of the languages of the world, covering 85% of the lineages of the world. They looked for specific segment frequencies in the 100-word Swadesh list (words forming part of the basic vocabulary of a language, e.g., tongue, bone, dog, etc.). For each concept, they compared the frequency of occurrence of each segment with the baseline frequency of occurrence of these segments in the words for other concepts[10]. Some interesting significant associations found across languages are reported in Table 3. The authors detected 74 '*signals*' (i.e. a signal corresponds to a segment that has a specific frequency for a given concept which is statistically significantly larger or smaller *'in contrast to their baseline occurrence in other words'*, p. 10819). They found segments that were specifically associated to some concepts (e.g. [l] for 'tongue'), as well as segments unlikely to appear in the words for some concepts (e.g. [k] for 'tongue').

Table 3. Vowels and consonants associated to some concepts across languages in Blasi et al. (2016)'s study.

Concepts	Vowels	Segment	Consonants	Segment
Little	High front vowel, rounded and unrounded	i	Voiceless palato-alveolar affricate	C
Round			All varieties of r sounds	r
Tongue			Voiced alveolar lateral approximant	l
Nose			Voiceless and voiced alveolar nasal	n
Breasts and Mother	high back vowel	u	Bilabial nasal	m
Fish	Low central vowel, unrounded	a		

This study shows that some basic concepts tend to contain or not to contain specific segments, worldwide and independently of phylogenetic lineages or geographical dispersion.

1.2.1.4. Animal names

Berlin (1994) analyzed the phonemic composition of words referring to fish and birds in Huambisa – a language spoken by the Huambisa people in Peru. He found different patterns

[10] The authors used a simplified phonological model in which segments were grouped, resulting in 34 categories of consonants and 7 categories of vowels. For example, the segment 'u' contained high back vowels, both rounded and unrounded (i.e. [u] and [ɯ], respectively).

of occurrence between these two groups of names, with some differences that appeared according to the syllabic position in words. As an example, bird names contain more [i] in comparison to fish names, which conversely contain more [a]. As regards consonants, final syllables of bird names contain more obstruents while those of fish names contain more nasals and continuants. Interestingly, there is also an internal pattern within biological classes expressing size – e.g. large birds contain more [a] and small fish contain more [i] – which is consistent with studies on size contrasts reported above.

Berlin presented a selection of the previous names by pair to students (the country it took place in or the language they spoke were unfortunately not stipulated). Each pair was constituted of names of a bird and a fish and he asked students to guess which one was referring to a bird – specifying that the second one referred to a fish (e.g. 'chichikía' vs. 'katan', the first one being the bird). Participants were able to guess the proper meaning with a performance of 54% of correct answers, which is significantly higher than chance.

Besides the existence of significations at the phonemic level found in the several studies reported above, this last study demonstrates that – to some extent – it is possible to guess the signification of a foreign word – presented within a pair – on the basis of its phonemic composition (while knowing the meaning of the second word). One can wonder what the performance would look like if the second meaning was not stipulated, since the presence of a contrast may possibly underlie the choice. The following section covers psycholinguistic studies which exploited the same type of experimentations that consists in showing pairs of foreign words. Some of them used real words of diverse languages, and others pseudo-words created for the sake of experimentation.

1.2.2. Experimentations that used real words

Brown, Black and Horowitz (1955) conducted an experimentation with English speakers. They presented to them pairs of antonyms from three different languages – Chinese, Czech and Hindi – which denominated sensitive experiences (e.g. 'hot' and 'cold'). On one page, the antonym pair appeared in English; on the other page, the same pair was presented in all the three other languages. Participants had to find the proper matching. These pairs were presented in a written form and were also pronounced by experimenters. The percentages of correct identifications exceeded chance level significantly for the three target languages: 59.6% for Hindi, 58.9% for Chinese, 53.7% for Czech. Authors suggested a potential pronunciation

bias and thus replicated the procedure by showing the pairs only in their written form. Results were even stronger: 60.7% for Hindi, 61.9% for Chinese, 61.9% for Czech.

Similarily, Kunihira (1971) obtained performances higher than chance with American students who guessed the meaning of Japanese words presented by pair of antonyms, in three different conditions: in Romanized written form only (i.e. in Latin alphabet), orally without expressive voice quality and orally with expressive voice quality (spoken words were also accompanied by written forms). The percentages of correct guesses were respectively 57.39%, 58.35% and 63.13%[11].

Word length can also be an indication of meaning: concrete and abstract words for example differ on word length in English (Reilly, Hung, & Westbury, 2017). The longer the word, the more abstract it tends to be (e.g. 'information' and 'cat'). Reilly et al. (2017) tested the possibility to guess whether a foreign word is abstract or concrete. English speakers indeed guessed above chance level the concreteness of words of four languages out of eight – namely Dutch (55% of accuracy), Hindi (52%), Russian (56%) and American Sign Language (ASL; 63%). These four languages exhibit, in fact, a correlation between word length and concreteness. For the first three languages, the pattern corresponds to the expected one: the longer the word, the more abstract it is. However, ASL presents a reverse pattern: concrete concepts are longer to express. However, another language presents the same pattern as the first three languages – Hebrew – but did not elicit any significant effect. The three other tested languages – Arabic, Korean and Mandarin – neither present a length pattern relation, nor they induced performance higher than chance. However, the iconic origin of this phenomenon is to be demonstrated since other factors could explain these results. For example, the two languages that exhibited the highest rates, Russian and Dutch, share a morphological property with English: the derivational morphology (i.e. the affixation of concrete words to produce abstract ones, e.g. 'friend' and 'friendliness')[12].

Imai et al. (2008) conducted a study with British English and Japanese speakers using mimetics created for the purpose of their experimentation[13] – thus avoiding a potential bias toward Japanese speakers. The pseudo-mimetics were created in order to convey information about two dimensions: speed and weight. They were made so that they matched video clips

[11] Based on my own calculations.
[12] Arabic, Korean and Mandarin grammatical systems differ in this regard, and those of Hindi and Hebrew are not described in Reilly et al.'s study.
[13] Mimetics were created from Hamano's analysis (1998, in Imai et al., 2008) of real Japanese mimetics.

displaying different ways of walking. For example, 'batobato' corresponded to the action of running with heavy steps – [t] expressing hitting and [b] heaviness – and 'nosunosu' corresponded to slow walking with very heavy steps – [n] expressing slowness and [s] friction. The task was either a matching judgment or a forced-choice task. In the matching judgment task, participant had to evaluate the matching between one pseudo-mimetic and one action using a scale from 1 to 7. The authors found the same orientation of responses in both groups of subjects, although the effect size was higher and the *p* value smaller in Japanese in comparison to English speakers ($d = 6.05$, $p < .001$ *vs*. $d = 0.60$, $p < .05$, respectively). For the forced-choice task, participants had to select the action – out of two – that was best depicted by one pseudo-mimetic. Japanese speakers consistently selected the action that matched the pseudo-mimetic (100%). English speakers did so with a smaller effect size that still significantly exceeded chance level (64%). Differences in effect size may be explained by culture and a greater sensibility to the segmental composition of pseudo-mimetics among Japanese speakers, but the results also show that the pairings do not require language exposure.

While the previous studies depicted commonalities across languages or speakers, the last one exhibited a quantitative difference between speakers, namely English and Japanese speakers. This may be explained by the greater proportion of *motivated* words in Japanese (chiefly, the number of mimetics). To further investigate this issue, the following section focuses on differences imputable to cultural and linguistic backgrounds.

1.3. *Incidence of cultural and linguistic backgrounds*

Iwasaki et al. (2007) compared English and Japanese speakers on their judgments about mimetic words that either mimic laughter (corresponding to giseigo mimetics) – loud (e.g. 'keta-keta') or quiet (e.g. 'kusu-kusu') – or manner of walking (corresponding to gitaigo mimetics) – capturing its auditory dimension (e.g. 'bata-bata') or its visual or affective dimension (e.g. 'yota-yota'). Participants had to evaluate each mimetic on different semantic dimensions for laughing (e.g., graceful vs. vulgar, excited vs. calm) and manner of walking (e.g., noisy vs. quiet, purposeful vs. aimless). Results are summarized in Table 4.

Table 4. Summary of Iwasaki et al. (2007)'s results. JS and ES stand for Japanese Speakers and English Speakers respectively. There was 24 mimetics and 20 dimensions for laughing (e.g. 'loudness'); 28 mimetics and 21 dimensions for walking (e.g. 'gracefulness').

	Laughing	Walking	Examples and precisions (only for laughing)
Number of correlated ratings on mimetics between JS and ES	12 out of 24	7 out of 28	'kusukusu' is the most correlated one
Number of correlated dimensions between JS and ES	6 out of 20	2 out of 21	'Loudness', 'openness of the mouth', 'continuity' and 'resonance' are positively correlated; 'beautiful' and 'graceful' are negatively correlated
Number of common associations between vowels and dimensions for both JS and ES	7 out of 20	0 out of 21	Two associations are similar: [a] is more 'amused' and 'cheerful' than [u] Five associations are more complex for JS than ES, e.g. for ES [a] is 'louder' than [u], while for JS [a] is 'louder' than [e], which is 'louder' than [u]

The languages of the world represent a treasure trove in the study of *motivation* and can provide hints about these associations (e.g. more [i] in bird names in Huambisa). However, using real languages also constitutes a limit in an experimental perspective (real words do not only result from *motivated* associations but also from arbitrariness, phonotactic rules, etc.). The use of pseudo-words permits the elaboration of multiple possibilities of segment combinations while controlling for linguistic constraints. Along these lines, a famous experimental task called 'bouba-kiki' has been used and replicated with a variety of vocalic and consonantal combinations, as well as with speakers of different languages from different continents. This paradigm brings insights on considerations about cultural and linguistic issues.

1.4. *Bouba-kiki – an experimental paradigm to seek proof of a universal phenomenon*

In the original experiment setup by Köhler (1929; 1947) the pseudo-words 'maluma' and 'takete' were used – the pseudo-words 'bouba' and 'kiki' later supplanted them. This experiment consisted in the presentation of these two pseudo-words and of two visual shapes: a round one and a spiky one. Participants were asked to associate one of the pseudo-words to

one of the shapes. Köhler reported consistency across participants in associating 'maluma' with the round shape, and 'takete' with the spiky one. While Ramachandran and Hubbard (2001b) claimed to find this effect in 95% of the population[14], a recent metanalysis conducted by Styles and Gawne (2017) – involving 16 studies with a total of 558 participants speaking different languages – estimates the prevalence of the expected associations from 84 to 94% of participants. This very simple experiment has been repeated with people speaking different languages and is the focus of this section.

In 1961, Davis studied the bouba-kiki effect with 335 African children from Tanzania (aged 8-14), whose mother tongue was Kitongwe (a bantu language) and who learned Swahili at school but not English (note, however, that Swahili is written in the Latin alphabet). Their results were compared to those of 281 English children (aged 11-14). There were two different conditions: pseudo-words were either presented visually and orally, or only orally. Overall, the author found the same matching pattern in both groups: 'takete' was associated to spiky and 'uloomu'[15] to round. However, response orientation was weaker in Kitongwe-speaking children and this may be mostly explained by an order effect, which is particularly present in this population. The pseudo-word 'takete' was always the first name to be pronounced, followed by 'uloomu', while the order of presentation of the shapes was counterbalanced. Significant effects appeared in Kitongwe-speaking children when the spiky shape was displayed on the left, otherwise it was never significant (the response orientation was nonetheless not contradictory).

Another evidence comes from Bremner et al. (2013). The authors tested the bouba-kiki effect with Himbas, people from Namibia and speakers of Otjiherero, who have no written language (five participants were excluded because they had been to school). They found the same results as other studies conducted with Westerners (at least for shapes/pseudo-words matchings[16] in 28 participants out of 34)[17]. The authors thus concluded that the bouba-kiki

[14] This amount should be considered cautiously since there is no description of the experiment that led to this result. More precisely, nothing is known about the number of people surveyed, their culture, the language they spoke, the experimental conditions, and so on.

[15] 'maluma' is a real word in Kitongwe. For this reason, the authors changed the pseudo-word for 'uloomu'.

[16] They presented differences – in comparison with Westerners – in the pairings implying different types of water (still and sparkling – for which there was no preferential mapping, whereas Westerners associate sparkling with spiky) and chocolate (sweet and bitter – for which they presented the reverse pattern, namely bitter with round). Unfortunately, they did not test the pairings between the pseudo-words and the different types of water and chocolate.

[17] It is important to note that according to Styles and Gawne's analyses (2017) Otjiherero does not distinguish voicing in plosives. The distinctive associations between [buba] and [kiki] should then only be interpretable in terms of place contrast rather than in terms of voicing contrast – place having been demonstrated as a source of variation of the associations (Nobile, 2015).

effect does not depend on the shape of letters (this will be further discussed in section 2.1) and is a universal phenomenon which could stem from phylogeny.

Rogers and Ross (1975) reported a counterexample: no effect was found in the Papua-New Guinea community called Songe. However, nothing was reported in their paper regarding either the methodology or the participants – and more importantly regarding the language they spoke. According to Maurer, Pathman and Mondloch (2006), one's knowledge of a language underlies her possible matching between pseudo-words and objects. Hence, it is possible that the sounds included in 'maluma' and 'takete' are not meaningful to Songe because some segments may not exist in their language. Imai et al. (2008) also support the existence of both universal associations and language-specific ones, relying on particular segments of a language. The following study further assesses this explanation.

Styles and Gawne (2017) also reported an absence of the bouba-kiki effect in another community. They tested speakers of Syuba in Nepal. They recorded a speaker of another dialect – with mutual intelligibility – pronouncing the pseudo-words 'kiki' and 'bubu', respecting the initial syllable tone implied by each consonant – which resulted in [khíkhí] and [bùbù] respectively. They expected an enhancement of the bouba-kiki effect due to pitch-shape correspondences (high-pitched sounds are associated with spiky and low-pitched sounds with round, see Marks (1987) and Parise and Spence (2009)'s studies reported below in section 3.4.5.). They presented two objects, a spiky one and a round one, and participants had to choose one of them given a pseudo-word presented orally through headphones. Results revealed no orientation in the choices (46% of agreement in the associations – compared to 92% obtained with English speaking participants using the same procedure and material).

In order to explain the previous discrepancies, that appear only in two studies within an otherwise widely consistent literature, Styles and Gawne (2017) looked for a linguistic explanation, more precisely coming from the sound structure of the languages spoken by the participants. They used the PHOIBLE dataset (Moran, McCloy, & Wright, 2014), which contains '*2,160 segments from 1,672 documented languages*' (Styles & Gawne, 2017, p. 3). First, the authors found that the most widespread segments are those mostly used in studies about *motivation*: [p, b, m, t, d, n, k, g, i, e, a, o, u]. These segments permit to contrast highly discriminable pseudo-words, and some of them turn out to be involved in the most robust *motivated* associations (voiceless plosives and sonorants are strongly associated to spiky and round shapes, respectively). Second, one major bias in the studies about *motivation* comes from the fact that most experimenters and participants are WEIRD (Western, Educated,

Industrialized, Rich, and Democratic), which represent a source of bias according to Henrich, Heine and Norenzayan (2010). Hence, the segments chosen within a specific study may not actually exist in the language spoken by the participants, which could compromise the whole study. Conversely, few or no studies tested segments of lower frequencies (such as ejectives, retroflex…). To go further, the authors considered recent findings about Hunjara (the language spoken by Songe people, with respect to Rogers and Ross' study in 1975. It turns out that this language does not contain the sounds [l] or [t^h]. Similarly, the pseudo-words used with Syuba speakers violate their language: 1) [k^h] does not occur word-medially; 2) [u] never occurs at the end of bi-syllabic words; 3) the tone should have been neutral in the second syllable. Thus, the absence of results may be due to phonetic and phonotactic violations of the participants' languages. Consequently, there are two possibilities: either the associations do not exist in these populations, or they actually require other stimuli to be revealed.

In addition to phonetic inventories and phonotactic rules, other cultural factors can influence the associations. Indeed, Chen, Huang, Woods and Spence (2016) studied the bouba-kiki effect in American (US) and Taiwanese participants while varying three visual parameters: spikiness, amplitude and frequency (i.e. number of branches) of shapes. The task consisted in choosing one pseudo-word ('bouba' or 'kiki') for a given visual stimulus. Overall, the increase of one parameter induced a gradual shift from 'bouba' to 'kiki' – 'kiki' was spikier, more elongated and had more extremities – in both groups. But there were also differences: amplitude influenced more Americans than Taiwanese, while spikiness influenced more Taiwanese than Americans. The authors explained this discrepancy with differences in visual processes, which would be more holistic in Taiwanese (relying on the global spikiness of the contour) and more analytic in Americans (relying on branches that are distinctly processed). Hence, they concluded that characteristics of shapes must be consciously chosen, having in mind the potential differences that visual frequency and amplitude can elicit on performance, at least in different cultures. More precisely, the number of branches and their amplitude have to be considered.

Similarly, Nobile (2015) tried to disentangle the effects of different visual and phonetic features. He proposed pairs of pseudo-words with pairs of shapes to participants and asked them to associate one of the pseudo-words to one of the shapes. Pseudo-words differed on voicing, manner, nasality or place of articulation, while shapes differed on spikiness, angle acuity (obtuse *vs*. acute), continuity and density. Several results ensued from his experiments. For example, voiced consonants are associated with curved, obtuse and continuous features, while

it is the reverse pattern for voiceless consonants. Following this line of investigation, the next section focuses on studies that introduce variations on phonetic features in order to determine those at play in *motivated* associations.

1.5. *Phonetic features and segments involved in* motivated *associations*

Several studies attempted to determine the phonetic features or segments implied in bouba-kiki associations. Some concluded to a greater vocalic effect (e.g. Tarte, 1974[18]) but the majority concluded to a superior consonantal one (e.g. Nielsen & Rendall, 2011). Aveyard (2012) also found a stronger effect of consonants in comparison to vowels but more specifically of continuants, compared to plosives. However, the author did not distinguish voiced from voiceless plosives in his analysis, while it has been shown that voicing impacts the associations (Nobile, 2015). The study of Fort, Martin and Peperkamp (2015) captured a subtler picture, rather than either a main vocalic or consonantal effect: the consonantal composition had a significant impact on participants' choices, and the effect of vowels differed in interaction with the consonantal context. The greater effect of consonants could not be explained by their occurrence as initials in pseudo-words, since the authors presented CVCV and VCV structures and the category of the first segment did not impact the effect. The study further revealed a continuum of manners of articulation: plosives are associated to spiky and sonorants to round, while fricatives are in-between, which is consistent with the continuum postulated by Styles and Gawne (2017). Using a judgement task, Knoeferle, Li, Maggioni and Spence (2017) identified a more precise gradient. From spiky to round, the associated consonants were as follows: voiceless plosives > voiced fricatives > voiced plosives > nasals > voiceless fricatives > liquids > glides[19].

All in all, most studies used the segments occurring most frequently in the languages of the world, and these segments seem to appear at the extremities of the continuum between spikiness and roundness. Both their high frequency and their position at such extremities may be due to their high discriminability (Styles & Gawne, 2017). As for the middle segments (in the continuum), their associations differ depending on the contrasts in which they are presented (Fort et al., 2015; Styles & Gawne, 2017). For example, a voiced plosive can be associated to either a spiky or a round shape depending on the contrasted consonant, a sonorant or a voiceless plosive, respectively. Also, one feature is not necessarily sufficient to predict associations. For

[18] Very few segments were tested: [w, d, k] as onsets and always [s] as coda. The tested vowels were [a, u, i].
[19] The authors also found a gradient for size contrasts, from large to small: sonorants > voiced fricatives and voiced plosives > voiceless fricatives and voiceless plosives.

example, the impact of voicing may differ according to manner: voiced plosives are more associated to round than voiced fricatives (Knoeferle et al., 2017), hence voicing may not be sufficient to predict the associations. For this reason, it seems more reasonable to consider segments rather than features. However, this can also be explained by another factor – the shapes of letters – as proposed by Cuskley et al. (2015). Indeed, they proposed as an explanation that the associations are induced by the shape of letters. For example, 'g' (a voiced plosive) is a letter that is rounder than 'z' (a voiced fricative) in the Latin alphabet. 'g' would be, as a result, more associated to a round shape.

In addition to these two explanations involving acoustic and graphemic influences, other explanations for the origin of *motivated* associations are proposed in the literature. The following section exposes them, beginning with graphemic biases.

2. Possible explanations for the existence of *motivated* signs

Observations about *motivation* are not explainable by any fortuitous phenomenon because of the consistency found across languages and speakers, no matter the linguistic structure (i.e. lexicons, phonemic frequencies, pseudo-words) or the conceptual object that is studied (e.g., size, emotions, shapes). This section focuses on the different explanations found in the literature, ranging from the shapes of letters to phylogenetic considerations.

2.1. *The potential graphemic origin of* motivated *associations*

Some authors have wondered if the bouba-kiki effect could be explained by graphemic biases. Cuskley et al. (2015) conducted a bouba-kiki judgment task using pseudo-words composed of either angular letters ('k, t, z, v') or curvy letters ('g, d, s, f'). They found that the first set of pseudo-words better fitted with the angular shape, while the second set better fitted with the round shape, based on judgments. In a second experiment, pseudo-words were presented orally, and the authors found similar results, but also an additional effect of voicing: voiced pseudo-words better fitted with the round shape, and the voiceless ones with the spiky shape. The authors argued in favor of a graphemic bias which would have mediated *motivated* associations in Westerners (or at least people knowing the Latin alphabet). This hypothesis would imply that: 1) *motivated* associations require written language acquisition, 2) these associations are specific to Westerners who use the Latin alphabet and 3) oral pseudo-words could evoke possible written forms (which is possible, on the basis of Chéreau, Gaskell and Dumay's experiment of 2007).

Different studies contradict this hypothesis and its implications. While some found the same effects no matter the modality of presentation (e.g., Davis, 1961; Nielsen & Rendall, 2011) – which could be questionable due to the knowledge of the Latin alphabet – Bremner et al.'s study with Himbas brought to light *motivated* associations in people who have neither a written system nor knowledge of the Latin alphabet. Moreover, some studies with children below the age of language acquisition suggest the existence of some associations before the learning of written forms (see section 2.5). Another strong evidence comes from Bottini et al. (2019) who compared sighted to early blind[20] Italian speakers in a bouba-kiki task. Spiky and round objects were presented by pairs and participants had to choose the object that matched best a given pseudo-word (the experimenter asked for the object that better matched either 'maluma' or 'takete'). They found consistent and expected matching in 83% of the sighted participants and 73% of the blind participants, without statistical difference between the two groups. Hence, vision – and more precisely graphemic shapes – does not seem to explain only by itself the correspondences between these pseudo-words and shapes. A second experiment aimed at better assessing the potential role of the shapes of letters, which could explain the small difference observed between blind and sighted people. CVCVCV pseudo-words were presented orally (using a large variety of segments) and participants – blind or sighted – had to determine if they were rather spiky or round. The authors found main effects of consonant manner (plosives with spiky shapes, sonorants with round shapes, and fricatives in-between), voicing (voiceless with spiky shapes and voiced with round shapes) and vowel backness (back vowels with round shapes and front vowels with spiky shapes). More interestingly, they also found an interaction between graphemic spikiness and group (blind *vs*. sighted). To further assess this interaction, they ran two different models – one per group – and found a simple effect of the shapes of letters in the sighted group only. This means that sighted people rely on – among other stronger factors – the shapes of letters, and this may explain the weaker bouba-kiki effects found in blind people.

Rather than demonstrating the graphemic origin of bouba-kiki associations, Cuskley et al. (2015)'s study highlighted the importance of the modality in which pseudo-words are presented to participants, and pointed to potential cumulative effects (since effects were stronger with the graphemic presentation). Moreover, Bottini et al. (2019)'s study pointed to the fact that effects may be expected to be stronger in literate subjects. All in all, the shapes of letters can influence associations but it cannot explain their existence, otherwise *motivated*

[20] They completely lost their sight at birth or before the age of four, and thus do not know the shapes of letters.

associations would be recent from a historical and evolutionary point of view. Rather, evidence coming from Turoman and Styles (2017)'s study highlights the opposite influence of *motivation* on the shapes of letters. The authors showed pairs of glyphs from several written traditions (e.g., Tamil, Mongolian) to participants using an online survey platform. One glyph contained the sound [i] and the other the sound [u]. Subjects had to guess which one contained the [i]-sound, or the [u]-sound, depending on the condition. Better-than-chance performances were obtained. This result leads to the assumption that the shapes of letters may be *motivated* by the sounds they represent. Hence, *motivated* associations may have constrained or at least influenced the shape of letters, as they have partially done so with the phonetic composition of words across the languages of the world.

In conclusion, the shape of letters cannot explain the origin of bouba-kiki correspondences. The following subsections focus on other hypotheses about the possible origins of diverse *motivated* associations, like size-sound correspondences.

2.2. *Size coding hypothesis*

In studies on language origin and evolution, the descent of the larynx has been of particular interest because it seemed to increase the likelihood of choking hazards. For a long time, following Lieberman (1984), researchers admitted that this problem was counterbalanced by the enabling of language, the main argument being that only a low larynx could enable the current vocalic space. However, more recent studies have come out with a more adequate scenario. Indeed, because a longer vocal tract lowers the pitch and a lower pitch is associated with larger individuals, the descent of the larynx might have happened as a way for smaller individuals to sound larger and scare away potential predators, or as a way for males to attract females. This would have represented a definitive advantage with a benefit exceeding the cost of this newly dangerous lowered position (Ohala, 1984; Fitch, 2010).

The *size coding hypothesis* proposed by Ohala (1984) refers to the advantage of the modulation of pitch in interactions. It was built from Morton's study (1977, in Ohala, 1984) which examined the vocalizations of different species in agonistic contexts. To summarize Morton's findings, low-pitched vocalizations convey aggressive behavior while high-pitched vocalizations convey submissive behavior through relative impressions of size. Usually, the larger individual (who is usually the older and more mature one) has the advantage, and, in general, pitch is inversely proportional to size: the larger the individual, the lower the vibration

rate of the vocal folds, resulting in a lower pitch. Visual (e.g. bristling) and acoustic (e.g. yells) manifestations can lead to fight avoidance – one side renouncing to fight in the face of an opponent who seems larger. Ohala extrapolated this inverse relation between size and pitch to human language with more subtle behaviors, extending it to smile (the lip spreading shortens the vocal tract), 'o-face' (the lip protrusion lengthens the vocal tract) and the distinction between question and statement (asking a question necessitates cooperation while making a declaration needs to convey a confident impression). Furthermore, segments that can be found in words relative to size, like [i] in 'little' and [a] in 'large', do not give an impression of the size of the speaker, but rather communicate about the size of an object. Hence, sound patterns about size generally speaking, as described in section 1.2.1.1, may derive from this phenomenon of communication about one's size that is found in humans and in other species. This theory links *motivated* associations at a linguistic level to behavioral and evolutionary phenomena through phylogenetic dynamics. It is, however, limited to impressions of size. The following subsection deals with other explanations related to articulatory and acoustic features.

2.3. *Possible articulatory and acoustic origins of associations*

Ramachandran and Hubbard (2001b) proposed as an explanation to some *motivated* associations – like those related to size – the articulatory imitation of physical gestures. More precisely, the vocalization of the vowel [i] would mimic a small pincer gesture, in contrast to the vocalization of [a].

This is in line with Sakamoto and Watanabe (2018)'s results about tactile sensations and the parallels they proposed with places of articulation. These authors studied *motivated* associations between mimetics and tactile sensations. Participants, who were Japanese, had to describe tactile sensations and were free to use pre-existing mimetics, to create new ones, or to use adjectives. They mostly used mimetics instead of adjectives and 80% of them were preexisting ones. After naming tactile sensations, they had to evaluate them according to eight pairs of adjectives (e.g. comfortable – uncomfortable) on scales from 1 to 7. The authors analyzed the relations between the latter adjectives and the mimetics participants produced in terms of syllables, clusters of segments and features – while restricting their analyzes to the first syllables. For the sake of simplicity, only a summary of the analyzes about features are compiled in Table 5. This shows that segmental features can bear sensitive and qualitative meanings.

Table 5. Summarized results of Sakamoto and Watanabe (2018). Actual results are more complex and those about consonants present some relativities depending on vowels that are not reported here.

	Segmental features	Associated concepts
Consonants	Voiceless	Comfort; Flat; Smooth; Slippery
	Voiced	Discomfort; Bumpy; Rough; Sticky
	Front + Nasal	Soft; Elastic; Sticky; Wet
	Back + Affricates + Fricatives	Hard; Inelastic; Slippery; Dry
Vowels	Back	Comfort
	Front	Discomfort

The most important distinction for tactile categorization seems to be comfort. These results confirm Hinton, Nichols and Ohala (1994) and Ohala (1983)'s assumption (in Sakamoto & Watanabe, 2018) according to which voiced consonants and anterior vowels are associated to negative emotions, because they require more pressure in their articulations. The authors also explained some of these results with perceptuomotor analogies: 1) bilabial and alveolar nasal consonants involve soft tissues which would be the reason why they are associated to soft, elastic, sticky and wet sensations; 2) alveolar affricate and fricative and velar plosive consonants are articulated with harder parts of the vocal tract, hence the reverse pattern of associations. It is also possible to make a parallel with Blasi et al. (2016)'s findings: concepts for 'tongue' tend to contain [l] – involving a tongue movement – while those for 'nose' tend to contain [n] – implying a nasal airflow.

In addition to articulatory explanations, acoustic ones can be proposed, as did some authors like Knoeferle et al. (2017) and Nielsen and Rendall (2011). For example, the burst by which plosives begin may be crossmodally similar to visual spikes. However, these explanations mostly rise from interpretations and, at this time, no experiment has assessed their validity.

So far, potential explanations lean on communication about one's size (through pitch), imitation (through vowel articulation) and perceptuomotor analogies (between meanings and either articulatory or linguistic features). The following subsection focuses on emotional explanations with two studies, one about emotional (and size) content expressed by acoustic

and articulatory features, and another about the potential involvement of emotions in speech emergence.

2.4. *Emotions*

Chuenwattanapranithi, Xu, Thipakorn and Maneewongvatana (2008) studied what they called the *size code hypothesis of emotional speech*, which refers to the coding of emotions (from anger to happiness) on the shared basis of body-size projection. They synthesized speech sounds of Thai vowels varying on three parameters: larynx height, lip protrusion and pitch (F_0). Each sound was presented either statically or dynamically regarding larynx height and pitch (i.e. with an initial acceleration and a final deceleration). The task consisted in choosing which vowel among two was spoken by a large or angry person, depending on the condition. All three parameters significantly influenced the responses. For size judgments, lower larynx height and lower pitch were associated to a larger person, and conversely, higher larynx height and higher pitch were judged as produced by a small person. There was also an interaction between larynx height and sound dynamism: larynx height provided better cues about size when the sound was static rather than dynamic. For judgments about emotions, lower larynx height and lower pitch were more associated to angry persons, and conversely, higher larynx height and higher pitch were more judged as produced by happy persons. Two interactions also appeared: 1) between laryngeal length and pitch: a lower larynx height accompanied by a lower pitch sounded the angriest; 2) between dynamics and laryngeal length: dynamic sounds produced by a lower larynx sounded the angriest. This study highlights the influence of acoustic modulations on judgements about size and emotions. Moreover, it confirms experimentally the size code hypothesis proposed by Ohala (1984) and goes beyond, proposing a potential ground for the perception of emotions (at least for the distinction of two emotions: happiness and anger). The following study clarifies the benefit of communicating about emotions in an evolutionary perspective.

Adelman et al. (2018)'s study reported in section 1.2.1.2 not only analyzed phonemic composition of words of different languages (English, Spanish, Dutch, German and Polish) expressing emotions, but also distinctly tested the first and final segments of words of different languages. The authors found that the first segments (R^2 = 1.16-3.86%) better predicted valency than the last ones (R^2 = 0.48-1.75%)[21] and explained it by a faster transmission of information, especially in case of danger, which outlines an adaptive advantage. To go further, they analyzed

[21] Polish is not included in this interval because last segments do not predict valency in this language.

pronunciation latency, i.e. the time between the stimulus onset – the beginning of the display of the segment – and the voice onset – the beginning of the answer, in a word pronunciation task. They investigated English and German, the only two languages for which they had data. Results showed that segments that are faster to pronounce tend to be at the beginning of negative words and conversely, those that are slower to pronounce tend to be at the beginning of positive words. Analyzes were significant for English and German and their respective effect sizes (Pearson's r) were 0.55 and 0.42. On this basis, and contrary to the position that *motivation* may ensue from generalization or analogical mechanisms (i.e. '*spandrel account*'[22]) which would have made *motivated* signals easier to learn, the authors support the assumption that it is actually an adaptive phenomenon (i.e. '*adaptation account*'). A selection pressure could have favorized individuals with greater communication abilities, more precisely who were able to better produce and perceive segments because they conveyed information about emotions, hence about potential dangers. This efficiency may depend on speed: the faster the negative emotional signal is received, the faster the proper response can be executed. This last point would in turn explain the benefit of beginning negative words with segments that are faster to pronounce. In other words, rather than deriving from language emergence, these authors support the idea that *motivation* might have underlain language emergence, at least in part, through natural selection.

As it is not possible to turn the clock backward in order to uncover the origins of language, paying attention to studies with children may provide information about the potential innateness of *motivated* associations, which is the subject of the following subsection.

2.5. *Ontogeny (and phylogeny)*

Based on their assumption that *motivation* may have facilitated language emergence, Imai et al. (2015, p. 2) further suggested that *motivation* '*may still facilitate synchronic language learning in infants and children*'. Hence, ontogenetic development may 'reflect' the phylogenetic evolution of language.

Several studies bear an interest in *motivated* associations within children aged from a few months to a few years. These studies contribute to settling the question whether correspondences are innate, or whether they are shaped from language exposure. This

[22] A spandrel is a character that appeared outside of an adaptation and that is considered as a byproduct of the evolution. For example, in deer, the overdevelopment of vertebra to support the antler has become a secondary sexual characteristic (Gould, 1997).

subsection exposes some studies within this framework, from natural crossmodal correspondences to bouba-kiki shapes.

2.5.1. Natural correspondences

Smith and Sera (1992) studied cross-modal correspondences, more precisely by testing size-darkness and size-loudness mappings in children of different ages (2, 3, 4 and 5-year-old) and in adults. Participants had to choose among two stimuli the one that best matched an object of a given size. They found that size-loudness mappings started at the age of 3 and that the mapping strength increased with age (large was associated to loud and small to quiet). The mappings between darkness and size appeared in 2-year-old children: they associated large to dark and small to light. This effect disappeared by the age of 3. However, in adults, three different patterns of mappings showed up: 1) large-dark and small-light; 2) large-light and small-dark; 3) no cross-modal correspondence. The authors explained the phenomenon about size-loudness mappings as a *'unified organization of cross-dimension similarities'* facilitated by language (p. 117). For the size-darkness mappings, they described an early perceptual organization (at the age of 2) that may be destabilized by language (by the age of 3), followed by an idiosyncratic organization in adults. Hence, mutual influences – either reinforcing or contradicting – may exist between natural perceptual biases and language development. Specifically, natural correspondences may influence language development and vocabulary growth may in turn modify perceptual correspondences.

Peña, Mehler and Nespor (2011) studied sound-size mappings in children who were 4-month-old and whose parents were Spanish speakers. They presented CV syllables composed of consonants [l], [f] or [d] and of vowels [i] and [o] in the first condition, and [e] and [a] in the second condition. At each trial, one syllable was exposed, accompanied by two objects, one small and one large. The authors analyzed the direction of the first gaze and total looking time, conveying preference for one of the two objects. They found significant differences according to the vowels[23]: children looked preferentially to small objects when accompanied by [i] or [e], and to large objects when accompanied by [o] or [a], as shown by both first-gaze direction and total looking times.

These experiments are of interest for two reasons: 1) they provide information about correspondences that exist prior to language acquisition or that arise from it, including one

[23] The authors did not analyze the influence of consonants.

involving pure auditive perception (i.e. loudness) (Smith & Sera, 1992); 2) they document the existence of early phonemic distinctions in interaction with other modalities (Peña et al., 2011).

2.5.2. Bouba-kiki shapes

Maurer et al. (2006) tested a group of children (average age: 2.8 years) with different round and spiky shapes displayed by pair. Each pair of shapes was accompanied by one pseudo-word among two, e.g. one pair of shapes with 'kaykee' for half of the participants and with 'bouba' for the other half. Children had to decide which form corresponded to the presented pseudo-word. Authors analyzed responses according to the vocalic roundness within the word (although consonants also differed within a pair). Overall, children made the expected associations (e.g. 'bouba' with the round shape or 'kaykee' with the spiky shape), except for one pair of shapes. However, one cannot claim that vowel roundness and the sight of the mouth shape that accompanied the pronunciation of vowels – which the authors tried to emphasize in their protocol – are the explaining factors in this study, since pseudo-words also differed on consonants. Furthermore, as the authors reported, they could not *'disentangle whether the child matched the sound to a shape based on its sound, the shape of the experimenter's lips as she spoke the word, and/or the feeling in the child's mouth of mimicking the sound'* (p. 321).

Spector and Maurer (2013) also tested vocalic influences but used pseudo-words that only differed on vowels (as 'kiki' and 'koko') with children aged from 2.5 to 3-year-old. The children were exposed to pairs of shapes accompanied by one pseudo-word selected among two (containing either [i] or [o]). Children answered to as much [i] as [o], these vowels being combined with four different consonants ([g], [b], [k] and [d]). The authors found consistent results: children associated pseudo-words containing [o] with round shapes and those containing [i] with spiky shapes. Although there were no statistical analyzes contrasting consonants, the pair [kiki]-[koko] departed from others (i.e. [g], [b] and [d]) by eliciting responses close to chance level. This may potentially point to consonantal influences in the 'bouba-kiki' effect.

Imai et al. (2015) used a preferential looking procedure with 14-month-old Japanese children with bouba-kiki shapes and the pseudo-words 'kipi' and 'moma'[24]. Children were assigned to a congruent (e.g. 'kipi' with spiky) or incongruent (e.g. 'kipi' with round) condition. A first phase consisted in presenting pairs, constituted of one shape and one pseudo-word, in

[24] They constructed different combinations of pseudo-words using different consonants (m, l, n, p, k), and vowels (a, o, i) and first tested them with adults speakers of Arabic, Japanese and English in order to select the two pseudo-words which presented the highest consistency for naming round and spiky shapes across these different speakers.

order to learn these associations (congruent or incongruent, depending on the allocated condition). After this first phase, the experimenters asked children in the test phase what object, among two, was the 'kipi' or the 'moma'. Overall, children in the congruent condition looked longer to the correct object compared to those in the incongruent condition. To assess the respective influences of *motivated* associations, learning, and their interaction, the authors compared different models, containing or not these different factors. They found that the most explicative model was the one that contained the three of them. They also noticed an influence of temporality in accordance with the congruent condition. They concluded about their results that *'sound symbolism provides additional boost to the training especially in the first 800ms such that the training effect was stronger for the infants in the match condition than those in the mismatch condition'* (p. 12).

Ozturk et al. (2013) also tested the bouba-kiki effect but with even younger children (the mean age was 4 months), using 'bubu' and 'kiki' as recorded pseudo-words. One pair, made of one shape and one pseudo-word, was shown at a time. Children looked longer at incongruent pairs than at congruent ones. To assess more specifically the role of vowels, the authors replicated the same task with pseudo-words differing only on vowels 'kiki' and 'kuku' and found no effect. Similarly, to assess the role of consonants, the authors compared pseudo-words that only differed on consonants: 'bubu' and 'kuku' but also found no effect. However, in the first case, vowels were presented in a consonantal context (i.e. [k]), and in the second case, the consonants were presented in a vocalic context (i.e. [u]). The segmental context may influence responses, something which was not taken into account (by also presenting 'bibi' and 'bubu' in the first case and 'bibi' and 'kiki' in the second). All in all, infants differentiated congruent from incongruent pairs only when both vowels and consonants differed.

In comparison to other studies, Ozturk et al. (2013) found longer response times for incongruent pairs, while the reverse was reported in other studies. These authors explained this by a difference in paradigm. Others studies (Imai et al., 2015; Maurer et al., 2006; Peña et al., 2011; Spector & Maurer, 2013) used protocols that consisted in presenting two shapes and one pseudo-word (this type of paradigm will be later referred as 2x1), while this study used a paradigm that consisted in presenting one shape and one pseudo-word (similarly, 1x1). In 2x1, the authors interpreted looking time as a marker of preference or choice. However, in 1x1 – in which one matching is presented *at a time* (made of one shape and one pseudo-word) – the authors interpreted the looking time as the expression of the detection of incongruity. They also added that *'infants presented with more variables and more complex stimuli tend to look longer*

at relatively familiar or congruent pairings, whereas infants presented with simpler stimuli tend to look longer at relatively novel or incongruent pairings' (Ozturk et al., 2013, p. 178).

Either way, these studies demonstrated that children by the age of four months are efficient at distinguishing congruent and incongruent pairs based on sound-shape and sound-size mappings.

Based on these results, we cannot firmly conclude to the existence of an innate inclination for *motivated* associations in children for the main reason that newborns already have a linguistic experience. They are indeed able to distinguish their mother tongue from other languages since birth (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993). In four months, newborns may have had sufficient linguistic exposure to demonstrate language-based cross-modal correspondences. What is more, the linguistic correspondences (e.g. between the vowel [i] and the concept 'small') may be more influential than other environmental correspondences (e.g. an object of a small size producing a high-pitched sound). This may explain why children are sensitive to linguistic correspondences between segments and size (Peña et al., 2011), or between segments and shape (Ozturk et al., 2013), as early as the age of four months, while environmental correspondences like between loudness and size do not appear before the age of three years (Smith & Sera, 1992). Altogether, these results are in favor of a *statistical* learning of correspondences, that may originate from linguistic exposure, in comparison to *semantic* learning (because it is unlikely that four-month-old children are able of having semantic representations). However, it does not rule out their possible *structural* (to a certain extent innate) origin, neither the coexistence of these three types of origin (which will be further reported in section 3.2). Moreover, even if language exposure underlies these correspondences, preceding cognitive biases stemming from phylogeny may play a part in their emergence (Imai et al., 2008). More precisely it may facilitate their learning, which is the topic of the following section.

2.6. *Learning facilitation*

Starting from aforementioned Kunihira's experiment (1971) in section 1.2.2 – about Japanese antonym pairs presented to American students (who guessed their meanings above chance level) – Nygaard, Cook and Namy (2009) conducted a study using Japanese antonyms in a learning paradigm. Participants, who were native speakers of American English, learned the meanings of these Japanese antonyms according to three conditions: match, mismatch and

random. In the match condition, participants learned the actual meaning of only one word within a pair (e.g. 'akarui' that means 'bright' *or* 'kurai' that means 'dark'). In the mismatch condition, they learned the meaning of the word's antonym (e.g. 'akarui' associated to 'dark' *or* 'kurai' to 'bright'). In the random condition, words were randomly associated to different meanings (e.g. 'akarui' associated to 'catch' *or* 'kurai' to 'wet'). Hence, participants learned one antonym of a pair (i.e. either 'akarui' or 'kurai') with one English target (its true meaning, the opposite meaning or an unrelated meaning). During the learning phase, the Japanese word was orally presented while the English target was simultaneously displayed in written form. The test phase consisted in presenting the Japanese word orally with two written possible English meanings, the target and a distractor. The learning and test phases were repeated three times. The authors analyzed response time and accuracy within participants who exhibited a minimum performance of 80% correct answers across the entire experiment (90 out of 104 participants). The accuracy analysis showed that the performance was significantly better in the match condition (94.3%), compared to the random one (91.6%). However, there was no significant difference either between match and mismatch (93%), or between mismatch and random. The analysis of response times also revealed a benefit both for the match and the mismatch conditions in comparison to the random one. However, there was no significant difference between match and mismatch conditions. These findings suggest that *motivated* correspondences can facilitate the access to specific semantic fields (e.g. brightness) whatever the polarity of words (i.e. both words 'akarui' and 'kurai' facilitate the linking with the concept of brightness). Hence, there is evidence that a word is easier to link to its semantic field rather than to a random one (Nygaard et al., 2009). Parallelly, other studies that used foreign words reported guesses higher than chance, which reveals that it is possible to guess the meaning of a foreign word *when presented in a pair* – with its opposite (i.e. antonym; Brown et al., 1955; Kunihira, 1971) or another related meaning (bird and fish; Berlin, 1994).

Nygaard et al. proposed as an explanation that '*such cross-modal correspondences may be achieved via some literal or figurative resemblance between the sound and meaning (e.g., vowel height may correlate with relative size), or may reflect an embodied representation involving simulation of the actual meaning*' (p. 185). The literal or figurative resemblance is in line with theories proposed by Ohala (1984), Ramachandran and Hubbard (2001b) and Sakamoto and Watanabe (2018) previously reported in section 2.3. Nygaard et al. (2009) also added that this phenomenon may have functional consequences such as language acquisition in children, which was investigated in Imai et al. (2008)'s study, already reported in section 1.2.2.

In addition to the comparison between English and Japanese adult speakers, Imai et al. (2008) also ran their clip task with Japanese children. They found similar results in two-year-old Japanese children and in English adults (65,7% vs. 64%). However, at three years of age, an augmentation of the effect (75%) expresses Japanese children's language exposure.

The authors also conducted another study in order to assess whether *motivated* associations help 3-year-old Japanese children in the generalization of action. Indeed, before the age of five, Japanese children present difficulties in generalizing an action verb learned with one object and one actor to the same action but with other objects or other actors. The authors showed the children video clips with oral description: either invented (*motivated*) mimetic verbs (e.g., tokutoku, batobato) or invented (non-*motivated*) verbs (e.g., chimoru, nuheru). Then, experimenters showed two clips, one that displayed the same action as in the first clip – but with a different character – and another that displayed another action – with the same character as in the first clip. Children had to determine which clip corresponded to the word they learned. The performance with mimetic verbs was better than the one with non-*motivated* verbs: with mimetics, children selected more the same action than the same character, which means that they generalized more the meaning of the action. To ascertain that this effect was not imputable to an online matching (i.e. a choice based on *motivated* associations during the matching task unrelated to the previous learning), the authors conducted another experiment in which children learned to associate a matching between mimetic with an action and a given mimetic while this mimetic congruently corresponded to the wrong answer of the test phase. In this condition, children did not choose preferentially the wrong – *motivated* – matching. Hence, *motivated* matching mimetics help Japanese children to extract action meaning – and to dissociate it from the actor. It thus facilitates word learning. However, the authors did not test non-matching mimetics, which could have informed us about the impact of word form – more precisely reduplication – and its potentially confounding effect on word learning.

In 2011, Kantartzis, Imai and Kita replicated the same procedure with 3-year-old English-speaking children, but they added the non-matching mimetic condition. They obtained the same results as in the former study: English children better learned and generalized the matching mimetic – better than chance – while they did not for verbs and non-matching mimetics. Hence, reduplication does not explain the enhancement of learning, but *motivated* associations do. According to the authors, these results support a universal influence from *motivated* associations instead of an effect explained by linguistic exposure (because Japanese children are exposed to a large number of mimetics, and could thus have learned regularities).

Nielsen and Rendall (2012) conducted a learning paradigm in which they presented one shape (spiky or round) and one pseudo-word composed of specific consonants – either [p, t, k] or [l, m, n]. Participants were either in the congruent condition or in the incongruent condition, the congruency being based on previous findings. The authors found a learning performance significantly higher than chance in the case of the congruent condition (53.3%) while the incongruent condition did not differ from chance (50.4%). Nielsen and Rendall concluded that the bouba-kiki effect has been overestimated due to explicit (associative) tasks, and that it also exists, more subtly, at an implicit level.

This study brought to light the advantage of *motivated* associations in the learning of categories (round and spiky). But another type of learning was not assessed in this study, namely the learning of individual stimuli. The following study, conducted by Monaghan, Mattock and Walker (2012), aimed at evaluating these two types of learning.

Monaghan et al. (2012) employed an implicit learning paradigm (without feedback): one pseudo-word (e.g. composed of plosives) was always presented with its target shape (e.g. a spiky shape), and either one exemplar of the other type of shapes (i.e. a round shape) or another exemplar of the same type of its target shape (i.e. another angular shape). The pairings between pseudo-words and target shapes were either congruent or incongruent (based on previous findings). The learning hence occurred over the four blocks of 64 trials. The authors found better performances (i.e. participants selected more the proper type shape) for congruent rather than incongruent pairings. Hence, congruent pairings enhanced learning. More interestingly, the congruence effect only appeared when the target shape was presented with a target of the other type rather than with another exemplar of its own type. This means that *motivated* pairings facilitate the learning of categories (categorization) rather than specific exemplars (individuation). Individuation can also be about identifying particular words, without confusion. For example, it is easier to distinct two animals called ‘cow’ and ‘sheep’ instead of using closed names as ‘feb’ and ‘peb’. According to Corballis (2002, in Monaghan, Christiansen, & Fitneva, 2011), the individuation of referents can sometimes be a matter of life and death. Indeed, it may prevent confusion between edible and poisonous plants, for example.

On this basis, Monaghan et al. (2011) conducted the first empirical study about the advantage of arbitrariness over systematic mappings between meanings and segments (which may include *motivated* ones, but also grammatical markers). Using computational and experimental methods, they found two major outcomes: 1) systematic mappings facilitate categorization; 2) arbitrariness facilitates the individuation of referents. More precisely, in the

experimentations, participants were exposed to a set of pseudo-words which referred to specific pictures of two possible categories (action and object). The link between pseudo-words and categories was either systematic (e.g. fricatives associated to action and plosives to objects) or arbitrary (i.e. fricatives with as much objects as actions). The authors analyzed the accuracy of learning per category (categorization) and per referent (individuation). There was an advantage of systematicity over arbitrariness for categorization. As for individuation, systematicity presented an initial advantage that was caught up by arbitrariness over blocks. The authors assessed, in a second experiment, the role of contextual information. They compared the same conditions while adding a category-marker (e.g. [wɛ] systematically preceded a word denoting an action and [mə] systematically a word denoting an object). In this case, they obtained an initial advantage for the systematic relation that was caught up by the arbitrary relation over blocks for categorization and that was even surpassed for individuation. Arbitrariness is thus advantageous since it provides complementary information that distinguishes referents while phonological systematicity provides redundant information with context. Because languages do not always provide contextual information denoting categories, the authors constructed a third experiment in which a pseudo-word contained both systematic and arbitrary information (e.g. one category had as codas [ʒ] and [f] and the other the codas [k] and [g], while the onsets [ʒ], [f], [k] and [g] occurred in both categories). This model resulted in the highest accuracy for individuation. All in all, systematicity facilitates the learning of categories and this consequently maximizes the information for individuation through arbitrariness. Hence, the combination of systematicity and arbitrariness enhances the learning of both categories and specific word meanings[25]. Either way, it is interesting to note that the advantage deriving from arbitrariness requires systematicity. As claimed by Dingemanse et al. (2015), respective or overlapping advantages of systematicity and *motivated* associations would need to be deepened. Lockwood and Dingemanse (2015)'s review of literature also points to the complementarity of arbitrariness and *motivation*. They outlined that, '*by supplying perceptual analogies for vivid communication, sound-symbolism allows for communication to be effective; by providing the lexicon with greater depth and distinction, arbitrariness allows for the efficient communication of concepts*' (p. 2).

[25] Additional analyzes of natural languages (English and French) also corroborate these findings: beginnings of words provide more information on their identity, which may speed their identification (which is consistent with Adelman et al. (2018) 's study about emotions), while endings predict their grammatical categories.

2.7. *Summary of the possible origins and implications of the associations*

To summarize, different potential underlying mechanisms for *motivated* associations have been proposed in the literature. It is clear that the shape of letters can influence *motivated* associations, but it cannot explain, by itself, their existence. Moreover, different acoustic, articulatory and emotional dimensions seem at play in different associations (e.g. the small aperture of [i]). It has been suggested that they could have played a role in language emergence (size code hypothesis, emotions), and there is evidence in favor of their implication in facilitating learning processes in children. Particularly, the early ability of children to distinguish between congruent and incongruent associations, as well as the learning advantage for congruent or systematic associations in adults, lead to inquiries about potential innate biases, and more generally the cognitive mechanisms at play. As it has been said earlier, the early advantage for congruent pairings that exists in children is not necessarily a proof of innate associations (through brain organization and/or specific cognitive mechanisms), since few-month-old children have already received language exposure. It may originate from environmental exposure – i.e. statistical environmental learning – by which people can learn, for example, that a spiky object may produce more high-pitched sounds. It may, however, also arise from linguistic exposure, through statistical or semantic learning. For example, Monaghan et al. (2012) conducted analyzes on the English lexicon of words related to spikiness and roundness and found two phonetic features related to these concepts: there are more velar consonants in words denoting spikiness, and more voiced consonants in words denoting roundness[26]. As a result, English speakers may rely on this statistical co-occurrence to make associations, which is in line with the advantage of systematicity in the learning of categories (which may also be the case with phonesthemes). However, this cannot entirely account for bouba-kiki associations, since they are found in other speakers of different languages. Rather, *motivated* associations may have had influenced the vocabulary about roundness and spikiness (just as it could have influenced the shapes of letters, as mentioned earlier), and the knowledge of these words denoting roundness and spikiness then, in turn, can reinforce the associations.

All in all, evidence of *motivation* tends towards a more general cognitive mechanism that consists in linking different modalities together. The following section addresses this issue.

[26] These effects were weak and disappeared after correction for multiple comparisons (for the reason that they were tested among 18 features). If the authors had restricted their analyses to the features known to be associated with these shapes, they would likely have been significant.

3. Potential cognitive mechanism(s) involved in *motivated* associations

3.1. *Two major candidates: crossmodal correspondences and synesthesia*

Motivated signs can easily be considered as examples of crossmodal correspondences (Spence, 2011), as they link sound properties – or segments – to meaning or, more generally, features of other modalities. They can appear in different types of contrasts: contrasts between auditory properties (e.g. sound pitch and size, as in Gallace and Spence's study in 2006); between segments (e.g. 'mil' and 'mal' for size contrasts, as used by Sapir in 1929); between pseudo-words, varying on several segments ('maluma' and 'takete', and then 'bouba' and 'kiki', with spiky and round shapes, in Köhler's original study in 1947). One particular type of crossmodal correspondences is *synesthesia* and Ramachandran and Hubbard (2001b) argued in favor of a synesthetic explanation for the bouba-kiki effect. Synesthesia is the phenomenon that consists in experiencing a secondary sensation (*concurrent*) when a first one (*inducer*) in the same or another modality is stimulated, for example letters inherently colored, tastes induced by spoken words, etc.

The following subsection exposes in greater details what crossmodal correspondences and synesthesia are, as well as their relation and the terminological issue that ensues.

3.2. *Crossmodal correspondences*

Crossmodal correspondences consist in relating different properties across different modalities (vision, audition, olfaction etc.), for example, the relation between a ball that is struck and the sound that comes with it. In his review, Spence (2011) starts with a distinction between two terms used in literature: *synesthetic correspondences* and *crossmodal correspondences*. The former only refer to sensory features that are not redundantly coded (e.g. sound pitch and visual brightness, which do not necessarily appear simultaneously in the environment), while the latter insure a broader inclusion since it includes both non-redundant and redundant[27] associated features (e.g. the previous example of the ball). The term crossmodal correspondence, being more inclusive, is thus preferred here.

There are three different types of crossmodal correspondences (Spence, 2011): structural, statistical and semantic correspondences.

[27] Redundant does not mean here systematic.

3.2.1. Structural correspondences

These correspondences stem from brain organization and functioning like synaptic connections or common cognitive mechanisms. They are possibly innate (Spence, 2011). According to Marks (2004), they may arise from intrinsic similarities between different modalities at the level of neural coding. For example, intensity would be encoded by the neural firing rate, whatever the modality, which would underlie the relation between loudness and brightness, for example. They also seem to correspond to *low synesthesia*, as defined by Mroczko-Wasowicz and Nikolić (2014), that arises from synaptic connections between perceptive areas.

3.2.2. Statistical correspondences

They arise from environmental exposure and, at a cognitive level, from a multisensory integration of redundant sources that permits a coherent representation of sensory signals (Ernst, 2007). Hence, a signal in one modality can be inferred after a related signal in another modality is perceived (depending on the strength of the coupling between the two sources). For example, touching an object produces a *haptic size estimate* which activates a *visual size estimate* through a *conceptual object size* developed in the brain.

3.2.3. Semantic correspondences

Semantic correspondences result from linguistic exposure and learning, and permit to link different modalities that have linguistic terms in common (e.g. pitch and spatial frequency, which are both described as ‘low’ and ‘high’) (Marks, 2004). However, one may wonder about the accidental or *motivated* origin of these commonalities across different modalities. Indeed, either the same terms may be used for different modalities just by chance (as allowed by arbitrariness), or the similarity across modalities may have *motivated* the use of the same terms in the first place. Either way, the use of common terms for different modalities may reinforce correspondences through semantics.

3.2.4. Individual cases

Some associations are hard to categorize. For example, Gallace and Spence (2006) studied the association between sound pitch and visual size using a categorization task. Participants were presented two consecutive circles and had to decide whether the second circle was ‘larger’ or ‘smaller’ than the first one. The second circle was simultaneously displayed with an auditory stimulus. They obtained the same results whether they used a high or low pitched-sound (experiment 1) or the linguistic – orally-presented – terms ‘high’ and ‘low’

(experiment 3)[28]. The correspondence between size and pitch is thus both statistical and semantic (as the one between pitch and visual elevation). According to Spence (2011), correspondences occur at different levels of cognitive processes: structural and statistical correspondences may take place at an *early* perceptual and a *later* decisional level, whereas semantic ones are primarily decisional. This seems to match the distinction between low- and high-level cognitive mechanisms, as involved for example when processing basic visual features of an object, or accessing its more abstract and conceptual properties, respectively. Section 3.4.5 is dedicated to these perceptual and decisional aspects, in order to deepen the level at which the correspondences occur. Meanwhile, a given association may be semantic in nature while coming from a statistical or structural learning. For example, a letter-color synesthesia, which implies concepts, may derive from a statistical correspondence (e.g. letters displayed in particular colors in an alphabet book for children). Hence, the nature and origin of the correspondences is not always crystal clear.

The following subsection outlines the more specific candidate – synesthesia – and arguments in favor of and in opposition to assigning *motivated* associations to this type of correspondence.

3.3. *Synesthesia and ideasthesia*

According to Ramachandran and Hubbard (2001a), synesthesia exists at a linguistic level as non-arbitrary neuronal connectivity between motor and auditory brain areas and would explain the 'bouba-kiki' phenomenon. However, synesthesia exists only in a small proportion of the population. Different estimations exist depending on the chosen criteria of assessment according to Simner et al. (2006), whose estimation is 4% of the population. There are two major differences between synesthetes and non-synesthetes. First, for the former, the synesthetic congruence is idiosyncratic, which leads to differences across individuals. On the contrary, *motivated* associations are consistent across people, including non-synesthetes (with some variations due to cultural, environmental and linguistic exposures)[29]. Second, synesthetes literally experience an additional sensation.

---

[28] The authors used the term 'synesthetic correspondences' whilst these correspondences are environmentally redundant: the larger an object is, the lower pitch sound it produces – crossmodal correspondences would thus be more appropriate, based on Spence (2011)'s dichotomy.

[29] A study partially belies this point: Moos, Smith, Miller and Simmons (2014) studied synesthetes and non-synesthetes and found some associations (e.g. [a] with red and [i] with yellow and green) for both groups, though these results were trends and did not reach significance. Since associations were stronger and more consistent in synesthetes, the authors concluded that motivation is the same, yet weaker, mechanism as synesthesia. This is further presented in the Discussion section.

Ramachandran and Hubbard studied a few synesthetes (2001b, 2001a) and argued in favor of a pure low-level perceptual origin, based on several reasons. First, it cannot be merely imputable to memory: a synesthete they described (Ramachandran & Hubbard, 2001a) did not 'see' the banana as yellow as the letter F while both were displayed in shades of grey. Since the memory of the color of the banana does not provoke the vision of its color, it is thus unlikely that the perception of the color of the letter is due to the recollection of an associated color which would take over the color in which the letter is displayed. If it was due to memory, it would likely not be restricted to specific objects. Moreover, these people were able to detect a shape formed by letters (among others) better than control participants (Ramachandran & Hubbard, 2001a) (see Figure 1). This tends towards considering a perceptual phenomenon that would probably stems from cross-activations between brain areas (Ramachandran & Hubbard, 2001b). The authors also argued against a potential conceptual explanation, based on the fact that roman numbers do not provoke color perception contrary to Arabic ones (Ramachandran & Hubbard, 2001b). Additionally, the fact that symbols like 'I' and 'V' provoke color perception when presented as letters but not when presented as roman numbers ('IV') would be due to top-down processes (see also Dixon, Smilek, Duffy, Zanna, & Merikle, 2006).

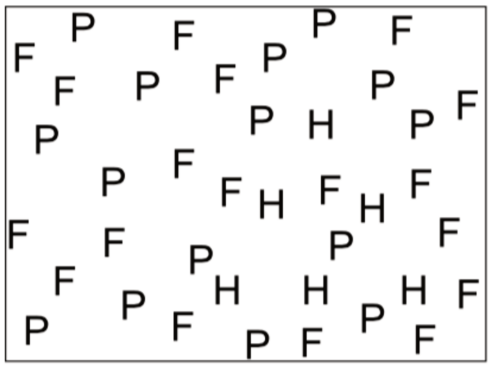


Figure 1. A letter cloud in which the letters 'H' form a triangle extracted from Ramachandran & Hubbard (2001a). Color-letter synesthetes detect the triangle faster in comparison to non-synesthetes.

Nikolić (2009) outlined an alternative to the low-level perceptual hypothesis: a high-level semantic hypothesis (see also Dixon et al., 2006; Martino & Marks, 2001). The inducer would be the *concept* of the letter rather than the basic feature (i.e. the shape of letter). Based on previous studies, Nikolic argued that synesthetes were not faster at detecting letters in comparison with non-synesthetes in a task as the one depicted in Figure 1 (e.g., Edquist, Rich, Brinkman, & Mattingley, 2006). The difference between synesthetes and non-synesthetes reported by Ramachandran and Hubbard (2001a) may be explained by its difficulty (due to the number of stimuli). Moreover, the color evoked by a given grapheme depends on its meaning (e.g. a shape representing either a letter or a digit, such as 'S' or '5'; 'O' or '0') (Dixon et al.,

2006[30]). Hence, Ramachandran and Hubbard (2001b)'s previous argument is questionable: Roman numbers are not as much encountered as Arabic ones, and therefore their significations may not be automatically accessed.

Another evidence comes from Asano and Yokosawa (2011), who investigated synesthesia in Japanese synesthetes with both Hiragana and Katakana syllabic writing systems[31]. They found consistency in color associations across the two writing systems and this is in favor of a phonemic – or conceptual – influence instead of a graphemic one. The authors supposed that synesthesia involves the IFG (inferior frontal gyrus) – which processes phonological information – and not only the VWFA (visual word form area), which processes graphemic shape. Hence, they settled in favor of a higher order processing.

In this regard, Nikolić proposed to take into account the implication of semantics – more precisely the semantic nature of the inducer – by a change of terminology. According to him, the term *ideaesthesia* (*'sensing concepts'* or *'perceiving meanings'*, Nikolić, 2009, p. 3) would be more accurate, since it sheds light on the representational or conceptual side of the associations. As for the pure low-level synesthesia, it would exist only in a context of drug use.

In general, inducers of synesthetic associations are concepts (e.g., letters, months) In another paper, Mroczko-Wasowicz and Nikolić (2014) opposed higher and lower synesthesia. These authors explained lower synesthesia by *'synaptic connections between neurons representing respectively the inducer and the concurrent'* (p. 2). As for higher synesthesia, it would derive from the organization of the brain system, which allows for *'more elaborated, distributed and flexible'* (p. 2) origins of the associations. These associations could be modulated by *'context, attentional mechanisms and interpretation of the stimuli'* (p. 2). It is similar to the distinction made by Dixon et al. (2006) using different names: *projectors* (between basic features) and *associators* (implying concepts).

Besides, according to Martino and Marks (2001), all cross-modal correspondences are similar to 'real' synesthesia and rely on the same neural mechanisms (e.g. *'temporal properties of neural impulses'*, p. 64). More precisely, they opposed 'strong' synesthesia (unusual experiences that exist in few people) and 'weak' synesthesia that corresponds to general crossmodal correspondences. Spence (2011), however, disagreed with the idea that the

[30] This study was conducted with only one participant, but it is a replication of previous studies that led to the same results.
[31] Hiragana is the writing system for Japanese words or particles, while Katakana is the one for borrowings. For example, が is the Hiragana grapheme for /ga/ and カ is its Katakana counterpart.

mechanisms are similar in both cases. He argued that facilitation (coming from congruent associations) and interference (from incongruent associations) effects would then be higher in synesthetes than in non-synesthetes, which would not be the case according to him (he only evoked preliminary testing using a speeded classification task). These facilitation and interference effects will be exposed in the following subsection – among other characteristics.

On the basis of Nikolić's dichotomy about synesthesia and ideasthesia (2009), it is unlikely that a pseudo-word like 'kiki' would automatically and irrepressibly activate – or even depict in mental imagery – the representation of a spiky shape through a synesthetic – perceptual – relation, at least in the larger range of the population. On the basis of this idea, ideasthesia seems more adequate because it goes by conceptual representations which serve as a bridge between a pseudo-word and a shape. However, speaking of crossmodal correspondences is less misleading, because it is more neutral. Concerning *motivated* signs, there is no specific type of correspondence that outperform the others (structural, statistical or semantic). Rather, different explanations relating to these different types can be proposed. For example, the association between loudness and size can be explained by a structural correspondence (through magnitude), the one between pitch and size by a statistical one (through environmental learning i.e. small objects tend to produce higher-pitched sounds, larger ones lower-pitched sounds), and the association between pitch and visual elevation can be explained by a semantic correspondence (the words 'low' and 'high' can be used in both contexts).

Studies about cross-modal correspondences can offer information that could be linked to some specific types of *motivated* associations.

3.4. *Crossmodal correspondences and* motivated *associations properties through diverse paradigms*

This subsection reports studies about cross-modal correspondences, not necessarily involving *motivated* associations. Those that do not directly imply *motivation* may, however, give some insights about it. More specifically, these studies explored the properties of cross-modal correspondence (e.g. the learnability of new correspondences), which may apply to other kind of correspondences, namely *motivated* relations. In any case, most studies involved the auditory modality, which is relevant even though it is not linguistic.

3.4.1. Learning of new correspondences

Correspondences are easily malleable, whether they are statistically or semantically learned. Indeed, Ernst (2007) exposed participants to a new statistical association (that does not

exist in the natural environment) between luminance and stiffness. Within only 1 hour, participants learned the proposed associations between features (stiffer-brighter or stiffer-darker). Afterwards, they better discriminated stimuli when they were presented with the associated feature, in comparison with a unimodal presentation (hence, there is no improvement in general discrimination) and with their baseline performance measured before learning. Another example, with a semantic transfer, comes from a study with synesthetes conducted by Mroczko, Metzinger, Singer and Nikolic (2009). The German participants learned three graphemes from the Glagolithic alphabet[32], two letters and a digit. They first learned the graphemes by handwriting them six times. Then, the participants learned the alphabetical correspondences of these graphemes by writing 20 words in German with a grapheme replaced by its Glagolithic equivalent (this took less than 10 min per grapheme). Then, the authors tested the learning via a Stroop task adapted to letter-color synesthetes: the naming of the ink color of a given grapheme is facilitated (i.e. faster) or impeded whether it corresponds to the color idiosyncratically induced within a particular synesthete. Participants were faster at naming the color ink that was congruent with the color elicited by the Glagolithic grapheme, which corresponded to the color elicited by the corresponding Roman grapheme. Hence, it is possible to transfer a synesthetic association within only 10 minutes – and it is a transfer and not a creation of a new association, since the concurrent color is the same for both graphemes (a given concept).

3.4.2. Facilitator and interferential congruency effects

Crossmodal correspondences can facilitate or interfere the processing of stimuli which are related to polarized dimensions. For example, pitch, loudness, brightness and size have two extreme poles, from the weakest intensity to the strongest intensity. A match between polarities of different dimensions leads to a congruent pair (e.g. loud and large), a mismatch leads to an incongruent pair (e.g. loud and small) (Marks, 2004). Different effects can appear through studies using discrimination or classification tasks, principally Garner interferences and congruence effects (see Marks, 2004 for a review).

The congruence effect refers to the facilitation to process a stimulus in one modality when this stimulus is accompanied by another congruent stimulus in another modality – congruent because they share congruent poles (e.g. a large object and a loud sound) rather than incongruent ones. This does not necessarily imply the reverse, i.e. interference effect from an

[32] The oldest Slavic alphabet known so far.

incongruent pairing. For example, Melara and O'Brien (1987) conducted a study in which visual position and pitch both varied and participants had to categorize one of the two modalities. The authors found facilitator effect between visual position and pitch for both judgment tasks when the stimuli varied congruently. However, they did not obtain interference (nor actually facilitation) from an incongruent pairing, i.e. the performance was similar to baseline.

Garner interference refers to a general performance decrease (longer response times and an increase in error rates[33]) that appears during the orthogonal presentation of two dimensions, indicating distributed attention (Garner, 1974, in Melara & O'Brien, 1987). For example, Melara (1989) compared four different presentations involving visual (colors) and auditory (pitch) dimensions: 1) the categorization of one dimension while the second is constant (baseline); 2) the categorization of one dimension while the second also varies between trials, but in a consistent way, either congruently, or 3) incongruently; 4) the categorization of one dimension while both dimensions vary (i.e. orthogonal presentation). When the second dimension is constant (baseline), there is no difference between congruent and incongruent pairings. However, the comparison between the baseline and the other conditions reveals that: 1) response times are faster in the case of congruency, regardless of which dimension is processed – visual or auditory; 2) response times are longer for incongruent pairings in color but not in pitch judgement task; 3) there is a Garner interference, i.e. in the orthogonal task, response times are overall longer.

However, some studies do not compare congruent and incongruent associations to a baseline (e.g., Marks, 1987), and provide only differences between these two types of associations. It is then complicated to conclude about the reality of facilitator effects deriving from congruent trials, and of interferential effects deriving from incongruent trials, without a comparison to a baseline. In fact, a difference between congruent and incongruent trials may originate from: 1) a facilitator effect from congruent trials; 2) an interference effect from incongruent trials; 3) both. The following studies have to be considered with this issue in mind.

In the size discrimination task of Gallace and Spence (2006), a congruent pitch sound increased response speed in comparison to an incongruent pitch and to the controlled condition (no sound). However, there was also an increase of speed with an incongruent pitch in comparison to the controlled condition, which would result from an *alerting effect* caused by

[33] They usually correlate positively (Marks, 2004).

the sudden onset of a sound (Posner, 1978, in Gallace & Spence, 2006). In this case, incongruent associations thus do not interfere with stimulus processing. However, Marks, Ben-Artzi and Lakatos (2003) reported differences between congruent, incongruent and baseline trials, using brightness and pitch differences. Generally, there were less errors for congruent trials, in comparison to baseline, which exhibited less errors than incongruent effects. These results show both facilitator and interference effects due to congruency. It is possible that these different patterns of results depend on the dimensions that are evaluated and their crossmodal relations (i.e. statistical, structural or semantic), but also on methodological differences and on the modalities of the stimuli (e.g. the different pitch values that are selected).

Melara and Marks (1990) conducted different experiments in order to evaluate the influence of semantics on a categorization task. The task was to categorize either linguistic stimuli (written syllables 'HI' and 'LO' or spoken words 'high' and 'low') or the other modality (pitch or spatial location – high and low). They reported the same results whatever the stimuli to categorize i.e. Garner interference and congruence effect. Hence, semantic labels ('high'/'HI' and 'low'/'LO') exhibited the same influence as pitch or spatial location. This may indicate that semantic processes are involved in perceptual cross-modal correspondences. This catches up with the semantic coding hypothesis proposed by Martino and Marks (1999), on which Nikolić (2009) leaned on for his hypothesis of ideasthesia.

Hirata, Ukita and Kita (2011) investigated *motivated* associations with Japanese speakers and more precisely the influence of consonants in discrimination of lightness, and of lightness on consonantal discrimination. The visual stimuli were white and black squares. The authors used oral syllables varying in consonants (because in Japanese a consonant is not '*usually*' pronounced alone) and opposed voiceless ones – *seion* or *han-dakuon* (か /ka/, さ /sa/, た /ta/, and ぱ /pa/) – to voiced ones – *dakuon* – (が /ga/, ざ /za/, だ /da/ and ば /ba/). One pair of syllables differing on voicing was assigned to each participant (e.g. /ka/ and /ga/) and she or he performed the two speeded discrimination tasks. In both discrimination tasks, there were four types of presentations: baseline, congruently correlated, incongruently correlated and orthogonal. In the lightness discrimination task, participants had to determine as fast as possible if the square was white or dark while one syllable was simultaneously presented; in the consonant discrimination task, they had to determine as fast as possible if the syllable

was composed of a voiceless or a voiced consonant[34], while one square was simultaneously displayed. The analyses revealed no congruence facilitation from consonants for the lightness discrimination task. However, for consonant discrimination, a facilitation effect appeared for congruent trials in the congruently correlated presentation and in the orthogonal presentations. The asymmetry between the two tasks may be explained by a difference in processing speed: visual stimuli were faster to categorize, which may not allow consonantal influence to take place. Even if pitch was similar across the syllables that were recorded for the purpose of this experimentation, the authors proposed as an explanation that the consonantal inherent pitch may explain the association. Indeed, voiceless consonants are usually pronounced with a higher pitch and high-pitched tones are associated to brightness, while voiced consonants are usually pronounced with a lower pitch and low-pitched tones are associated to darkness (Marks, 1987). Moreover, these results cannot be explained by a linguistic bias, since words related to lightness and darkness in Japanese only contain voiceless obstruents (*shiro* – white, *kuro* – black, *akarui* – bright, *kurai* – dark).

Another study is of particular interest, since it involves bouba-kiki associations. Kovic, Plunkett and Westermann (2010) evaluated learning of associations between pseudo-words and visual stimuli (composed of several spiky and curvy elements) via a categorization task. The learning phase was implicit but followed by a feedback, indicating if the answer was correct or not. Participants were assigned to a congruent or an incongruent condition. Based on previous studies, in the congruent condition, stimuli with round head-element had to be classified as 'mot' and those with a spiky head-element had to be classified as 'riff'. Then, the test phase consisted in categorizing pairs as matching or not matching. Results of the test phase showed a congruency effect on response time, but no difference in error rates. Participants were faster to answer to congruent trials when they had previously learned congruent pairings. Interestingly, the authors also found an interference effect of congruency: participants were slower to categorize congruent pairs as '*mismatching*' when they had learned incongruent pairings. Thus, congruent pairings exhibited a categorizing bias with different manifestations depending on learning: congruent pairings are faster to categorize as a match and slower to categorize as a mismatch.

[34] Japanese speakers are aware of the difference of voicing between two syllables like /ka/ and /ga/, which are visually marked in Hiragana (か *vs.* が) and belong to different categories as mentioned above.

Overall, while interference effects seem to vary according to studies, facilitator effects deriving from congruent associations seem well established. They depend on the dynamic or constant presentation of the second modality to appear. In other words, these effects are relative rather than absolute, which is the subject of the following subsection.

3.4.3. Relativeness or absoluteness of effects

In another experiment of Gallace and Spence (2006)'s study (see section 3.2.4), participants still had to discriminate sizes of circles (as 'larger' or 'smaller') while pitch sounds did not differ within a block of trials – only size did. There was thus only variation of pitch across blocks. In this experiment, there was no congruency effect. This suggests relative effects instead of absolute ones. Indeed, both modalities have to vary in order to bring congruent associations to light. If a modality is constant, it does not influence the processing of the second (i.e. a non-relevant modality is easily discarded when it does not vary).

Similarily, Melara (1989) reported systematically faster response times for congruent trials rather than for incongruent trials, when both dimensions varied (congruently, incongruently and orthogonally) (see section 3.4.2). There was no speed increase for congruent pairings in the baseline condition (i.e. when the second modality was constant). Hence, the facilitator effect for congruent trials is relative and not absolute, since it requires variation.

Marks (1987) also concluded in favor of relative effects. He tested the influence of pitch on judgments about lightness in two different experiments. The first one used only two different pitch frequencies: 200 and 360 Hz. The second one added two other frequencies: 100 and 800 Hz. The difference he found between 200 and 360 Hz in the first experiment was similar to the one found in the second experiment between 100 and 800 Hz (i.e. extreme poles). Hence, the effects depend on the extremities of the pitch range presented within a study and not on their actual values.

The last point leads to considerations about the polarity or continuity of the associations, which is the focus of the following subsection.

3.4.4. Polarity or continuity

While most studies focused on extremities of continua (e.g. 'high' vs. 'low'; antonyms; etc.), an alternative perspective comes from a study conducted by Thompson and Estes (2011). These authors investigated whether *motivated* associations about size were categorical or graded. They created CVCVCV pseudo-words composed of six, four, three, two or zero

segments known to be associated to largeness or smallness. Segments denoting largeness were [a, u, o, m, l, w, b, d, g] and those denoting smallness were [i, e, t, k]. For example, [wodolo] contains six segments associated to largeness, while [kuloti] contains three of those segments and [kitete] contains none. In one trial, an object was presented on the screen, beside a cow which served as a reference point. This object had five different possible sizes: 100%, 66%, 50%, 33% and 10% (50% was a size similar to the cow's). Participants had to choose the written pseudo-word that best matched the visual stimulus, among five possibilities. The authors obtained a gradient: the size of the object predicted linearly the number of segments denoting largeness (i.e. the larger the object, the larger the number of segments denoting largeness was). However, there was a positive and significant correlation between the width in pixels of pseudo-words and the number of segments denoting largeness, a potential confounding explanation. The results were thus replicated with spoken pseudo-words, though the paradigm differed on several points. The authors used CVCV pseudo-words composed of two, one or none 'large' syllables (one syllable was exclusively composed of 'large' segments [b, d, g, u, o] or 'small' segments [p, t, k, i, e]). In this experiment, only three different sizes were presented with a different reference point (i.e. smaller, larger or as large as a human being). First, the visual stimulus was displayed and then three spoken pseudo-words were successively presented to participants, who had to decide which one best matched the object. As the authors concluded, '*rather than crudely dichotomizing graded dimensions of objects (e.g., small and large), sound symbolism reliably conveys relatively fine distinctions along those graded dimensions*' (p. 2403).

A continuum in *motivated* associations relative to size instead of extreme poles was thus brought to light. However, one may wonder what would have been the results if the task was to choose one size among five for a given pseudo-word. Moreover, even if a continuum exists for *motivated* associations about size, it may not exist for other dimensions. Similarly, a continuum may appear or not according to the methodology and the task to complete. In Thompson and Estes (2011)'s study, participants had to associate a pseudo-word to a size. In Marks (1987)'s study – in which participants had to categorize lightness while a pitch sound was simultaneously produced – results were in favor of extreme poles instead of graded influences. For the previous stated reasons, these two studies cannot be directly compared.

In addition to methodological differences, the question of the cognitive level at which the associations appear is also of interest.

3.4.5. Perceptual or decisional influence

Some authors have tried to disentangle the possible cognitive origins of the previous effects and, more precisely, have distinguished between a perceptual level and a decisional one. Evidence in favor of both levels are reported in this subsection.

Marks (2004) proposed three levels at which cross-modal correspondences may occur in a speeded classification task: a perceptual level, a post-perceptual level and a decisional one. At a perceptual level, different modalities of a given stimulus could be simultaneously processed, which could provoke an interference effect. However, it is unlikely since early processes do not decompose a stimulus in its different modalities. Another possibility is that different modalities may have mutual effects on each other, i.e. the perception of a modality increasing or decreasing would be enhanced by the increasing or decreasing stimulation of another modality. At a post-perceptual level, an interaction could occur at a more abstract level, linguistic or semantic. Finally, it could appear at a decisional level based on perceptual or post-perceptual information. These different levels at which the effects could occur are not exclusive to each other, and this could depend on the modality of the stimuli and on their crossmodal relation.

Parise and Spence (2009)'s study investigated the association between pitch and size in a spatial identification task and a temporal identification task. In the temporal task, participants were exposed to two stimuli (one per modality) in different orders, i.e. the auditory stimulus was either the first one or the second one. The task consisted in determining whether the auditory stimulus was the second one to occur. When both stimuli were simultaneously presented, performance was similar to chance (50%). When there was an order, incongruently associated stimuli facilitated the task in comparison to congruent ones. The authors' explanation is that a multisensory integration masks the temporal sequence of congruent stimuli in different modalities. The discriminability of these stimuli is thus harder to process. In the spatial version task, participants had to determine the provenance (left or right) of an auditory stimulus (a low- or high-pitch sound) while a visual stimulus varied in size – congruently or incongruently – with the sound. Congruent pairings led to better discriminability in the spatial localization of sounds. These results also support a multimodal integration and thus a perceptual influence from associations, instead of a decisional one.

Marks deepened the question whether effects appeared at a sensory or at a decisional level in two studies (1987, 2004). In a speeded (the maximum response time was 1 sec.) sensory discrimination task (1987), participants had to categorize either a visual stimulus or an auditory one, while respectively an irrelevant auditory or visual stimulus was simultaneously varying. He found bidirectional effects: an irrelevant visual stimulus influences the response to an auditory one, and conversely, an irrelevant auditory stimulus influences the response to a visual one. However, in an unspeeded sensory discrimination task (2004), he found a unidirectional effect instead: the visual modality influenced the auditory discrimination, i.e. congruent stimuli provoked more hits (i.e. good answers) and incongruent ones provoked more false alarms. However, the reverse pattern did not show up, i.e. irrelevant variations of pitch or loudness did not systematically influence brightness discrimination. According to Marks, this difference may point to a decisional bias rather than a sensitive one, due to speed stress. An additional experiment in his 2004's study involved a different (unspeeded) task: instead of answering 'high' or 'low', participants had to judge two stimuli and decide if they were the same or different (on brightness or pitch). In this case, no congruency effect appeared at all (i.e. no difference between congruent and incongruent trials). All in all, Marks concluded in favor of late decisional processes to explain cross-modal interactions.

However, in Gallace and Spence (2006)'s study, another discrimination task belies this last finding: participants had to determine whether a stimulus was 'different' or 'similar' instead of 'larger' or 'smaller' than the other stimulus (with the simultaneous presentation of high- or low-pitched sounds). Hence, there was no congruency between the visual stimulus and the answer. A change in the type of response did not change the results (i.e. the pattern of results with 'different' or 'similar' was similar to those obtained with the answers 'larger' or 'smaller') – although response times were longer than in the three other experiments – which sustains the perceptual hypothesis. This means that the congruency is not between the answer and the visual stimulus, but rather between the visual stimulus and the auditory stimulus. Nonetheless, a difficulty arises when it comes to comparing different studies, using different methodologies and material. What holds with brightness or pitch contrasts might not with size contrasts.

Either way, the two following studies conducted by Marks (1987) and Parise and Spence (2009) both brought to light congruency effects between pitch and shapes ('bouba-kiki'-like shapes).

Marks (1987) used reversed and asymmetric ‘U’ and ‘V’ shapes that were simultaneously displayed with high- or low-pitched sounds. Participants had to decide whether the shape was round or spiky. Interactions between pitch and shapes appeared both in reaction times and error rates. More precisely, we can see from the graphics (see Figure 2) – no statistical tests have been conducted – that the high pitch sound (360 Hz) seems to speed up the recognition of a spiky shape and to decrease the amount of errors compared to low pitch sound while it increases the number of errors for round shapes (without increasing reaction times much though).

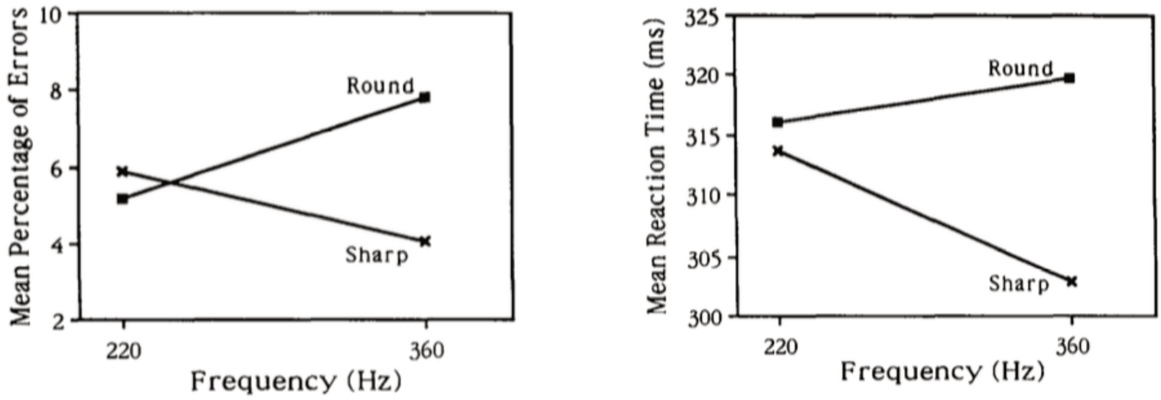


Figure 2. Mean percentages of errors and response times in Marks' study (1987). ‘Sharp’ corresponds to a spiky shape.

Parise and Spence (2009)’s study also tested the congruency between pitch and ‘bouba-kiki’-like shapes using a temporal identification task. They found results similar to the pitch-size experiment: congruent associations were harder to process than incongruent ones, which supports a multisensory integration between pitch and visual shapes, more precisely between high-pitch tone and angular shapes, and between low-pitch tone and curvilinear shapes.

All in all, the categorization of spiky and round shapes is modulated by pitch (Marks, 1987) and the temporality of congruent associations between shapes and pitch is harder to process (Parise & Spence, 2009). This last point supports a multisensory integration, hence processes at a perceptual level, even though it is hard to decide whether the influence from one modality to another appears at a perceptual or a decisional level because of the discrepancies that appear between the different studies reported above.

The next subsection focuses on other experimental paradigms that aim at better figuring out the level at which *motivated* correspondences can occur. Instead of perceptual and decisional effects, the next studies address *motivated* associations using implicit paradigms instead of associative tasks.

3.4.6. Implicitness vs. Explicitness

Schmidtke, Conrad and Jacobs (2014) pointed at the major limit of most studies about the bouba-kiki effect (which used associative tasks): they are '*offline*' measures that can reflect metacognitive strategies. However, some studies have attempted to assess the implicit existence of some *motivated* associations. This subsection focuses on two types of implicit paradigms, lexical decision task and priming.

3.4.6.1. Lexical decision task

Westbury (2005) conducted a study that consisted in a lexical decision task with strings of characters displayed in either spiky or round frames. The strings of letters – words and pseudo-words – varied in their phonological compositions: they were composed of either plosives or sonorants (or mixed in the case of words). Westbury obtained an interaction between the type of frame and the phonological composition of words: pseudo-words composed of plosives were faster to categorize (as pseudo-words) when they were displayed in spiky frames, while those composed of sonorants were faster to categorize in round frames. This experiment constituted a major evidence in favor of low-level integration of *motivated* associations, even though effect sizes were quite weak.

3.4.6.2. Priming

Another evidence of implicit associations comes from a priming task involving phonesthemes which was conducted by Bergen (2004). Participants had to decide if a string of characters was a real word or not. This string was preceded by a written prime stimulus for 150 ms (*'just long enough to be barely perceived by the subject'*, p. 196), which was followed by an interstimulus interval of 300 ms. Primes and targets could share phonological or semantic properties, or both, or none (see Table 6).

Table 6. Conditions used in Bergen's study (2004).

Conditions	Semantic feature sharing	Phonological onset sharing	Phonestheme sharing	Examples of pairs
Phonestheme				glitter - glow
Form				druid - drip
Meaning				cord - rope
Pseudo-phonestheme				crony - crook
Baseline				frill - barn

Participants were faster to recognize words preceded by words with which they share a phonestheme, in comparison to the four other conditions. Bergen concluded that the knowledge

of a phonestheme meaning is accessed unconsciously (because it was barely perceived), even under a time pressure, and that this is not imputable to individual semantic or phonological priming effects. To quote Bergen, *'phonesthemes are a testament to the diligence of the human ability to encode and use subtle statistical associations in the linguistic environment*' (p. 307).

Another study, conducted by Hung, Styles and Hsieh (2017), aimed at investigating the possibility to highlight the implicit nature of the bouba-kiki effect with two types of paradigms: continuous flash-suppression and visual masking. The first one consists in displaying simultaneously two stimuli, one dynamically changing to the dominant eye and another static to the nondominant eye. The static one is the target and is gradually intensified until it is perceived despite the dynamic one. In practice, the target was a pseudo-word ('bubu' or 'kiki') that was displayed within a shape (round or spiky) while a Mondrian stimulus[35] was presented to the dominant eye. Participants had to answer when they detected the target. Congruent stimuli were detected faster than incongruent ones. The authors therefore concluded that congruence is processed before conscious perception. These results were replicated in a second experiment that used another writing system whose letters exhibited as much curves as acute angles, which meant that the initial effects were not imputable to the shape of letters. Participants learned to match two written pseudo-words in this alphabet with the two auditory pseudo-words 'bubu' and 'kiki'. Hence, results are due to the relation between phonological composition and visual features – without incidence from the shapes of letters. The second paradigm – visual masking – consisted in presenting an auditory word form 150 ms before, simultaneously or 150 ms after a brief masked visual shape presentation (33 ms), on the right or on the left of a fixation point. Participants had to determine if they had seen something and if it was on the right or on the left (asking to specify the location allowed the identification of false alarms). The only congruence effect was found when the auditory word preceded the visual shape. Hence, a congruent auditory form reduced the detection threshold of a visual stimulus not consciously perceived.

Similarly, Sidhu and Pexman (2017) investigated the influence of the supraliminal priming of a pseudo-word on the categorization of shapes. Pseudo-words were either written or oral. In the case of written pseudo-words, they were displayed for 1500 ms, followed by a blank screen (500 ms) and then by a shape to categorize. In the case of oral pseudo-words, the shape to categorize immediately followed the auditory presentation. In both cases, the composition of the pseudo-words influenced the categorization of an ambiguous shape (that was as round as

[35] A composition of colored rectangles.

spiky), consistently with previous findings in this line of experiments: [b, m, o, u, a] are related to round shapes, [t, k, i, ə, eI] to spiky shapes.

Implicit paradigms – lexical decision task, continuous flash-suppression, visual masking and priming – thus bring evidence that associations do not derive from metacognitive strategies and exist at a lower level.

3.5. *Summary on the nature and effects of crossmodal correspondences*

Section 3.4 exposed that: 1) correspondences are easily modulated, 2) both facilitator and interferential effects are possible, depending on paradigm and material, 3) associations are relative and not absolute, hence they always have to be considered in context, 4) it is not clear whether these effects appear at a perceptual or decisional level but some evidence are in favor of low-level processes (lexical decision task and priming) in addition to high-level ones (associative task). All in all, experimental studies about crossmodal correspondences, and more specially those implying *motivated* associations, involve very diverse paradigms resulting in various, sometimes contradictory, outcomes. The next section further exposes this diversity, the difficulties it leads to, and finally the contribution of this thesis.

4. Experimental and methodological contributions in the study of *motivation*

4.1. *Discussion about the previous findings*

The bouba-kiki task is a very harnessed paradigm that allows one to obtain results across populations of speakers of different languages and at different ages (including young children). These studies constitute a way to gather cross-cultural and cross-linguistic evidence, and to formulate hypotheses about the ontogenetic and phylogenic emergence of language. Most studies converge in their results and highlight the *motivated* nature of some segments (the best-established associations are the voiceless plosives with spiky shapes and the sonorants with round shapes). However, most studies involve participants from Western countries and pseudo-words compatible with the phonology of the experimenters' languages. The few discrepancies found in the literature highlight the necessity of conforming the material to the population being studied. Meanwhile, the conceptual contrast that is widely used – the spiky and round shapes – may seem quite distant from phylogenetic language emergence and from evolutionary hypotheses. It is unlikely that the first interactions were about such types of shapes. For this

reason, it is difficult to elaborate hypotheses on the emergence and evolution of language on this basis. Beyond the common association task, these studies differ on several points: 1) the contrasted segments or phonetic features; 2) the populations studied (culture, language, age…); 3) the type of presentation (see Table 7 which compiles examples of diverse studies and accounts for a wide variety of phonetic material, populations and paradigms). Regarding the later especially, there are four major types of stimuli presentation. First, the original bouba-kiki paradigm used a 2x2 paradigm (Köhler, 1947): two shapes and two pseudo-words were presented and participants had to decide which shape and which pseudo-word matched best. Hence, two contrasts were present at a time. Second, it is possible to show a pair of shapes with only one pseudo-word (2x1), hence no phonetic contrast (e.g. Fort et al., 2015). Third, the opposite, presenting only one shape but a phonetic contrast, is a 1x2 paradigm (e.g. Chen et al., 2016). Finally, a trial that consists in presenting only one shape and one pseudo-word (1x1) exhibits no contrast within trial (e.g. Asano et al., 2015), but the contrast can appear across trials (a succession of spiky and round shapes).

It can be hard to disentangle the respective impacts of these experimental differences. While some studies have already aimed at assessing the complexity induced by phonetic contrasts and the diversity of populations with more systematic approaches or meta-analyses, the impact of the paradigm has never been delved into, despite the differences across studies. Nevertheless, the study of Aveyard (2012) provided insights about the impact of methodological differences. The author conducted two paradigms, one in which one pseudo-word was presented with two visual stimuli (a round and a spiky shape), and a second one in which one pseudo-word was presented with four visual stimuli (a target, a distractor of the same category of shape and two distractors of the other category). Participants had to choose the proper shape for each pseudo-word through several trials, based on repeated post-trial feedback and across three blocks of learning. The matching pairs were either congruent or incongruent, according to findings of previous studies. In the first paradigm, there was an advantage (i.e. faster learning) for congruent matching pairings from the first block, but the performance also improved for incongruent matching pairings across the three blocks. In the second paradigm, participants were less effective and slower in the learning of the matching pairs. Improvement only occurred for the congruent matching pairings. This study thus points to the impact of experimental settings, which can modulate the highlighting of *motivated* associations.

Table 7. Examples of studies differing on population, paradigms and phonetic material. C. stands for consonants and V. for vowels.

Publication	Population	Paradigm	Phonetic material
Ahlner & Zlatev, 2010	Swedish speakers	2x2 presentation but 2x1 task[36]	C: [p, t, k, tʃ, l, m, n, ŋ] V: [i, u]
Asano et al., 2015	11-month-old Japanese children	1x1 (EEG – preferential looking paradigm)	'moma' and 'kipi'
Aveyard, 2012	English speakers of Sharjah university (United Arab Emirates)	2x1 and 4x1	C: [p, b, t, d, k, g, l, w, r, s, f, h]
Bottini et al., 2019	Sighted and blind Italian speakers	2x2	'maluma' and 'takete'
Bremner et al., 2013	Himbas (Otjiherero speakers)	2x2	'bouba' and 'kiki'
Chen et al., 2016	American and Taiwanase speakers	1x2	'bouba' and 'kiki'
Fort et al., 2015	French speakers	2x1	C: [p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, l, m, n] V: [i, e, o, u]
Knoeferle et al., 2017	English speakers	5x1 (two shapes, a 5-point scale)	C: [w, j, l, r, m, n, z, v, ʒ, ð, b, d, g, s, f, ʃ, θ, p, t, k] V: [a, u, o, e, i]
Kovic et al., 2010	English speakers	1x1 (decision matching task after learning phase)	'mot' and 'riff'; 'dom' and 'shick'
Sidhu & Pexman, 2017	English-speaking Canadian	2x1 and 1x1 (categorization of ambiguous shapes)	C: [b, m, t, k, dʒ, f, h] V: [o, u, ɑ, i, ə, eI]
Vainio, Tiainen, Tiippana, Rantala, & Vainio, 2017	Finish speakers	1x1 (pronunciation task)	[ti] and [mɑ]; [i] and [ɑ]; [te] and [me]; [i] and [u]

[36] Participants had to choose one shape given a pseudo-word selected by the experimenter among two.

Similarly, different paradigms may induce different cognitive processes, depending on which type of stimulus is the target of the choice. Indeed, choosing between two shapes for one pseudo-word is different from choosing between two pseudo-words for one shape. Also, an associative task is different from a judgment task about the matching between a shape and a pseudo-word. At a cognitive level, particularly interesting studies are those using a 1x1 paradigm, because they make use of different types of cognitive processing: judgment (e.g. Perfors, 2004), categorization (e.g. Kovic et al., 2010; Sidhu & Pexman, 2017), learning (Nielsen & Rendall, 2012), lexical decision task (Westbury, 2005). Also, 1x1 may allow to record '*online*' – direct – associations (via limitations of time response, priming…) by preventing metacognitive strategies, because of the lack of overt contrast. The use of this type of presentation is thus interesting in order to investigate the existence of some associations, and the level at which they appear.

4.2. *A more ecological approach in an evolutionary perspective*

While bouba-kiki shapes are distant from scenarios of language emergence and evolution, communications about emotions, body size or animals may be more relevant. For this reason, the first study that is outlined in this thesis used animals as conceptual material instead of shapes, in order to adopt a more ecological approach. Animals actually allow to study several contrasts simultaneously: dangerousness, repulsiveness, size and biological class (e.g. fish *vs*. birds). Accordingly, we based our hypotheses on studies about emotions (Adelman et al., 2018; Fónagy, 1961, 1983), size (Brent Berlin, 1994; Blasi et al., 2016; Haynie et al., 2014; Iwasaki et al., 2007; Knoeferle et al., 2017; Ohala, 1997; Sapir, 1929; Vainio et al., 2017) and biological class (Brent Berlin, 1994; Blasi et al., 2016).

4.3. *A wide variety of paradigms in the study of motivation*

The second study focused on the impact of the variety of paradigms of presentation of stimuli, in the continuity of the first study. With a comparison of the four types of presentations introduced above in section 4.1, it aimed at evaluating the importance of contrasts between pseudo-words and contrasts between concepts to be associated to these pseudo-words. The population, monolingual French speakers, as well as the segmental and conceptual (labels denoting types of animals) material were similar across the four protocols, in order to clearly assess and single out the influence of the presence of a phonetic contrast (1x2), of a conceptual contrast (2x1), of both (2x2) and of their absence (1x1).

4.4. *The cognitive level involved in the emergence of associations*

The third study is based on Westbury's study (2005). It is a decision lexical task, with strings of letters displayed in frames with a spiky or round shape. Westbury found an interaction between the phonological composition of pseudo-words and frames: participants were faster to answer to a pseudo-word composed of voiceless plosives in a spiky frame, and to answer to those composed of sonorants in a round frame. The same methodology was used but another parameter was added: the font in which strings were displayed, in order to assess Cuskley et al., (2015)'s hypothesis according to which the shapes of letters induce associations. Hence, the font was either angular or curvy. This study aimed to investigate the existence of bouba-kiki associations at a low-level via an implicit paradigm, and to evaluate the potential influence of the shapes of letters.

4.5. *The purpose of this thesis*

In line with our three studies, the main questions this work aims to answer are the following:

1) Which *motivated* associations may emerge from an ecological approach using concepts and dimensions potentially relevant for early human communication? What do they possibly tell us about the origin and evolution of language?
2) Do differences in the presentation of stimuli influence the highlighting of *motivated* associations, more specifically depending on the presence – or absence – of linguistic segmental contrasts *or* of conceptual contrasts?
3) Does a supraliminal priming of shape influence the processing of pseudo-words according to their phonetic composition, or according to their graphemic composition, in a lexical decision task?

Experimentations

First study:

De Carolis, L., Marsico, E. & Coupé, C., 2017, Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features, *Language & Communication, 54*, pp. 21-35

Second study:

De Carolis, L., & Coupé, C., submitted to Cognition in May 2019, Phonetic and conceptual contrasts in the assessment of sound symbolic associations: comparing protocols and inferring cognitive processes

Third study:

De Carolis, L., Marsico, E., Arnaud, V. & Coupé, C., 2018, Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm, *PLoS ONE*, 13:12, e0208874

1. First study: Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features

Léa De Carolis[1], Egidio Marsico[1] & Christophe Coupé[1]

[1] Laboratoire Dynamique du Langage, CNRS & Université de Lyon, Lyon, France

Article published in *Language & Communication, 54*, pp. 21-35, in 2017

NB: in this article, the use of '1x2' differs from the rest of the thesis: it refers to one pseudo-word and two concepts instead of one concept and two pseudo-words.

Language & Communication 54 (2017) 21–35

Contents lists available at ScienceDirect

Language & Communication

journal homepage: www.elsevier.com/locate/langcom

Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features

Léa De Carolis, Egidio Marsico, Christophe Coupé*

Laboratoire Dynamique du Langage, Institut des Sciences de l'Homme, 14 Avenue Berthelot, 69007 Lyon, France

ARTICLE INFO

Article history:
Available online 10 January 2017

Keywords:
Non-arbitrariness
Multimodality
Sound symbolism
Origins of language
Association task
Animal names

ABSTRACT

Contradicting Saussure's arbitrariness of the linguistic sign, sound symbolism – the systematic association of sounds with meanings – is consistently found across languages. It may have offered a ground for our ancestors to develop an initial communication system, and later move toward symbolic signs. We tested sound symbolic associations in French between phonetic segments or phonetic features and various attributes of animals (size, dangerousness…). A first experimental setting revealed no significant association, while a second did. These associations furthermore do not appear in French animal names. We discuss these results in the light of scenarios of language origins and evolution.

1. Background: sound symbolism and the origins of language

1.1. Studies on sound symbolism

1.1.1. How arbitrary are linguistic signs?

For a century, theoretical linguistics has built on Saussure's heritage about the arbitrariness of linguistic signs. In doing so, it took for granted that the relation between the 'signifier' and the 'signified' is arbitrary and conventionalized (De Saussure, 1916). This undermined the possibility of what philosophers of language have called 'natural' or 'motivated' signs, i.e. signs underpinned by some non-arbitrary principle(s) of association.

Although a major part of the lexicon of a language clearly relies on arbitrary associations between a mental representation and an 'acoustic image', there is however evidence of the existence of motivated signs in human languages. Most of them – if not all – use at least some signs that present a non-arbitrary relation between their sounds and their meaning. In onomatopoeia for example, an iconic relationship exists between the phonetic shape of the word and the sound emitted by what it refers to, e.g. the sound produced by an animal. In ideophones or phonaestemes, a strong similarity of shape between the signifier and the signified may not exist, but recurrent principles of association do, as for example when words related to 'vision' and 'light' in English often contain the phoneme cluster /gl/ (Schmidtke et al., 2014). The frequency of occurrence of these phenomena varies across languages, with some languages such as Japanese being very rich in sound symbolic words (Kantartzis et al., 2011).

A few conceptual differences need to be highlighted. First, motivation and iconicity may be synonyms in some texts, but while iconic signs are motivated, not all motivated signs are necessarily iconic. Second, Ahlner and Zlatev (2010) stress that

* Corresponding author.
E-mail addresses: Lea.De-Carolis@univ-lyon2.fr (L. De Carolis), Egidio.Marsico@cnrs.fr (E. Marsico), Christophe.Coupe@cnrs.fr (C. Coupé).

http://dx.doi.org/10.1016/j.langcom.2016.10.003
0271-5309/

the notion of convention has often been wrongly equated to arbitrariness. However, that non-arbitrary sign be conventionalized in a population of speakers is a perfectly viable option.

Sound symbolism is the expression commonly used to refer to non-arbitrariness. According to Ohala (1997), this is 'the term for a hypothesized systematic relationship between sound and meaning'. Different articulatory or acoustic properties of speech sounds may be associated to various *ontological*[1] properties of objects or events.

The previous considerations relate to oral languages, but sign languages should not be left aside. Indeed, they are known to contain a large number of non-arbitrary signs in addition to arbitrary ones. For many, the *visual* shape of the signifier resembles those of the signified, in a similar way the *acoustic* shape of a vocal sign may resemble the sound(s) made by a referent.

1.1.2. Evidence of sound symbolism

Beyond the previous evidence readily available in the lexicon of many languages, Sapir (1929) showed experimentally nearly a century ago that phonemic contrasts could also be mapped to physical and more generally ontological properties of things. Among others, he explained how English speakers associated [a] with large things, and [i] with small ones. His investigations were the first steps of a series of experimental studies that gave further voice to the idea of non-arbitrariness in language. To this day, two main experimental protocols have been employed: *association tasks* and *phonetic judgment tasks*.

A famous example based on an association task is the so-called '*bouba-kiki*' effect. It consists in the simultaneous presentation of two shapes, one spiky and the other curvy, and two pseudo-words 'bouba' and 'kiki' – in the original study, Köhler (1947) actually used 'maluma' and 'takete'. According to Ramachandran and Hubbard (2001), speakers of different languages and cultures are overall very consistent in their choice: 95% of surveyed people choose to associate the curvy shape with 'bouba' (or 'maluma') and the spiky one with 'kiki' (or 'takete'). Although this study has often been cited after its publication, it does not offer details about the experimental setting, the number of people surveyed, their gender, the language(s) they spoke etc. Other scholars have however refined the basic setting upon two particular aspects: on the one hand the phonetic differences between the pseudo-words, and on the other hand the explicit nature of the task. Regarding phonetic differences, subsequent studies implemented better control of the phonetic content of the pseudo-words in order to disentangle the role of consonants and vowels – if bouba is associated with round shape, is it because of the [b] or of (one of) the vowel(s)? This implied using only one vowel and one consonant, and also descending at the level of phonetic features (Nielsen and Rendall, 2011; Nobile, 2015; Ozturk et al., 2013). Regarding the task itself, the *bouba-kiki* experiment is very explicit in the sense that subjects can immediately notice the difference(s) between the two shapes, and between the two pseudo-words. This can induce the use of elaborate strategies which depart from the more intuitive judgments of sound symbolism made by subjects in other contexts. To avoid such strategies, Westbury (2005) used a lexical decision task, which was much more implicit in its design, and still obtained significant associative effects.

Other experiments are based on phonetic judgment tasks. This is how Sapir and his student Newman shed light on the relations between some phonemes and some ontological properties of things (Newman, 1933). In reviews of such studies (Nuckolls, 1999; Ozturk et al., 2013; Spector and Maurer, 2013), as well as in cross-linguistic surveys of the relevant vocabulary (Nuckolls, 1999; Tanz, 1971), different types of correspondences are reported across languages: some phonemes relate to distance, others to brightness or elevation, others yet to properties such as nice, bitter etc. For example, Fónagy (1983) collected subjective judgments about phonemes from Hungarian speakers. He found that [i] is 'little', 'agile', 'nice' and that [u] is 'big', 'corpulent', 'obtuse', 'sad', 'dark', 'strong', 'bitter'. The same author also compared the distribution of phonemes in poems judged as 'aggressive' or 'tender', and found that aggressive poems contained a greater proportion of voiceless plosives like [t, k], whereas tender ones included more sonorants like [m, n, l] (Fónagy, 1961).

Berlin (1994) noted some specific phoneme distributions in animal names in south-American languages like Huambisa. In this language, names of fish contain more phonemes and syllables of low acoustic frequency ([a], [ku], [ka]), more nasals ([n], [m]) and more continuants ([s], [r]). Names of birds contain more phonemes and syllables of high frequency ([i], [pi], [t], [ts]), stops and affricates (with some differences between initial and final positions in words). Moreover, the frequency of [i] correlates with size across species: the names of the smaller fish and birds contain more [i] and the names of the bigger ones more [e], [a] and [u]. These distributions are also found in three other languages, two from South America, and one from Mexico. To elaborate on these findings, Berlin conducted an experiment with English-speaking students. They were presented a list of pairs of Huambisa names (explicitly one bird and one fish), and had to indicate which one referred to a bird. Their accuracy rate reached 8% above chance threshold and was highly statistically significant. This meant that English speakers could most often guess the biological class of an animal only from the phonetic composition of its name in Huambisa. Cross-culturally again, experiments with the Himba people of Northern Namibia revealed that they produce the same answers as Westerners for the bouba/kiki association task, but differ from them when it comes to associating angular and round shapes with water carbonation or food bitterness (Bremner et al., 2013). Himba have had little contact with Western culture, and the study shows that cross-modal associations are not necessarily universal.

Some of these associations appear early during development. In particular, Ozturk et al. (2013) showed that 4-month infants consistently distinguish between congruent and incongruent sound-shape mappings in a looking time task. That

[1] We use this term to encompass physical properties such as size, shape or color, as well as properties such as dangerousness, attractiveness, beauty, value etc.

sound symbolism be present in a wide range of languages, but also early in cognitive development, suggests that it rests on well-anchored cognitive processes that deserve investigation.

1.1.3. Possible mechanisms underlying intra-modal and cross-modal sound symbolism

The debate is still open with respect to the cognitive processes which explain sound symbolic associations. Different processes may actually underlie these associations depending on the relationship between the signifier and the signified.

In some cases, an obvious iconic relationship – i.e. a relationship of similarity – can take place *within* the acoustic modality, as in the case of onomatopoeia. In other cases, the association is *cross-modal*, and relies on iconicity to varying extent. For example, a relation of similarity may exist between the referent and the physical articulation of the sound in the vocal tract: in the 'bouba-kiki' experiment, the rounding of the lips for [u] can be said to resemble the curvy shape, while the spikes relate to the closure of the velar plosive [k]. Berlin also argued that the particular distributions of phonemes observed in animal names in Huambisa is reminiscent of the movements of fish and birds: fish display a sinusoidal motion at low visual points, which resembles the articulation of fricatives, and birds an energetic and fast displacement at high visual points, which is reminiscent of the mode of articulation of plosives. In yet other cases, a cross-modal relationship may be based on an indexical relationship: the fact that large individuals tend to produce lower frequency sounds, and small individuals higher frequency ones, is the underlying basis of the 'frequency code' hypothesis, which states that the value of the fundamental frequency of the acoustic signal is associated with the size of its referent (Ohala, 1994). Some other systematic relationships between a sound and an ontological property may be hard to account for with an iconic or indexical relationship, e.g. phonemes expressing a degree of brightness or a distance.

The proposed nature of a relationship needs to be assessed carefully. Recently, Cuskley et al. (2015) have argued that participants to the bouba/kiki experiments simply relied on the visual similarity between the shapes and the letters of the pseudo-words, rather than on sound symbolism per se. What may seem to be sound symbolic at first might in some cases turn out not to be.

Explanations for cross-modal sound symbolism can be considered at a neurophysiological and/or cognitive level. They can then be linked to a broader cognitive phenomenon, namely cross-modal correspondences (Spence, 2011). These cross-modal correspondences establish links between concepts and percepts, and may thus provide a unified representation of a multimodal entity (hearing the bouncing of a ball also activates the vision of the ball, and to some extent the motor actions to make it bounce). Three major kinds of cross-modal correspondences can be defined: structural, statistical and semantic. Structural correspondences are possibly innate and depend on neural systems through spreading activation between nearby brain areas, or similar mechanisms overlapping different modalities. For example, cognitively assessing either the magnitude of a sound (loudness) or the magnitude of a light (luminance) produces an increased neural firing in both respective areas. This in turn creates a correspondence between these two sensory inputs. Statistical correspondences result from associative learning and natural correlations in the environment. Finally, semantic correspondences result from language acquisition and correspond to common descriptions of distinct perceptual modalities. However, it is often difficult to assess whether a semantic correspondence does not derive from a structural or statistical correspondence, and whether the former does not in return reinforce the latter. For example, a lot of languages use words as 'low' and 'high' for both visual elevation and sound pitch. Is it the case purely because of linguistic reasons, or because of partial shared brain encoding for vision and sound? Whatever the answer, different correspondences may lead to different forms of sound symbolism, which may coexist in speakers and in their lexicon. This may explain why some associations may be universal (those based on structural correspondences) and others culture-specific (based on semantic correspondences).

Synaesthesia is a well-known manifestation of cross-modal correspondences and, according to Ramachandran and Hubbard (2001), it is the key element to explaining sound symbolism. It is a cognitive phenomenon in which two or more senses are associated at the experiential level, e.g. letters and colors, numbers and spatial positions, music and shapes etc. Hence, hearing a word can, for example, elicit a taste. Studies using imaging reveal activations of the expected areas (given the stimulus presented) and of additional brain areas corresponding to perceptions reported by synaesthetes (Spector and Maurer, 2009). Synaesthesia is much more common than previously believed (Simner et al., 2006). It seems that mechanisms underlying synaesthesia are universal, but particularly pronounced in children and in adult synaesthetes.

1.2. Non-arbitrary signs as a prerequisite for the emergence of oral communication

1.2.1. The emergence of a symbolic vocal code

Beyond the accumulation of evidence for non-arbitrary linguistic signs in modern languages, arbitrariness and non-arbitrariness point at the origins of our communication system. Indeed, one of the key questions regarding the origins of language concerns the emergence of linguistic conventions: **how did a common symbolic vocal code initially appear in a group of humans?** This issue has been addressed abundantly in the literature. It has been shown that many animals, whether primates, dogs, parrots, dolphins etc. (Herman, 2009; Kaminski et al., 2004; Pepperberg, 2000; Savage-Rumbaugh and Lewin, 1996), are able to learn symbols for communicative means. Nevertheless, this behavior is the result of human intervention, and does not occur in the wild, i.e. animals do not create symbols by themselves. In archaeology, early shell beads and other ornaments suggest symbolic behaviors more than 100,000 years ago (Botha, 2008; Bouzouggar et al., 2007; Vanhaeren et al., 2006). However, relating such behaviors to symbolic communication or a 'fully syntactical language' rests on weak inferences (Botha, 2008), and the lack of 'linguistic fossils' blinds us with respect to what existed before and after such clues appear in the

archaeological register. In the field of computer modeling, it has been demonstrated how conventions shared by a whole group can emerge from repeated pairwise interactions, i.e. without collective mechanisms of convergence (de Boer and Zuidema, 2010; Kirby and Hurford, 2002; Oudeyer, 2013; Steels, 2008). However, the symbolic nature of the communication code is hardcoded in the simulation from the start, and does not emerge from repeated interactions. All in all, the emergence of symbolic vocal communication therefore still remains elusive.

1.2.2. From iconic to symbolic codes of communication

Authors like Bouchard (Bouchard, 2013a, 2013b) have explicitly addressed how the code may have experienced changes in its nature through time. Following him and others, we argue that lexicons have not always predominantly relied on arbitrariness. On the contrary, we state that the initial linguistic signs were most probably iconic, or that there existed other principles of association between them and various properties of their referents, i.e. a 'motivation'.

In his seminal book 'The order of things', Foucault sketched out a possible scenario of the origins of language, linking the 18th century French philosophical tradition with the (yet to come) modern cognitive science (Foucault, 1989):

> *'As long as it is a simple extension of the body,* ***action*** *has no power to speak: it is not language. It becomes language, but only at the end of definite and complex operations: the* ***notation of an analogy of relations*** *(the other's cry is to what he is experiencing – that which is unknown – what my cry is to my* ***appetite*** *or my* ***fear****); **inversion of time** and voluntary use of the sign before the representation it designates (before experiencing a sensation of hunger strong enough to make me cry out, I emit* ***the cry that is associated with it****)'* (Emphasis is ours).

According to Foucault, language would have originated *via* an analogy made by a human being between the overt expression of an emotion or a sensation (a 'cry') emitted by someone else *and* his own cry in the same state of hunger or fear.

This analogy is possible only because 'the cry is associated with' the sensation by a special link. For Foucault, this link lies in 'action' (which in more contemporary terms could translate as a 'language of action' and a form of 'embodied cognition') and guaranties that the cry will always be more or less the same in a given situation. This is the first fundamental condition for this cry to later 'represent' the sensation. This point is crucial to us: for a convention to emerge, i.e. to reach a collective agreement on a sound-to-meaning association, vocal signals must be as stable as possible. Our hypothesis is that in the first stages of language, a motivated relationship existed between sound sequences and meanings, i.e. sound symbolic relations, and warranted this requirement of stability.

The second condition is the cognitive ability to build an analogy between somebody else's cry and someone's own sensation. At the level of the neural equipment, mirror neurons seem a likely candidate, as by linking production and perception they somehow assume the role of an internal representational bridge (Rizzolatti and Arbib, 1998). Crucially, this bridge can be crossed both ways, from external stimuli to internal states and vice-versa, something Arbib refers to as the parity requirement – what counts for the speaker must count approximately the same for the hearer (Arbib, 2005). This creates the path toward communication, and also more generally to an understanding of someone else's internal state (Gallese and Goldman, 1998).

An important question is left unanswered by neurophysiological approaches: why would there be a link associating a mental representation with a particular vocal response at all? As Bouchard states: 'a sign is a link between elements from domains of very different natures – physical/perceptual and psychological/conceptual. […] The key question for the origin of language is how these very different elements came to meet in the brains of humans to form linguistic signs' (Bouchard, 2013b, p. xi). Bouchard's main argument involves the development of representations via an 'Offline Brain System (OBS)', which stores representations that can be activated even without the related percept: 'With OBS, it is not the percept per se that is linked with a concept in a linguistic sign, but a representation of the percept; i.e., a mental state corresponding to it'. In other words, the link between a cry and its meaning is mediated through a representation. This proposal adds a representational layer to the involvement of mirror neurons in connecting communication signals to internal states, or to the synchronization of neural discharges as previously mentioned for multi-sensorial integration.

1.2.3. From ontogeny to phylogeny

Children are sensitive to specific correspondences between sound and shape (Maurer et al., 2006) and it has been shown that sound symbolism guides infants' word learning (Imai et al., 2008; Miyazaki et al., 2013). Although ontogeny does not recapitulate phylogeny, this can be seen as an argument in favor of a scenario of language emergence anchored in sound symbolic associations. Indeed, infants face the same problem as our ancestors: how to make sense of their environment and build a system of signs to communicate with others. In their *sound symbolism bootstrapping hypothesis*, Imai and Kita (2014) stress that sound symbolism help infants grasp the referential nature of speech sounds and navigate between the many possible referents usually available for the words they hear. This could have been also the case for early humans, on the basis of a biological substrate to associate sounds and information in other modalities. However, where modern children benefit from their parents' own stable communication system, our distant ancestors had to develop signs from scratch. In the same way we don't understand yet what takes place in the infant's mind as they grasp the symbolic and referential nature of words, it is difficult to conceive of how this process unfolded across different generations of ancient humans, in a gradual manner rather than in a sudden 'aha' moment.

One can argue that only when a basis was firmly in place could our ancestors move toward more arbitrary signs. The non-arbitrary signs found in today's languages would therefore be reminders of much earlier stages, and their presence could still be explained by the help they provide, no longer for the emergence of novel communicative signs, but for their acquisition during childhood. At the same time, the need to distinguish between close concepts as language developed in the past may explain the evolutionary advantage arbitrary signs gradually gained, and why sound symbolism is not more prevalent in today's languages (Imai and Kita, 2014; Monaghan et al., 2012). Such proposals echo Ahlner and Zlatev (2010)'s broader evolutionary explanation for sound symbolism. According to them, 'iconicity is a key factor in the emergence of new expressions', but loses its functionality as conventionalization of these expressions occurs in the speakers' community. The two authors mention how writing, language standardization and language contact can in some cases further contribute to this evolution.

1.2.4. Sound symbolism and the multimodal origins of language

Some scholars have long defended a scenario of the origins of language based on manual rather than vocal communication (Corballis, 2003). This is in line with ape's greater flexibility in the use of their body compared to their vocal tract (Arbib et al., 2008), and experimental studies suggest that motivated signs are more easily produced in the gestural domain than in the vocalic one (Fay et al., 2013). How did our ancestors then move from iconic gestures to symbolic vocalizations? Did it imply first the development of symbolic gestures? Were there no iconic vocalizations at first, and why? Perlman et al. (2015) consider different scenarios: on the one hand, scenarios in which vocalizations are intrinsically meant to function symbolically rather than as iconic signs, and would have emerged from an intrinsically iconic gestural communication system; on the other hand, scenarios in which iconic communicative signs developed both in gestural and vocal communication, depending on what was easiest given the referent to be expressed, e.g. gestures for actions and spatial relationships, and vocalizations for objects or events identifiable by distinctive sounds.

Previous proposals regarding the emergence and properties of motivated communication signs apply equally well to gestural and vocal signs. What changes is the nature of the cross-modal relationships, and the expression with non-arbitrary signs of some objects, events or ontological properties may be easier in one modality than in the other.

1.3. Early sound symbolic associations

To what did early humans relate the phonetic or gestural features of their communication signs? On the one side, sensations and emotions were surely as important as they are today in modern humans, and could form part of what was expressed. On the other hand, in a world yet lacking sophisticated technology, farming etc., animals, and especially the dichotomy between prey and predator, were a significant part of life. Their perceived nature (fish, birds, mammals etc.) and how dangerous/harmless, attractive/repulsive, big/small they were are therefore plausible candidates for early sound symbolic associations. Such cognitive sensitivity to animals seems well alive in today's humans, as suggested by various aspects of folk psychology regarding the essence and properties of animals (Boyer et al., 2000; Gelman, 2003). For these reasons of evolutionary and cognitive relevance, we thought relevant and more 'ecological' to experimentally investigate representations of animals rather than more abstract stimuli such as angular or round shapes.

We investigated the relationships between phonetic features of linguistic units and animal size, dangerousness, attractiveness and biological class in French. Our starting hypothesis was that a number of sound symbolic associations could be highlighted, much as what was shown by Berlin in Huambisa.

2. Experiment

2.1. Overview & rationale

The experiment consisted in an association task aimed at shedding light on sound symbolic associations between phonetic features and animal size, dangerousness, attractiveness and biological class. We presented subjects with images of animals and oral pseudo-words, in a 1×2 design: the pseudo-word was presented first, then the two images simultaneously. Subjects had to choose which association seemed intuitively more natural to them. The experiment was reviewed and accepted by the *Comité de Protection des Personnes Sud-Est IV* (the relevant local ethical committee, following current French procedures).

Mixing various pairs of images with different ontological contrasts (size, attractiveness etc.) made guessing what was going on more difficult for subjects. The task was therefore more implicit than the *bouba-kiki* association task, in the sense that features to be associated were partly hidden from the participants.

2.2. Hypotheses

We designed stimuli to provide a testbed for various hypotheses regarding sound symbolism. Investigating specific words or segment interactions would have required a large number of subjects, and we focused instead on simpler hypotheses based on phonetic features or segments. Nine were chosen according to sound symbolic associations proposed in the literature, and are summarized in Table 1.

Table 1
Hypotheses tested in each ontological category.

Category	Hypothesis		Reference
Size	[i] – small	[a] – big	Sapir (1929), etc.
	Voiced consonants – big	Unvoiced consonants – small	Ozturk et al. (2013)
Dangerousness	Back vowels – dangerous	Front vowels – harmless	Fónagy (1983, 1961)
	Plosives – dangerous	Sonorants – harmless	Fónagy, (1983, 1961)
	Back consonants – dangerous	Front consonants – harmless	Fónagy, (1983, 1961)
Repulsion	Plosives – repulsive	Sonorants – attractive	Fónagy, (1983, 1961)
	[a] – repulsive	[i] – attractive	Fónagy, (1983, 1961)
Class	Fricatives – fish	Plosives – birds	Berlin (1994)
	[a] or [u] – fish	[i] – birds	Berlin (1994)

2.3. Material

2.3.1. Images of animals

The depicted animals were selected from those most frequently mentioned by French subjects in a fluency task, as available from the BASETY database (Léger et al., 2008). Out of a total of 379 animals, we initially selected the 200 most quoted to ensure that our stimuli would be well-known by our French participants.

216 pictures were selected from the website 123rf.com to match 166 of these animals – for some animals, several different images were selected. A few pictures had to be modified so that all eventually consisted of a single animal against a white background. Several studies and cross-linguistic surveys have shed light on possible associations between sounds and visual properties such as brightness (Nuckolls, 1999; Wicker, 1968). The hue or lightness of our animals' skins, furs or feathers could therefore have had an influence on the task, and pictures were therefore turned into shades of gray with equal lightness.

The images were evaluated through an online survey – those surveyed did not participate in later experiments. Five dimensions were assessed with Likert scales by the following numbers of subjects (some subjects participated in several evaluations):

- Size: from 1 (smallest) to 9 (biggest) – 32 subjects (15 males).
- Dangerousness: from 1 (most harmless) to 5 (most dangerous) – 32 subjects (16 males).
- Repulsion: from 1 (most attractive) to 5 (most repulsive) – 31 subjects (16 males).

2.3.2. Pairs of images of animals

Pairs of images were created in four different categories:

- Size: 16 pairs.
- Dangerousness: 16 pairs.
- Repulsion: 11 pairs.
- Class (Fish vs. birds): 16 pairs.

In pairs addressing size, dangerousness and repulsion, the two animals always belonged to the same biological class (mammals, fish, insects, birds…). For each pair, images were contrasted on the target feature, and balanced on the others, according to the outputs of the previous online survey. For example, in pairs testing size, the size difference was maximized, while contrasts in terms of dangerousness and repulsion were minimized. In pairs opposing birds to fish, it proved somewhat difficult to balance repulsion. Table 2 summarizes the output of the process.

Table 2
Average difference of assessments of paired images for the experimental target categories and the assessed dimensions.

	Repulsion (1–5)	Dangerousness (1–5)	Size (1–9)
Repulsion	**1.93**	.53	.59
Dangerousness	.32	**2.41**	.53
Size	.44	.49	**4.14**
Class	.75	.29	.46

Each row corresponds to an experimental target category, each column to an assessed dimension. The numbers in bold highlight maximized differences, others correspond to minimized differences.

59 target pairs were therefore constituted to test sound symbolic association. To further hide from subjects the target contrasts of the experiment, 45 fillers made of randomly associated pictures were added to the experiment, to reach a total of 104 trials. 4 trial pairs were added for training purpose.

Riou et al. (2011) have shown that perceiving an object reactivates information of its dimensions stored in memory. There are thus interactions between perceptual and memory mechanisms. Moreover, Paivio (1975) has shown that judgment tasks

on animal size require less time when the relationship between physical and memory-stored sizes are congruent rather than incongruent.

In order to avoid a potential cognitive correction of size difference inside pairs that contrasted this feature, we chose to approximate participants' mental representation of animal size – as indicated by our initial survey – with the onscreen size of our animal pictures. Onscreen sizes of animals were thus power-law transformations of size judgments obtained with previous evaluations. As a result, for example, a bee has a screen area of 9,336 pixels, while a roe has an area of 48,930 pixels. This ensured that animals of different sizes presented in pairs appeared with a simulated size difference likely reflecting participants' mental representations.

2.3.3. Pseudo-words

64 pseudo-words were built with a VCVC (vowel-consonant-vowel-consonant) structure where both a single vowel and a single consonant were reduplicated. They consisted in one of the 4 vowels [i], [a], [u] and [y], and one of 16 possible consonants. All these phonemes belong to the French phonological inventory, and the VCVC structure does not violate phonotactic constraints of this language. The goal was to achieve all possible combinations between these vowels and consonants. Reduplication of the vowel and of the consonant was chosen to avoid complex interactions between multiple phonemes.

The VCVC structure is not very frequent in French. According to the Lexique 3.81 database (New et al., 2001), there are 1532 VCVC words in the 142,694 words of the database, compared to 8759 CVCV words. Three words were found when we looked at all possible VCVC forms with our reduplicated 4 vowels and 16 consonants: two loanwords from Arabic – 'hallal' and 'hammam' – and an inflected verbal form – 'hulule'. We replaced these three words with three pseudo-words using [ɲ] as their first consonant: /aɲal/, /aɲam/ and /yɲyl/. Finally, 'Hittite' existed in French, but was kept as a pseudo-word given its very low frequency of occurrence and the fact that most speakers do not know its meaning (Table 3).

The 4 vowels allowed comparisons on three partially redundant phonetic features: aperture, anteriority and rounding. The 16 consonants allowed investigations of place, manner and voicing.

Following Stevens and House (1963), the 6 different consonantal places of articulation were grouped into 3 categories: front (bilabial and labiodental, 20 pseudo-words), medium (i.e. coronals, alveolar and postalveolar, 28 pseudo-words) and back (velar and uvular, 12 pseudo-words). The three pseudo-words /aɲal/, /aɲam/ and /yɲyl/ were excluded from this classification.

The 5 different manners of articulations were grouped into 3 categories: plosives (24 pseudo-words, either voiced or voiceless), fricatives (24 pseudo-words, voiced or voiceless) and sonorants (nasal and lateral, 12 pseudo-words, all voiced).

44 non-target pseudo-words with a VCVC structure but varying and non-reduplicated consonants and vowels (e.g. /unɛg/, /yzak/ etc.) were added. 40 were used as fillers, 4 for the training trials.

Pseudo-words were presented in the auditory modality in order to prevent potential orthographic bias. They were recorded with a ZOOM H4 digital recorder by a 28-year-old male native speaker of French, from the center of France, unaware of the experiment and of its hypotheses. Recordings were segmented with Praat (Boersma, 2001), and normalized for amplitude. Background noise was filtered. The mean pitch was about 122 Hz.

2.3.4. Associations of pseudo-words and pairs of images

Target pseudo-words and pairs of images were associated randomly for each participant, and presented in a pseudo-random fashion – some constraints were added to prevent categories of judgments to occur too many times successively. Images of each pair were randomly shown on the left or the right side of the screen.

Table 3
Pseudo-words built for the experiment, voiced ones in bold.

Mode	Place	i	a	u	y
Plosive	Bilabial	ipip	apap	upup	ypyp
		ibib	**abab**	**ubub**	**ybyb**
	Alveolar	itit	atat	utut	ytyt
		idid	**adad**	**udud**	**ydyd**
	Velar	ikik	akak	ukuk	ykyk
		igig	**agag**	**ugug**	**ygyg**
Fricative	Labiodental	ifif	afaf	Ufuf	yfyf
		iviv	**avav**	**uvuv**	**yvyv**
	Alveolar	isis	asas	usus	ysys
		iziz	**azaz**	**uzuz**	**yzyz**
	Postalveolar	iʃiʃ	aʃaʃ	uʃuʃ	yʃyʃ
		iʒiʒ	**aʒaʒ**	**uʒuʒ**	**yʒyʒ**
	Uvular	**iʁiʁ**	**aʁaʁ**	**uʁuʁ**	**yʁyʁ**
Sonorant	Bilabial	**imim**		**umum**	**ymym**
	Alveolar	**inin**	**anan**	**unun**	**ynyn**
		ilil		**ulul**	
	–		**aɲal**		**yɲyl**
			aɲam		

2.4. Participants

We collected answers from 53 monolingual native French speakers (22 males and 31 females) aged from 16 to 34 years (mean = 21.36 years). None of them had language, visual or auditory impairment. The large majority of subjects were undergraduate students in psychology, cognitive science or linguistics, who participated for partial course credit. Other subjects were graduate students or young professionals with a university background.

2.5. Procedure

The software *OpenSesame* was used for stimuli presentation and data recording.

In a quiet experimental room, after completing a consent form and a laterality test, participants were seated in front of a 17" screen in an experimental booth, and fitted with an audio headset. Onscreen instructions told them that the goal was to choose the best association between a word – of an unknown language – presented orally and one of two animals presented visually. They were also instructed to answer as fast as possible using two keyboard keys indicated with color stickers: the left key for the animal on the left and the right key for the animal on the right.

Four training trials immediately followed the instructions, and a comfortable listening volume was ensured. Target stimuli and fillers were then presented. For each trial, a fixation dot was presented in the center of the screen. After a varying duration between 1200 and 1600 ms, the pseudo-word was orally presented and the fixation dot disappeared 300 ms after the end of the audio recording. Then, two images were simultaneously presented on both sides of the screen until one of the two relevant keys was pressed. A 2500 ms upper threshold was defined in case subjects failed to answer. The images were immediately followed by a mask for 80 ms, in order to erase the visual memory of the previous event. This mask was made of bars and crescents of differing orientations and sizes, in order to cover low and high visual frequencies. The next trial followed automatically.

2.6. Data analytic procedure

Since the associations collected during our experiment were grouped by subjects, pairs of images and pseudo-words, and therefore not independent from each other, we could not rely on contingency tables and chi^2 tests. Instead, we relied on mixed logistic regressions as follows:

- The dependent variable was the binary choice between one of two images.
- The fixed effect was the target phonetic element or class involved in the tested hypothesis, e.g. vowel anteriority to assess the choice of dangerous or harmless animals.
- Subjects and pairs of images were included in the model as random effects.

Given the lack of hypotheses in the literature regarding possible interactions between the various fixed effects in a category, we decided to set these interactions aside and consider fixed effects separately. Pseudo-words were not included as random effects, since this meant introducing interactions between vowels and consonants in the model, although not with fixed effects, and partly mask the main effects we were primarily interested in.

Statistical analyses were conducted with R (R Development Core Team, 2008), more precisely the *lmer* and *glmer* function of the lme4 package (Bates et al., 2015). For each model, in order to assess the significance of the fixed effect, we considered the values returned by the glmer function with sum contrasts (equivalent in this situation to the p-value of a Wald chi^2 test given by a type-III analysis of variance). Since it would only have changed results marginally at the cost of additional complexity, reported results do not take into account the fact that our hypotheses were all directional (e.g. 'big animals are associated with [a] and small animals with [i]', rather than 'there is an interaction between the size of an animal and the vowels found in its name').

To address the issue of false positives (type I errors) likely to occur with multiple tests (Hochberg and Tamhane, 1987), we chose to control for the *family-wise error rate* (FWER) with the Holm–Bonferroni method (Holm, 1979).

2.7. Analysis and results

2.7.1. Influence of laterality

A linear regression was used to check a potential impact of participants' laterality (as obtained from the laterality test) on responses, since they involved both hands. After one subject was removed because he appeared to have given up the task half-done, laterality was not a significant predictor and we decided not to include it in the main regression models.

2.7.2. Absence of response and reaction times

We obtained 5408 trials for our 52 subjects – 3068 targets and 2340 fillers. 104 entries – 63 for target trials and 51 for fillers – corresponded to an absence of response within the 2500 ms time frame.

On the basis of 3003 targets and 2291 fillers, we removed extreme reaction times (±3 times the standard deviation): only 2 fillers were removed, for which the reaction time was above 2383 ms. Average response time was then equal to 1185 ms.

Once fillers were removed, non-paired two-sided t-tests were run between the reaction times of the pairs of our different categories of images. No significant difference was observed between any two categories.

2.7.3. Logistic regression models

We tested our 9 hypotheses with as many regression models, and did not find statistical support for any of them. The absence of significant results was not due to the application of the Holm–Bonferroni method, except for the preference of [a] and [u] for fish, and [i] for birds, which was significant before the family-wise error rate was controlled for. Results are summarized in Table 4.

Table 4
Results of the significance tests for the various hypotheses.

Category	Hypothesis		Est.	Std. Err.	z Value	Pr (>|z|)
Size	[i] – small	[a] – big	.023#	.102	.228	.820
	Voiced C – big	Unvoiced C – small	.041#	.076	.546	.585
Dangerous-ness	Back V – dangerous	Front V – harmless	−.121#	.087	−1.395	.163
	Plosives – dangerous	Sonorants – harmless	.037#	.098	.384	.701
	Back C – dangerous	Front C – harmless	.048#	.104	.463	.643
Repulsion	Plosives – repulsive	Sonorants – attractive	−.036#	.120	−.298	.765
	[a] – repulsive	[i] – attractive	.103	.128	.804	.421
Class	Fricatives – fish	Plosives – birds	−.072#	.082	−.873	.382
	[a] or [u] – fish	[i] – birds	.227#	.094	2.410	.016

Values reported correspond to the fixed effect of each logistic regression model (values for intercepts not reported). No result was significant after application of the Holm–Bonferroni method. C stands for consonants, V for vowels. # Indicates that the content of the contingency table and the sign of the estimate corresponded to the directionality of the hypothesis.

Overall, we thus globally did not find evidence of sound symbolic associations with our experiment.

3. Discussion

3.1. Hypotheses regarding the previous results

We thought of several reasons why our experiment failed to reveal sound symbolic associations.

A first reason could be that we did not have enough subjects to test our hypotheses. However, given that our hypotheses were mostly about features and sounds, and not about complex interactions, our models were applied to rather large numbers of data. In the case of voicing for big or small animals, there were 294 trials with unvoiced consonants, and 519 trials with voiced consonants. In the case of [a] vs. [i] with big or small animals, we had 745 trials for [a], and 755 for [i] (on average more than 14 instances per subject). If effects had been strong, they would have shown in the logistic regressions despite possible interactions. Weaker effects could require more subjects and/or more trials per subjects, but the literature suggests rather strong associations, if one thinks for example of how 95% – not 65% or 60% – of subjects associate a rounded shape with 'bouba', and a spiky shape with 'kiki'.

Other reasons for the lack of significant associations could lie in the design of the experiment. First, the 2 × 1 design differed from association tasks like the *bouba-kiki* experiment, where two non-linguistic stimuli and two linguistic stimuli are presented together – i.e. a 2 × 2 design. In the latter case, subjects can explicitly compare two linguistic stimuli, and therefore focus on a contrast, even if they cannot describe it in terms of rounding, voicing or place of articulation. Being able to contrast sounds could be key to the mental processing of sound symbolic associations in such a task, although the results of phonetic judgment tasks suggest rather otherwise. In any case, a 2 × 1 design is more 'implicit', which may hinder cognitive multimodal associations. Also, without explicit sound contrast, participants possibly developed idiosyncratic strategies distinct from the sound symbolic associations they would have produced in another setting. A combination of such simpler matching strategies could have produced patterns of responses very distinct from what was expected.

Subjects also had limited time to make their choice. They were told in the instructions to be as quick as possible, and reaching the 2500 ms threshold could act as a reminder of this temporal constraint (there was however no time counter displayed). The VCVC structure is not very frequent in French, and could therefore have required some time to process. If sound symbolic associations appear late in the cognitive processing of the linguistic and non-linguistic inputs, participants could have lacked time to provide the expected responses. This however seems to run counter to the fact that even very young children, likely without elaborate strategies, favor sound symbolic associations. Also, participants to the *bouba-kiki* association task usually have strong intuitions about the 'correct' pairs, without necessarily providing a rationale for their choice. Additionally, the average response time was close to 1200 ms, which does not seem fast for a decision task that only requires pressing one key or another.

Finally, the last issue could be the nature of the non-linguistic stimuli: despite our efforts at maximizing and minimizing contrasts of size, dangerousness or repulsion according to our categories, images of animals are complex visual stimuli, which

activate very rich representations. Whether subjects actually identified and focused on the targeted contrasts can therefore be questioned.

Beyond the experimental design, it seemed curious that French speakers did not produce sound symbolic associations, given how common they are cross-culturally, and given previous experimental evidence with these speakers for the *bouba-kiki* task (Nobile, 2015).

We conducted two additional investigations to delve into these issues. We first ran a second experiment with a more explicit association task. We then analyzed animal names in French in search of sound symbolic associations.

3.2. A more explicit association task

We conducted a second experiment to further investigate the presence or absence of sound symbolic associations in French speakers.

3.2.1. Overview & rationale

For the sake of comparison, we decided to keep the 2 × 1 experimental design and the VCVC pseudo-words used in the first experiment. However, rather than using pictures of animals, we chose to simply speak of 'big animals' vs. 'small animals', 'dangerous animals' vs. 'harmless animals' etc. By doing so, we explicitly presented the ontological domain of investigation to our subjects, who had to decide whether a given pseudo-word was rather depicting a small or a big animal, an attractive or a repulsive one etc.

Rather than an experiment in front of a computer, we distributed questionnaires with nine questions to answer. Pseudo-words were therefore written and not presented auditorily. We assumed that reading a word activated the related phonological representation in the subject's mind.

3.2.2. Material

36 pseudo-words of the first experiment were used:

- 8 for judgments of size: 2 vowels ([i] and [a]) and voicing were contrasted, with 4 pseudo-words made of plosive consonants ([p] vs. [b]), and 4 made of fricatives ([ʃ] vs. [ʒ]).
- 12 for judgments of dangerousness: 2 vowels ([u] and [i]) were contrasted within 10 pseudo-words and [y] and [a] were used each in a single word. Voicing ([b] vs. [p]; [g] vs. [k]) and place ([b] vs. [g]; [p] vs. [k]) were contrasted in 8 pseudo-words. 4 other pseudo-words containing sonorants ([l], [m] and [n]) were added to be compared on manner with the 8 first.
- 8 for judgments of repulsion: half of the pseudo-words contained the vowel [i], the other half the vowel [a]. Orthogonally, half were composed of sonorants ([l], [m] and [n]), and half of the two consonants [k] and [ʁ].
- 8 for judgments of biological class (fish or bird): half of the pseudo-words contained plosives ([p] and [t]), the other half fricatives ([s] and [f]). Orthogonally, half contained the vowel [a] and half the vowel [i].

Since pseudo-words were to be presented on paper, we had to devise acceptable written forms for them. Forms like 'ikik' would likely have been judged as weird by subjects, since very few French syllables if not none except in borrowed words end with a [k]. We therefore adopted 'pseudo-written forms', i.e. forms that appeared to respect what written words usually look like in French.

These 36 pseudo-words were grouped into 4 different lists to prepare as many questionnaires. Each questionnaire contained 2 judgments for size, 3 for dangerousness, 2 for repulsion and 2 for class (Table 5).

The four different questionnaires were built so that the 9 hypotheses of the first experiment could be tested. Keeping the number of questions low for each of the participants additionally ensured that they did not form meta-strategies to give a response to a specific contrast. More precisely, a subject could only see one side of a phonetic contrast, whether it was [i] vs. [a], voiced vs. unvoiced consonants etc. This prevented them from making a second choice taking the first one into account, at least inside a given category (Table 6).

The pseudo-words were presented in a pseudo-random order to avoid several judgments of a given ontological category to appear successively. It aimed at minimizing possible strategies such as 'I have already chosen something for a bird, I assign this other word to a fish'.

Table 5
The 4 lists of written pseudo-words used in the questionnaires of the second experiment.

	Size		Dangerousness			Class		Repulsion	
V1	ipipe	ajage	ouboube	iguigue	ananne	atate	ississe	hacac	ilile
V2	ibibe	achache	ougougue	iquique	ilile	ipipe	affafe	arrare	innine
V3	apape	ijige	oupoupe	ibibe	oumoume	itite	assasse	iquique	alame
V4	ababe	ichiche	oucouc	ipipe	ununne	apape	iffife	irrire	annane

Table 6
Corresponding phonological forms for the written pseudo-words used in the four versions of the second experiment.

	Size		Dangerousness			Class		Repulsion	
V1	ipip	aʒaʒ	ubub	igig	anan	atat	isis	akak	ilil
V2	ibib	aʃaʃ	ugug	ikik	ilil	ipip	afaf	aʁaʁ	inin
V3	apap	lʒiʒ	upup	ibib	umum	itit	asas	ikik	alam
V4	abab	iʃiʃ	ukuk	ipip	ynyn	apap	ifif	iʁiʁ	anan

3.2.3. Participants

A total of 132 students (20 males and 112 females), aged 17–49 years (mean = 21.7 years), participated in the protocol. All were French native speakers and had not participated in our earlier surveys and experiments.

3.2.4. Procedure

The questionnaires were filled at the end of a class. The 4 different versions were distributed randomly. Subjects were asked to provide help with understanding the words of a newly discovered language, by deciding whether proposed words better fitted one proposition or the other, e.g. a big or a small animal.

3.2.5. Hypotheses and data analytic procedure

We tested the same hypotheses as previously, although with fewer conditions than in the first experiment. We again relied on logistic regression models with a single fixed effect to see how subjects' choices were predicted by phonetic features of the pseudo-words. The Holm–Bonferroni method was applied to determine which associations were significant.

3.2.6. Results

Results of the assessment of our 9 hypotheses are reported in Table 7.

We found seven statistically significant sound symbolic associations after application of the Holm–Bonferroni method.

Table 7
Results of the significance tests for the various hypotheses tested with questionnaires.

Category	Hypothesis		Est.	Std. Err.	z Value	Pr (>|z|)
Size	[i] – small	[a] – big	.984#	.139	7.100	**1.2e−12**
	Voiced C – big	Unvoiced C – small	.306#	.125	2.455	**.014**
Dangerous-ness	Back V – dangerous	Front V – harmless	.051#	.110	.463	.643
	Plosives – dangerous	Sonorants – harmless	.243#	.116	2.102	.035
	Back C – dangerous	Front C – harmless	.499#	.123	4.048	**5.2e−5**
Repulsion	Plosives – repulsive	Sonorants – attractive	−.468#	.158	−2.962	**.003**
	[a] – repulsive	[i] – attractive	−.490#	.128	−3.815	**1.4e−4**
Class	Fricatives – fish	Plosives – birds	.588#	.136	4.309	**1.6e−5**
	[a] or [u] – fish	[i] – birds	.372#	.132	2.820	**.005**

Values reported correspond to the fixed effect of each logistic regression model (values for intercepts not reported). p-Values in bold indicate significant results after application of the Holm–Bonferroni method. C stands for consonants, V for vowels. # Indicates that the content of the contingency table and the sign of the estimate corresponded to the directionality of the hypothesis.

Therefore, this new enquiry provided different results from the first experiment. The null hypothesis could be rejected in 7 out of 9 of our tests, and we found no results in contradiction with the directionality of our hypotheses.

It has been assumed that consonants have more influence than vowels when it comes to sound symbolic associations (Nielsen and Rendall, 2011, 2013). This could have resulted in more associations involving consonants than vowels. This is not really the case, since 3 of the 7 significant associations involve vowels. This may be due to the VCVC pattern we chose for our pseudo-words. Conversely, it is perhaps the use of CVC or CVCV patterns in other studies which explains best why consonants seem more influent than vowels in sound symbolic associations. We can also stress that both place and mode of consonants occur in these associations.

3.3. Exploring animal names in French

Studying Huambisa, Berlin (1994) found some specific proportions of phonemes in animal names related to the animal being a bird or a fish, or being big or small. He focused in particular on 175 bird names and 85 fish names. This prompted us to pay attention to animal names in French.

We considered the 166 animals extracted from the BASETY database that we used in our first experiment (62 mammals, 39 birds, 27 fishes, 28 arthropods, 7 reptiles, 2 worms and a slug). In relation with the judgments on size, dangerousness and repulsion, and with the biological class, we coded the frequency of occurrence of the segments and features related to our

nine hypotheses of sound symbolic associations. Words were much more varied in terms of structure and content than the pseudo-words tailored for our experiments, and we therefore had to adapt these hypotheses, e.g. consider the relationship between size and the degree of aperture of vowels, rather than the relationship between size and the number of [i] and [a] in the words. For example, to test the association between voicing and size, we computed for each word the ratio of the number of voiced consonants to the total number of consonants: in 'chameau' ([ʃamo] – camel), one consonant was voiced, and one unvoiced, thus a ratio of .5. For vowel anteriority or aperture, and for consonant place of articulation, we devised a score taking the different vowels or consonants of the word into account. As for vowel aperture for example, the score was maximal (100%) if all vowels in the name were high (close) (e.g. hibou/[ibu] – owl), minimal (0%) if they were all low (open) (e.g. canard/[kanaʁ] – duck), and intermediate in other configurations.

For each hypothesis, we ran a logistic regression to see whether how big/small, harmless/dangerous etc. the animal was according to subjective judgments predicted the expected phonetic distributions in the words. Results are summarized in Table 8.

Table 8
Results of the significance tests for various hypotheses applied to French animal names.

Category	Hypothesis		Est.	Std. Err.	z Value	Pr (>|z|)
Size	High V – small	Low V – big	−.009	.029	−.327	.744
	Voiced C – big	Unvoiced C – small	.006	.037	.160	.873
Danger	Back V – dangerous	Front V – harmless	−.037	.064	−.581	.561
	Plosives – dangerous	Sonorants – harmless	.035	.091	.385	.700
	Back C – dangerous	Front C – harmless	−.061	.058	−1.053	.292
Repulsion	Plosives – repulsive	Sonorants – attractive	−.009	.103	−.092	.927
	Low V – repulsive	High V – attractive	.058	.053	1.104	.270
Class	Fricatives – fish	Plosives – birds	−.032	.150	−.212	.832
	Back V – fish	Front V – birds	.044	.105	.420	.674

Values reported correspond to the fixed effect of each logistic regression model (values for intercepts not reported). No statistically significant association was obtained. C stands for consonants, V for vowels.

None of the regression models turned out to give significant results. A certain degree of sound symbolism that appears in the lexicon of Huambisa and of other languages therefore does not in the French lexicon of animal names.

3.4. Factoring in the various approaches

The second experiment showed that French speakers actually produce sound symbolic associations for various attributes of animals: size, biological class, but also aspects of their behavior and/or appearance.

3.4.1. Reinterpreting the results of the first experiment

The second experiment allowed to prune some of the possible explanations for the lack of significant results in the first experiment.

First, the 2 × 1 design alone was not the cause of missing sound symbolic associations; indeed, the last experiment showed that the lack of an explicit phonetic contrast did not prevent subjects to establish cross-modal associations. Second, the disyllabic VCVC structure was not the sole determining factor either, since it was 'successfully' used in the second experiment, despite not being a very frequent word structure in French. Additionally, the absence of expected sound symbolic associations in the French lexicon of animal names did not hinder French speaking participants to make sound symbolic associations.

What is left as a plausible cause of the absence of sound symbolic associations in the first experiment is the use of pictures that carried too much information at the perceptual and semantic levels. A zebra was for example a four-leg mammal, lived in specific environments and had a specific coat, could be dangerous yet attractive etc. The pictures were perhaps too hard to harness, even with quantitative contrast and balancing across categories of ontological features. Alternatively, a combination of several of the previous factors, each adding to the difficulty of the task, could explain the absence of detected associations.

3.4.2. Cognitive vs. lexicalized processes of sound symbolism

As previously mentioned, some explanations can be formulated why sound symbolic associations are rare in today's languages, despite the role they likely played in the initial steps of our communication system. The absence of sound symbolic associations in French animal names suggests that this part of the lexicon in French has indeed lost or 'got rid' of such associations. This however does not mean that French speakers do not produce these associations when prompted to do so in an appropriate task. What exists at a cognitive level therefore does not necessarily appears in the lexicon.

The question is then why some cognitive basements of sound symbolism get reflected in the lexicon of some languages, and not in the lexicon of others. To elaborate on the contrasted situations of Huambisa and French, sound symbolism may be more developed in languages where it provides a greater advantage. For animals, sound symbolic names may be more helpful in a population of hunters than in a population where hunting activities have not been at the center of everyday life for the majority of people and for a long time.

3.4.3. Evolutionary processes shaping lexical sound symbolism

How did past languages evolve from dominant sound symbolic lexical associations to dominant arbitrariness? How did this happen while sound symbolic capacities were preserved at the cognitive level? For some scholars, arbitrariness in speech is a consequence of the vocal-auditory nature of the channel (Galantucci et al., 2012). Does this mean that a process actively erases sound symbolic associations from the lexicon?

Roberts et al. (2015) consider two competing pressures for a communication system: a pressure for referential efficiency, which can be efficiently satisfied by iconic signs, and a pressure for transmission efficiency, which benefits from combinatoriality – the productive combination of meaningless units into meaningful forms – and suffers from iconicity. Iconicity does not strictly equal sound symbolism, but the previous tradeoff seems to extend to the latter. In early forms of communication, sound symbolism may thus have hindered the emergence of combinatoriality, a central mechanism to support the multiplication of expressed concepts, events and objects. We may conceive of a drift, which gradually transformed primary sound symbolic signs to give way to combinatoriality, but not to the extent that sound symbolism disappeared completely. In addition, if forces favoring the lexical expression of sound symbolism only weakly bear on the evolution of the lexicon, they can be overcome by stronger constraints. If these stronger constraints are not universal, but only appear in specific sociocultural contexts, this could explain why some languages are more prone than others to exhibit sound symbolic associations in their lexicon. Such constraints may be limited to specific domains such as the expression of movements, or of specific entities like animals.

This teleological view means that languages only lose inherited sound symbolic associations, but never create new ones. Rather than only focusing on mechanisms building up arbitrariness, it makes sense to also investigate the existence of processes favoring the emergence of new sound symbolic associations in a (mostly arbitrary) communication system. If such mechanisms do exist, they would point at the existence of a dynamic equilibrium between constraints for or against sound symbolism in the lexicon, rather than an endless decay. Given how much time language(s) had to evolve until today, this may be a more plausible scenario.

3.4.4. Sound symbolism of animal biological classes and emotions

In the second experiment, we found a partial replication of Berlin's results in Huambisa: our subjects significantly associated birds to fricatives and [i] and fish to plosives and [a] or [u]. In addition to English-speaking subjects correctly guessing the class of Huambisa animal names (Berlin, 1994), this suggests that the fricative–plosive and [i]–[a/u] distinctions are present in a wide range of languages.

Fónagy's sound symbolic associations with emotions or aspects such as dangerousness or attractiveness have not been extensively studied with experimental approaches. Our second experiment is a first step in this direction, and reinforces the idea that such associations do exist at the cognitive level, also they may not always surface in the vocabulary.

3.4.5. Sound symbolism of individual segments or of word acoustic shapes?

We reported the tests of nine hypotheses, without describing in details patterns of interactions of phonetic features. Such interactions however do exist, but are not easily interpretable.

Phonemes are the building blocks of words in a language. They are restricted by articulatory constraints, and as such cannot fully replicate what is perceived or stored. Combining phonemes gives rise to co-articulatory effects, and to some extent refine the phonetic options for sound symbolism. The possibility exists that the phonetic basic units of sound symbolic associations are not confined within segmental boundaries, and encompass several segments or parts of different segments.

Along these lines, if one thinks of the size-code hypothesis of emotional speech, which relates emotions such as anger or happiness to the fundamental frequency of the voice (Chuenwattanapranithi et al., 2008), it makes sense to go beyond segments and analyze the evolution of F0 during the production of a word. Vowels are known to have their own intrinsic frequencies (Whalen and Levitt, 1995), but interactions with different consonants can modify theses frequencies. Voiced plosives and voiceless plosives for example exert their influence in different directions (Stevens and House, 1963). We think that the fundamental frequency, but perhaps also the characteristics of the spectral acoustic envelope of words, deserve more attention. Sound symbolic associations established at the level of words, or of chunks of segments, also resonate with early linguistic forms that might have been holistic and not segmented into segmental components. If early signs were non-compositional, it makes sense to think of early sound symbolic associations at the word level rather than at the segmental one.

4. Conclusion & perspectives

Sound symbolism requires further studies to be better understood. This is true both for how it takes place at the cognitive level and in the lexicon of today's languages, but also to understand which processes shaped languages to their current state, given individual cognitive processes but also other, especially social, constraints on communication.

Among the questions still lacking a consensual answer: at which level do the cognitive processes underlying sound symbolism take place? What are the phonetic dimensions involved in these associations, and how far do they range in terms of segments, chunks of segments or words? Are variations of the fundamental frequencies or some parameters of the spectral envelope of a sound signal the relevant cues that are analyzed by speakers? How do phonetic features associated with consonants or vowels interact with the position in a lexical unit? Which associations are universal, and which are culture- or language-specific?

Besides complementing the experiments presented in this article to address the previous issues, we are also considering the dimensions of valency and arousal for pictures of animals or other visual stimuli, and the role they play in sound symbolic associations. Preliminary investigations seem to suggest that subjects relate how arousing animal pictures are to variations of the fundamental frequency of the voice. This suggests new directions for the study of sound symbolism.

Acknowledgments

The authors are grateful to the *LABEX ASLAN* (ANR-10-LABX-0081) of *Université de Lyon* for its financial support within the program *Investissements d'Avenir* (ANR-11-IDEX-0007) of the French government operated by the *National Research Agency* (ANR). The authors also thank Slawomir Wacewicz, Nathalie Bedoin, Vincent Arnaud, François Pellegrino, Bruno Galantucci, Dan Dediu and two anonymous reviewers for their very useful comments and suggestions.

References

Ahlner, F., Zlatev, J., 2010. Cross-modal iconicity: a cognitive semiotic approach to sound symbolism. Sign Syst. Stud. 38, 298–348.
Arbib, M.A., 2005. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav. Brain Sci. 28, 105–124. http://dx.doi.org/10.1017/S0140525X05000038.
Arbib, M.A., Liebal, K., Pika, S., 2008. Primate vocalization, gesture, and the evolution of human language. Curr. Anthropol. 49, 1053–1076. http://dx.doi.org/10.1086/593015.
Bates, D., Maechler, M., Bolker, B.M., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. http://dx.doi.org/10.18637/jss.v067.i01.
Berlin, B., 1994. Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In: Hinton, L., Nichols, J., Ohala, J. (Eds.), Sound Symbolism. Cambridge University Press, pp. 76–93.
Boersma, P., 2001. PRAAT, a system for doing phonetics by computer. Glot Int. 5, 341–345.
Botha, R., 2008. Prehistoric shell beads as a window on language evolution. Lang. Commun. 28, 197–212. http://dx.doi.org/10.1016/j.langcom.2007.05.002.
Bouchard, D., 2013a. Arbitrary signs and the emergence of language. In: Lefebvre, C., Comrie, B., Cohen, H. (Eds.), New Perspectives on the Origins of Language. Studies in Language Companion Series 144. John Benjamins, pp. 407–440.
Bouchard, D., 2013b. The Nature and Origin of language. Oxford University Press.
Bouzouggar, A., Barton, N., Vanhaeren, M., D'Errico, F., Collcutt, S., Higham, T., Hodge, E., Parfitt, S., Rhodes, E., Schwenninger, J.-L., Stringer, C.B., Turner, E., Ward, S., Moutmir, A., Stambouli, A., 2007. 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. Proc. Natl. Acad. Sci. USA 104, 9964–9969. http://dx.doi.org/10.1073/pnas.0703877104.
Boyer, P., Bedoin, N., Honoré, S., 2000. Relative contributions of kind- and domain-level concepts to expectations concerning unfamiliar exemplars: developmental change and domain differences. Cogn. Dev. 15, 457–479.
Bremner, A.J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K.J., Spence, C., 2013. "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. Cognition 126, 165–172. http://dx.doi.org/10.1016/j.cognition.2012.09.007.
Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., Maneewongvatana, S., 2008. Encoding emotions in speech with the size code. A perceptual investigation. Phonetica 65, 210–230.
Corballis, M., 2003. From hand to mouth: the gestural origins of language. In: Christiansen, M.H., Kirby, S. (Eds.), Language Evolution. Oxford University Press, pp. 201–218.
Cuskley, C., Simner, J., Kirby, S., 2015. Phonological and orthographic influences in the bouba–kiki effect. Psychol. Res. 9, 389–397. http://dx.doi.org/10.1007/s00426-015-0709-2.
de Boer, B., Zuidema, W., 2010. Multi-agent simulations of the evolution of combinatorial phonology. Adapt. Behav 18, 141–154. http://dx.doi.org/10.1177/1059712309345789.
De Saussure, F., 1916. Cours de linguistique générale. Payot.
Fay, N., Arbib, M., Garrod, S., 2013. How to bootstrap a human communication system. Cogn. Sci. 37, 1356–1367. http://dx.doi.org/10.1111/cogs.12048.
Fónagy, I., 1983. La vive voix. Essais de psycho-phonétique. Payot.
Fónagy, I., 1961. Communication in poetry. Word 17, 194–218.
Foucault, M., 1989. The Order of Things: An Archaeology of the Human Sciences. Routledge.
Galantucci, B., Garrod, S., Roberts, G., 2012. Experimental semiotics. Lang. Linguist. Compass 6, 477–493. http://dx.doi.org/10.1002/lnc3.351.
Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2, 493–501.
Gelman, S.A., 2003. The Essential Child: Origins of Essentialism in Everyday Thought. Oxford University Press.
Herman, L.M., 2009. Language learning and cognitive skills. In: Perrin, W.F., Wursig, B., Thewissen, J.G.M. (Eds.), Encyclopedia of Marine Mammal, second ed. Academic Press, New York, pp. 657–663.
Hochberg, Y., Tamhane, A., 1987. Multiple Comparison Procedures. John Wiley & Sons.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70. http://dx.doi.org/10.2307/4615733.
Imai, M., Kita, S., 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 369, 20130298. http://dx.doi.org/10.1098/rstb.2013.0298.
Imai, M., Kita, S., Nagumo, M., Okada, H., 2008. Sound symbolism facilitates early verb learning. Cognition 109, 54–65. http://dx.doi.org/10.1016/j.cognition.2008.07.015.
Kaminski, J., Call, J., Fischer, J., 2004. Word learning in a domestic dog: evidence for "fast mapping". Science 304, 1682–1683 http://dx.doi.org/10.1126/science.1097859.
Kantartzis, K., Imai, M., Kita, S., 2011. Japanese sound symbolism facilitates word learning in English speaking children. Cogn. Sci. 35, 575–586.
Kirby, S., Hurford, J., 2002. The emergence of linguistic structure: an overview of the Iterated learning model. In: Cangelosi, A., Parisi, D. (Eds.), Simulating the Evolution of Language. Springer Verlag, London, pp. 121–148.
Köhler, W., 1947. Gestalt Psychology, second ed. Liveright.
Léger, L., Boumlak, H., Tijus, C., 2008. BASETY: Extension and typicality of the specimens for 21 categories of objects. Can. J. Exp. Psychol. 62, 223–232. http://dx.doi.org/10.1037/a0012885.
Maurer, D., Pathman, T., Mondloch, C.J., 2006. The shape of boubas: sound-shape correspondences in toddlers and adults. Dev. Sci. 9, 316–322. http://dx.doi.org/10.1111/j.1467-7687.2006.00495.x.
Miyazaki, M., Hidaka, S., Imai, M., Yeung, H.H., Kantartzis, K., Okada, H., Kita, S., 2013. The facilitatory role of sound symbolism in infant word learning. In: Knauff, M., Pauen, M., Sebanz, N., Wachsmuth, I. (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin, Texas, pp. 3080–3085.
Monaghan, P., Mattock, K., Walker, P., 2012. The role of sound symbolism in language learning. J. Exp. Psychol. Learn. Mem. Cogn. 38, 1152–1164. http://dx.doi.org/10.1037/a0027747.

New, B., Pallier, C., Ferrand, L., Matos, R., 2001. Une base de données lexicales du français contemporain sur internet : LEXIQUE™. Annee Psychol. 101, 447–462. http://dx.doi.org/10.3406/psy.2001.1341.
Newman, S.S., 1933. Further experiments in phonetic symbolism. Am. J. Psychol. 45, 53–75.
Nielsen, A.K.S., Rendall, D., 2013. Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. Can. J. Exp. Psychol. 67, 153–163. http://dx.doi.org/10.1037/a0030553.
Nielsen, A.K.S., Rendall, D., 2011. The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. Can. J. Exp. Psychol. 65, 115–124. http://dx.doi.org/10.1037/a0022268.
Nobile, L., 2015. Phonemes as images. An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In: Hiraga, M.K., Herlofsky, W.J., Shinoara, K., Akita, K. (Eds.), Iconicity: East Meets West. John Benjamins, pp. 71–91. http://dx.doi.org/10.1075/ill.14.04nob.
Nuckolls, J.B., 1999. The case for sound symbolism. Annu. Rev. Anthropol. 28, 225–252.
Ohala, J.J., 1997. Sound symbolism. In: Proceedings of the 4th Seoul International Conference on Linguistics [SICOL], pp. 98–103.
Ohala, J.J., 1994. The frequency code underlies the sound-symbolic use of voice pitch. In: Ohala, J.J., Hinton, L., Nichols, J. (Eds.), Sound Symbolism. Cambridge University Press, pp. 325–347.
Oudeyer, P.-Y., 2013. Self-organization: complex dynamical systems in the evolution of speech. In: Binder, P.-M., Smith, K. (Eds.), The Language Phenomenon, the Frontiers Collection. Springer-Verlag, pp. 191–216. http://dx.doi.org/10.1007/978-3-642-36086-2.
Ozturk, O., Krehm, M., Vouloumanos, A., 2013. Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. J. Exp. Child Psychol. 114, 173–186. http://dx.doi.org/10.1016/j.jecp.2012.05.004.
Paivio, A., 1975. Perceptual comparisons through the mind's eye. Mem. Cogn. 3, 635–647. http://dx.doi.org/10.3758/BF03198229.
Pepperberg, I.M., 2000. The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots. Harvard University Press.
Perlman, M., Dale, R., Lupyan, G., 2015. Iconicity can ground the creation of vocal symbols. R. Soc. Open Sci. 2, 150152. http://dx.doi.org/10.1098/rsos.150152.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing.
Ramachandran, V.S., Hubbard, E.M., 2001. Synaesthesia — a window into perception, thought and language. J. Conscious. Stud. 8, 3–34.
Riou, B., Lesourd, M., Brunel, L., Versace, R., 2011. Visual memory and visual perception: when memory improves visual search. Mem. Cogn. 39, 1094–1102. http://dx.doi.org/10.3758/s13421-011-0075-2.
Rizzolatti, G., Arbib, M.A., 1998. Language within our grasp. Trends Neurosci. 21, 188–194.
Roberts, G., Lewandowski, J., Galantucci, B., 2015. How communication changes when we cannot mime the world: experimental evidence for the effect of iconicity on combinatoriality. Cognition 141, 52–66. http://dx.doi.org/10.1016/j.cognition.2015.04.001.
Sapir, B.Y.E., 1929. A study in phonetic symbolism. J. Exp. Psychol. 12, 225–239.
Savage-Rumbaugh, E.S., Lewin, R., 1996. Kanzi: The Ape at the Brink of the Human Mind. John Wiley and Sons.
Schmidtke, D.S., Conrad, M., Jacobs, A.M., 2014. Phonological iconicity. Front. Psychol. 5, 80. http://dx.doi.org/10.3389/fpsyg.2014.00080.
Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S.A., Fraser, C., Scott, K., Ward, J., 2006. Synaesthesia: the prevalence of atypical cross-modal experiences. Perception 35, 1024–1033. http://dx.doi.org/10.1068/p5469.
Spector, F., Maurer, D., 2013. Early sound symbolism for vowel sounds. i-Perception 4, 239–241. http://dx.doi.org/10.1068/i0535.
Spector, F., Maurer, D., 2009. Synesthesia: a new approach to understanding the development of perception. Dev. Psychol. 45, 175–189. http://dx.doi.org/10.1037/a0014171.
Spence, C., 2011. Crossmodal correspondences: a tutorial review. Atten. Percept. Psychophys. 73, 971–995. http://dx.doi.org/10.3758/s13414-010-0073-7.
Steels, L., 2008. The symbol grounding problem has been solved, so what's next ? In: de Vega, M. (Ed.), Symbols and Embodiment: Debates on Meaning and Cognition. Oxford University Press, pp. 223–244.
Stevens, K.N., House, A.S., 1963. Perturbation of vowel articulations by consonantal context: an acoustical study. J. Speech Lang. Hear. Res. 6, 111–128.
Tanz, C., 1971. Sound symbolism in words relating to proximity and distance. Lang. Speech 14, 266–276. http://dx.doi.org/10.1177/002383097101400307.
Vanhaeren, M., d'Errico, F., Stringer, C.B., James, S.L., Todd, J.A., Mienis, H.K., 2006. Middle Paleolithic shell beads in Israel and Algeria. Science 312, 1785–1788. http://dx.doi.org/10.1126/science.1128139.
Westbury, C., 2005. Implicit sound symbolism in lexical access: evidence from an interference task. Brain Lang. 93, 10–19. http://dx.doi.org/10.1016/j.bandl.2004.07.006.
Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic F0 of vowels. J. Phonetics 23, 349–366.
Wicker, F.W., 1968. Mapping the intersensory regions of perceptual space. Am. J. Psychol. 81, 178–188.

2. Second study: Phonetic and conceptual contrasts in the assessment of sound symbolic associations: comparing protocols and inferring cognitive processes

Léa De Carolis[1] & Christophe Coupé[2]

[1] Laboratoire Dynamique du Langage, CNRS & Université de Lyon, Lyon, France

[2]Department of Linguistics, The University of Hong Kong, Hong Kong SAR, China

Article submitted to Cognition in May 2019

Phonetic and conceptual contrasts in the assessment of sound symbolic associations: Comparing protocols and inferring cognitive processes

Authors

Léa De Carolis[1]

Christophe Coupé[1,2]

Affiliation

[1]Laboratoire Dynamique du Langage, CNRS & Université de Lyon, Lyon, France

[2]Department of Linguistics, The University of Hong Kong, Hong Kong SAR, China

Abstract

The best-known paradigm in the study of motivated associations is the bouba-kiki task, which consists in associating a pseudo-word to a visual shape – round or spiky. There are many variations across experimentations regarding the experimental settings and the population being tested, such as: the language spoken by participants (e.g., English, French, Dutch), their age (adults and children), the segments composing the pseudo-words (multiple vowels and consonants), the visual or conceptual contrasts (mostly graphical shapes, but also bird and fish names in Berlin's study (1994)). The specific task to achieve also differ between studies: a choice between two pseudo-words and two shapes (2x2), a choice between two pseudo-words for one shape (1x2), a choice between two shapes for one pseudo-word (2x1), and finally a judgment (for example) about a matching between one pseudo-word and one shape.

This study aims to assess the influence of the presence or the absence of the two possible contrasts, between segments and between concepts, by comparing the four possible types of presentation (2x2, 1x2, 2x1 and 1x1). The concepts are animal features – biological class, i.e. bird vs. fish, size, dangerousness and repulsiveness – and ten hypotheses found in the literature on sound symbolism are assessed, e.g. the association between large objects and [a], and between small objects and [i]. Segments, concepts and profiles of participants are kept constant across four protocols. Results reveal obvious differences between the protocols, more precisely different numbers of statistically validated hypotheses, and weaker or stronger average effect sizes. Each of our ten hypotheses is confirmed at least once across the four protocols, and no contradiction appears. Some associations are systematically confirmed, as the one previously mentioned between vowels and size. With respect to effect sizes, the presence of both conceptual and phonetic contrasts (2x2) leads to the highest values, followed by the presence of the phonetic contrast only (1x2). This suggests a facilitating effect of the phonetic contrast in the detection of sound symbolic associations. Overall, the differences point to the relevance of investigating the cognitive processes at play in the production of sound symbolic associations.

Keywords

Sound symbolism; phonetic contrast; conceptual contrast; association and judgment tasks; methodology; protocols

Introduction

Overview about sound symbolism

Sound symbolism is a cognitive process consisting in linking linguistic sounds to other modalities (as visual stimuli of different shapes or sizes). It refers in particular to the hypothesis that some phonetic units intrinsically carry semantic content. Among various proposals found today in the literature, it has been for example demonstrated more 90 years ago that Western participants tend to associate the vowel [a] with large objects, and the vowel [i] with small objects (Sapir, 1929).

Sound symbolism can be studied from different perspectives and with different methodologies. In particular, experimental protocols in psycholinguistics are commonly relied upon. The best-known of these protocols is undoubtedly the bouba-kiki task, which consists in the presentation of two shapes – a round one and a spiky one – and two pseudo-words – 'bouba' and 'kiki', although early versions of the task actually relied on different items, e.g. 'maluma' and 'takete' (Köhler, 1947). Participants have to decide which association, among four possibilities, is the best. Most surveyed people associate 'bouba' with the round shape, and 'kiki' with the spiky shape. This has been shown to be the case in up to 95 % of Western participants, according to Ramachandran & Hubbard (2001), but also in other populations and cultures (Bremner et al., 2013; Chen, Huang, Woods, & Spence, 2016; Davis, 1961).

While the description above is the canonical bouba-kiki paradigm, many variations can be found today in the literature. At the expense of comparability, they reflect various approaches to assessing occurrences of sound symbolism, their properties and the underpinning processes.

On the one hand, the stimuli presented to participants can differ. First, experimenters use various pseudo-words. A reason for this is that the aforementioned patterns of association involving 'bouba' and 'kiki' (or similar "complex" words with several different consonants or vowels) cannot be readily explained: is the association between 'bouba' and round shapes due mostly to the consonant [b], or to one of the two vowels [u] and [a]? Conversely, is the association between 'kiki' due mostly to the consonant [k] or to the vowel [i]? At a deeper level, which articulatory or auditory features of these units are involved in the cognitive operations leading to the associations? Experimenters thus choose various linguistic stimuli to address these issues and focus on specific units with respect to sound symbolic properties. These units and their contrast, which are thus embedded in pseudo-words (e.g. contrasting [i] and [a] in a [_p_p] context, resulting in the two stimuli [ipip] and [apap]), can be specific consonants or vowels, or articulatory or auditory features of these consonants and vowels: aperture and anteriority for vowels, mode and place of articulation for consonants, voicing, frequency of formants, etc. (Aveyard, 2012; Knoeferle, Li, Maggioni, & Spence, 2017; Nielsen & Rendall, 2011) General questions such as the relative weight of consonants and vowels in sound symbolism have been raised (Fort, Martin, & Peperkamp, 2015; Tarte, 1974), but aspects like sonority (Westbury, 2005) or the length of the vocal tract (Chuenwattanapranithi, Xu, Thipakorn, & Maneewongvatana, 2008) have also been investigated. Additionally, pseudo-words are sometimes presented orally, e.g. (Fort et al., 2015; Monaghan, Mattock, & Walker, 2012), while in other experiments they appear as written forms, e.g. (Cuskley, Simmer, & Kirby, 2015; Nielsen & Rendall, 2012, 2013; Westbury, 2005). Second, there is diversity in the non-linguistic material. Many studies rely on spiky and round shapes but create sets of shapes with graphical variations (Fort et al., 2015). Other visual forms or finer-grained visual properties can also be investigated, as reported in the next section (Nobile, 2015). But, as it will be the case in this contribution, non-visual conceptual material can also be used, as exemplified by Berlin (1994)'s study which focused on animals and the sound symbolic properties of their names.

On the other hand, beyond the choice of stimuli, the protocols used to introduce them to participants are also varied. The associations may in particular be collected in an explicit way, as it is the case in

most associative tasks, or in an implicit way (Vainio, Tiainen, Tiippana, Rantala, & Vainio, 2017; Westbury, 2005). More generally, the various paradigms or tasks can be grouped in three main categories:

- learning paradigms (Aveyard, 2012; Monaghan et al., 2012; Nielsen & Rendall, 2012);
- judgment tasks (Knoeferle et al., 2017; Perfors, 2004);
- associative tasks (Berlin, 1994; Bremner et al., 2013; Chen et al., 2016; De Carolis, Marsico, & Coupé, 2017; Fort et al., 2015; Nielsen & Rendall, 2011; Nobile, 2015; Sapir, 1929; Sidhu & Pexman, 2017; Tarte, 1974; Turoman & Styles, 2017).

Alongside these categories of tasks, there are differences in the phonetic or conceptual contrasts used in each study. Each of these two types of contrast can be present or absent, resulting in four possible presentations to the participants as illustrated by Fig. 1: 2x2 (two images and two pseudo-words), 2x1 (two images and one pseudo-word) and 1x2 (one image and two pseudo-words) in the case of associative tasks, and 1x1 (one image and one pseudo-word) in categorization/learning tasks (non-exhaustively).

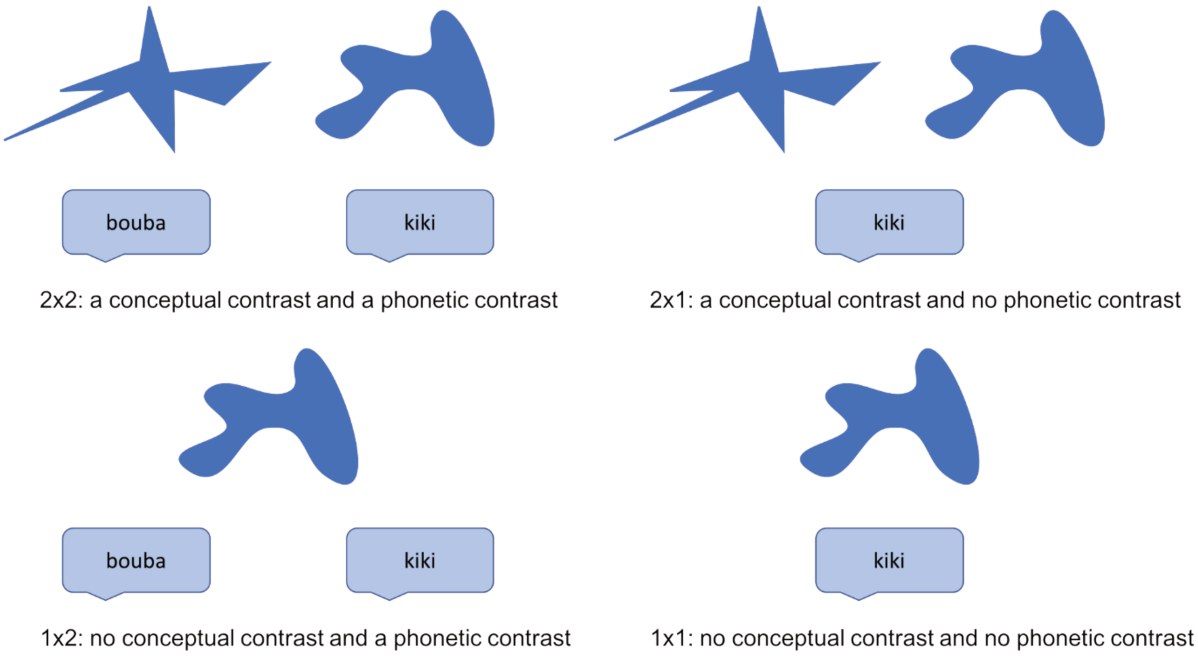


Figure 1. Four possibilities of presentation in the study of bouba-kiki associations

Overall, there is thus considerable variability in the experimental protocols, which leads to two different concerns: first, can we compare results on sound symbolism based on different methodologies, and if yes, how? Second, are these different protocols congruent regarding sound symbolism and the corresponding cognitive processes? Our goal is to provide answers to the second question, with a focus on the four possible presentations mentioned before and illustrated in Fig. 1. These four presentations and various results that were obtained with them are further described in the next section. We then present our own methodology – our experimental stimuli, our procedures and our participants –, before reporting our results both across and within protocols. We discuss these results in the light of cross-study comparisons and necessary phonetic and/or conceptual contrasts for subjects to produce sound symbolic associations.

Presentation of the different protocols

This section reports different studies according to the presentation of a conceptual and/or linguistic contrast in their experimental protocols. Overall considerations about these different protocols and what they suggest are then discussed in the following section.

2x2: two visual/conceptual stimuli, two pseudo-words

In 2015, Nobile conducted a study that aimed at simultaneously evaluating visual and phonetic features. Participants assigned one pseudo-word of a pair (differing in one phonetic feature) to one of two different shapes. The pseudo-words were displayed in a written form and uttered by the experimenter. The consonantal features evaluated were: (1) voicing (voiced vs. voiceless), (2) manner (fricative vs. plosive), (3) nasality (oral vs. nasal), (4) place of articulation (palato-velar vs. alveo-dental). The visual features contrasted were the following: (1) shape curvature (curved vs. angular), (2) acuity (obtuse vs. acute), (3) continuity (continuous vs. discontinuous), (4) density (dense vs. sparse) and (5) regularity (regular vs. irregular). Results show, for example, that voiced consonants are associated with curved, obtuse and continuous features, and, conversely, voiceless features are associated with angular, acute and discontinuous features (Nobile, 2015).

As already mentioned, Berlin (1994)'s study is of particular interest to us because it deals with animal names, which are the subject of the present study. He collected names of birds and fish in Huambisa, a Peruvian language. Pairs of names (each containing a name for a bird and a name for a fish) were presented to non-Huambisa-speaking students (in the same condition as in Nobile's experiment, in written forms and orally pronounced). The students had to decide which one refers to a bird, knowing that the second refers to a fish. The overall performance was significantly higher than chance (58%), which suggested cross-linguistic sound symbolic associations. However, one can ask what would have been the results if the participants had not been told explicitly that the second word designated a fish. Indeed, one cannot rule out that at least some participants' choice was based on both animals, rather than only the bird. Phonological analyses of Huambisa names revealed specific patterns of frequencies of occurrences for a range of phonetic units. For example, regarding the initial syllable, there is more [i] in bird names than in fish names, themselves containing more [a] than [i]. Regarding consonants, there are more nasals and continuants in the final syllables of fish names than in the final syllables of bird names, and more obstruents in the final syllables of bird names than in the final syllables of fish names. The fact that the frequencies of occurrence of some phonetic categories differ between two biological classes of animals within a language may indicate some motivated relation between sounds and meaning, but this is not directly examined in this paper – see (De Carolis et al., 2017) for a discussion of this issue. However, the fact that Huambisa speakers can rely on 'symbolic' phonetic features when accessing a meaning and that these associations seem shared by speakers of other languages suggests the possibility of linking this phenomenon to the one underlying bouba-kiki associations.

1x2: one visual/conceptual stimulus, two pseudo-words

In Chen et al. (2016)'s experiment, each trial consisted in giving the subjects a choice between the two classical pseudo-words 'bouba' and 'kiki' (presented orally) for a particular visual stimulus. There was a large variety of visual stimuli, with the two original shapes used in Bremner et al. (2013) modified to differ additionally in the number, amplitude and spikiness of the points of the star-shaped figures. The aim of the study was to investigate cultural differences in patterns of association. Participants from the United States (recruited through Amazon Mechanical Turk) and Taiwanese students were tested. Both groups associated the classical spiky shape with 'kiki' (respectively 90.8 and 86.7%) rather than 'bouba'. However, only the Taiwanese participants associated the round shape to 'bouba' rather than to 'kiki' (60.9 vs. 50.6%). Moreover, the three visual parameters (frequency, amplitude and spikiness)

influenced the answers for both populations: the enhancement of any of them increased the proportion of 'kiki' answers. Another study conducted by Nielsen and Rendall (2011) also consisted in the presentation of one visual stimulus (either spiky or round) and of two pseudo-words, also presented orally, contrasted either on consonants or on vowels. The results revealed a larger effect of consonants than of vowels (59% vs. 51%), and, in consonantal contrasts, choices were more sound symbolic for spiky shapes than for round shapes (63% vs. 54%), similar to the previous study. Moreover, these results were compared to those of a 2x2 protocol, using similar material, but with pseudo-words presented as written forms. The comparison revealed a weaker overall effect in the case of the 1x2 protocol. Two possible explanations were: 1) the opacity of the protocol (because of its lack of visual contrast); 2) the oral modality for the presentation of pseudo-words. However, the second explanation seems less likely given Sidhu and Pexman (2017)'s results with a 1x1 protocol where pseudo-words were presented either orally or in written forms, suggesting stronger effects in the oral condition than in the written one, even if there was no direct comparative test between the two.

2x1: two visual/conceptual stimuli, one pseudo-word

In the study of Fort et al. (2015), an associative task that implied a choice between two shapes for a given oral pseudo-word led to analyses of consonantal and vocalic influences, both independently and in interaction. Two structures of pseudo-words (CVCV and VCV) were investigated. The results revealed a larger impact of consonants on the participants' choices, whatever their position as first or second phoneme in the pseudo-word. There were no effects from vowels considered independently, only effects in interaction with consonants. With the same approach of contrasting two visual stimuli, Turoman and Styles (2017) focused on the role of visual features. They presented pairs of glyphs from different writing systems. In each pair, the pronunciation of one glyph contained the sound [i], and the pronunciation of the other the sound [u]. The task for the participants consisted in guessing which glyph had a pronunciation containing [u] or [i], depending on the experiment. Participants presented a higher than chance performance in both conditions. Then, the authors studied the spatial frequencies of glyphs and highlighted a difference in complexity and line length in the more guessable pairs: the glyphs that contained [u] sounds were more complex and had greater line length in comparison with [i] sounds.

1x1: one visual/conceptual stimulus, one pseudo-word

Protocols of type 1x1 can take several forms. In 2005, Westbury conducted a lexical decision task in which words and pseudo-words appeared in round or spiky frames. The results revealed implicit sound symbolic associations: pseudo-words composed of sonorants were more quickly categorized as pseudo-words when presented in round frames, as were those composed of plosives when appearing in spiky frames (Westbury, 2005). However, De Carolis et al. (2018) did not obtain similar results with a derived task presented to French speakers, where an additional variable – the font used to write words and pseudo-words – was considered. No sound symbolic association was detected, but an unexpected interaction between the angular font and spiky frames was shown, reviving discussion about the impact of the shapes of letters. In 2012, Nielsen and Rendall tested the impact of congruency between shapes (round and spiky) and phonemes in a learning paradigm. Pairs – composed of a shape and a pseudo-word – were presented sequentially with an indication about their correctness. The participants were split into two conditions: the learning was either congruent (according to previous studies) or incongruent. After the learning phase, they had to decide about the correctness of new pairs. The participants in the congruent condition presented a significantly higher performance than chance (53.3%), contrary to the performance in the incongruent condition (50.4%) (Nielsen & Rendall, 2012). These results may seem weak but the difficulty of the task and its implicit nature attest to the existence of sound symbolic associations beside metacognitive strategies, which may amplify these associations. Kovic et al. (2010) proposed another implicit paradigm: in a learning phase, a cartoon

creature composed of different round and spiky elements was presented with two different pseudo-words. These pseudo-words contained sounds which were related to either spiky or round shapes according to well-established sound symbolic associations. The participants had to assign the stimulus to one of the two pseudo-words, and feedback indicating the correctness of their answers followed. This correctness depended only implicitly on the shape of the head-element of the creature but this was not indicated. Then, in the test phase, a pair of a visual stimulus and a pseudo-word were displayed and the participants had to decide if it was a match or a mismatch. The learning phase was therefore 1x2, but the test phase 1x1. The responses were faster in the congruent condition – congruent according to previous studies. Moreover, in the incongruent condition, the participants were slower to reject pairs that were sound symbolically congruent. As a result, a bias was said to exist in favor of sound symbolic associations.

Building consensus across protocols

Besides the differing linguistic and non-linguistic material, the principal difference between the previous subtypes of paradigms is the presence or the absence of a phonetic contrast and/or a visual/conceptual contrast explicitly proposed to participants. All the results that arise from these studies must be contextualized, in the sense that a preferred association between, for example, the vowel [i] and 'small' is relative to another vowel in the 2x2 and 1x2 protocols, relative to 'large' in the 2x2 and 2x1 protocols, and 'intrinsic' in the 1x1 protocol. At the same time, however, even if contrasts are not explicitly presented to the participants, successive trials in an experiment allow experimenters to study it in a between-trial rather than within-trial fashion.

In relation to the preceding point, as for 2x2 protocols, Nobile and Berlin's studies reveal sound symbolic associations with a within-trial approach for both phonetic and visual/conceptual parameters. As for the 1x2 paradigm, Chen et al. (2016)'s and Nielsen and Rendall (2011)'s experiments both contrast phonetic features but their conclusions focus on two different aspects: visual features for the former with a between-trial perspective, and phonemes for the latter with a within-trial perspective. The same dichotomy occurs in 2x1 paradigms – while contrasting visual features, Fort et al. (2015) present results about consonants and vowels with a between-trial analysis, whereas Turoman and Styles (2017) focus on visual parameters with a within-trial analysis. Finally, in 1x1 paradigms, authors (Kovic et al., 2010; Nielsen & Rendall, 2012; Westbury, 2005) formulate results both on visual and phonetic features, necessarily with a between-trial approach. As a conclusion, the analyses that can applied to participants' responses do not necessarily mirror the structure of the task and whether contrasts are explicitly presented to participants. A single protocol can enlighten distinct purposes, and vice versa. Nevertheless, one can ask whether results are identical in the different approaches, in particular whether contrasts are explicitly presented to the participants or not.

The benefit of a 2x2 experiment is the simultaneous evaluation of two parameters in different modalities, here phonetic and visual. However, most studies request a single association from participants. Therefore, for example, more associations may be made between [i] and spiky when presented with [a] and round, but this does not imply a sound symbolic association between the latter for the participants. Nobile (2015)'s study differs here in that two associations are requested from participants, one per pseudo-word.

Hence, this study aims at comparing different types of paradigms, preserving the same phonetic and conceptual material, experimental conditions and population through the investigation of sound symbolic associations, using associative judgement, and memory/recognition tasks. Contrary to most studies on sound symbolism, we do not use visual shapes but labels which refer to different types of animals. This study evaluates several conceptual contrasts: size ('large' vs. 'small'); dangerousness ('dangerous' vs. 'harmless'); repulsiveness ('repulsive' vs. 'attractive'); biological class ('fish' vs. 'bird').

Except for the latter, other contrasts are inherently contrastive (i.e. large is naturally opposed to small). This approach offers shared foundations to compare different protocols used in literature, and to evaluate their respective propensity to shed light on sound symbolic associations.

Methodology

Hypotheses

In order to compare the various protocols, one needs a set of hypotheses to be assessed similarly in each of them. Ten hypotheses applicable to animals were selected on the basis of previous studies in the literature and related to conceptual categories for specific segments (vowels or consonants). Pseudo-words that contrasted on the relevant consonants and/or vowels were then built to create the adequate experimental material (see below for more details). Given the various resulting conceptual and phonetic contrasts, one could then expect 'sound symbolic congruent' or 'incongruent' answers from the subjects.

The hypotheses are first summarized in general terms in Table 1, and illustrated with pseudo-words in Table 2. In some cases, the general hypotheses are stated in terms of phonetic features such as voiced, voiceless, plosive, sonorant, front, back etc. In others, the opposition takes place directly between segments, i.e. the opposition between [i] and [a] for size is not explained in terms of differences in aperture or frontness.

Conceptual category	**'General' hypotheses**		**Reference**
Size	Voiced C – large	Voiceless C – small	(Sapir, 1929) etc.
	[i] – small	[a] – large	(Ozturk, Krehm, & Vouloumanos, 2013)
Class	Fricatives – fish	Plosives – bird	(Berlin, 1994)
	[a] or [u] – fish	[i] – bird	
Repulsiveness	Plosives – repulsive	Sonorants – attractive	(Fónagy, 1961, 1983)
	[a] – repulsive	[i] – attractive	
Dangerousness	Back V – dangerous	Front V – harmless	(Fónagy, 1961, 1983)
	Voiced C - dangerous	Voiceless C - harmless	
	Plosives – dangerous	Sonorants – harmless	
	Back C – dangerous	Front C - harmless	

Table 1. General hypotheses for assessing sound symbolism in the different protocols. C stands for consonants, V for vowels.

Conceptual category	**Pseudo-words**	**'Instantiated' hypotheses with pseudo-words**	
Size	[ipip] [ibib], [apap], [abab]	[abab] & [ibib] – large	[apap] & [ipip] – small
		[ipip] & [ibib] – small	[apap] & [abab] – large
Class	[isis], [asas], [itit], [atat]	[isis] & [asas] – fish	[itit] & [atat] – bird
		[atat] & [asas] – fish	[isis] & [itit] – bird
Repulsiveness	[inin], [ikik], [anan], [akak]	[ikik] & [akak] – repulsive	[inin] & [anan] – attractive
		[akak] & [anan] – repulsive	[ikik] & [inin] – attractive
Dangerousness	[igig], [imim], [ugug], [umum]	[ugug] & [umum] – dangerous	[igig] & [imim] – harmless
	[ubub], [ugug], [upup], [ukuk]	[ubub] & [ugug] - dangerous	[upup]& [ukuk] - harmless
	[idid], [ubub], [ilil], [umum]	[idid] & [ubub] – dangerous	[ilil] & [umum] – harmless
	[upup], [ukuk]	[ukuk] – dangerous	[upup] - harmless

Table 2. Hypotheses instantiated with pseudo-words for assessing sound symbolism in the different protocols.

Material

Labels

Eight different labels, i.e. eight written expressions referring to different types of animals, were used in all four tasks, with two for each conceptual contrast: 'a small animal' and 'a large animal' for size, 'a dangerous animal' and 'a harmless animal' for dangerousness, 'a repulsive animal' and 'an attractive animal' for repulsiveness and 'a fish' and 'a bird' for the biological class. Labels were either presented alone (1x2 and 1x1) in the center of the screen, or by pair (2x1 and 2x2) (one on the left, the other on the right according to a random selection in the script). They appeared in white on a black background with the font Mono 30 pts.

Visual stimulus for oral pseudo-words

A visual stimulus representing a loudspeaker was used to represent an oral pseudo-word. In the case of a vocalic contrast, two icons were present, one on the left for the first pseudo-word, and one on the right for the second one. Each icon was enlarged while the pseudo-word was played (or during 764ms in 2x2, because of scripting limitations, see below). The baseline size was 210*210px, the enlarged size 300*300px.

Pseudo-words

Twenty-one VCVC pseudo-words were generated, using three vowels ([i], [a] and [u]) and ten consonants ([b], [d], [g], [p], [t], [k], [s], [m], [n], [l]). In 2x1 and 1x1, for any conceptual contrast, participants (described below) were exposed only once to a target segment. They therefore dealt with a selection of pseudo-words. For example, four pseudo-words were used for size contrasts: [abab], [apap], [ibib] and [ipip]. Half of participants only encountered [abab] and [ipip] (version #1) and the other half only [ibib] and [apap] (version #2). In 2x2 and 1x2, participants saw all pseudo-words and were exposed twice to a target contrast, more precisely in two different contexts. For size for example, half of the participants encountered a vocalic contrast in two different consonantal contexts ([ipip] – [apap] and [ibib] – [abab]), and the other half a consonantal contrast in two vocalic contexts ([ibib] – [ipip] and [abab] – [apap]). Adopting this strategy led participants to be exposed to very similar number of trials, whatever the protocol they were subjected to. The pseudo-words and their repartition according to the conceptual contrasts and versions are presented in supplementary information. The number of trials and the pseudo-words differed according to the protocols and versions (see Table 3). Only one pseudo-word ([ikik]) appeared for two different conceptual contrasts (dangerousness and repulsiveness), but only once for each subject.

Eight additional pseudo-words were used for the training phase: [yzyz], [usus], [ypyp], [adad], [agag], [udud], [ifif], [aʁaʁ] (the four first in 1x1 and 2x1; all of them in 1x2 and 2x2).

	2x1	**1x2**	**2x2**	**1x1 part 1**	**1x1 part 2**
Nb of trials	12	10	10	12	25
Nb of pseudo-words	12	20	20		

Table 3. Number of pseudo-words and trials in each protocol

For the recognition test following the 1x1 protocol, the nine pseudo-words from the set of 21 pseudo-words that were not heard in the first phase were added, as well as four more unused pseudo-words: [afaf], [uʃuʃ], [yʁyʁ] and [yvyv].

Audio recording

A male French native speaker unaware of the hypotheses of the experiment recorded the pseudo-words in a quiet isolated room. The mean duration of a recording was 634 ms (SD: 54). The mean F_0 was 122 Hz (SD: 8.11).

For the recognition test in the 1x1 protocol, another set of recordings from another male French native speaker was used. The mean duration of a recording was 562 ms (SD: 67). The mean F_0 was 150 Hz (SD: 30.5).

Procedures

Each task began with a training phase with four trials, in order to ensure that the participants did understand the instructions. The experimenters thus had the possibility to check whether they used the correct medium to answer (the mouse or the keyboard, depending on the protocol). The participants who misunderstood at this stage were invited to read the instructions again, then to perform the training phase once again. The pseudo-words heard at this stage were not repeated later during the experiment. In 2x1, 1x2 and 1x1, the duration of the icon enlargement was adjusted according to the duration of the pseudo-words in OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). However, in 2x2, the script was more complicated since it involved two answers, and we were not able to adapt the duration of the icon enlargement to the duration of the pseudo-words. We instead determined a default duration (764 ms) that covered every possible pseudo-word (the maximum pseudo-word duration was 736 ms). The maximum duration to answer was adjusted according to the difficulty of the task and the feedback collected from colleagues during the testing phase.

2x2 (see Figure 1)

Icons were displayed in the same size until the mouse selection: when an icon was selected, it was enlarged; when a label was selected, it was framed by a green rectangle.

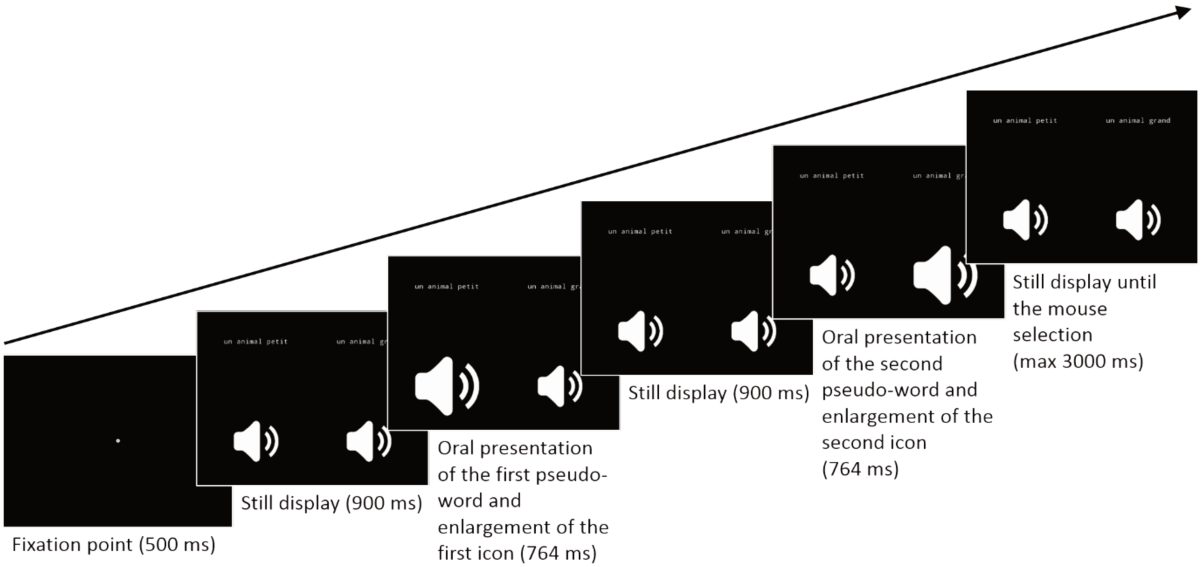


Figure 2. Representation of the succession of experimental phases in a single trial of the 2x2 protocol

1x2 (see Figure 2)

Icons were displayed in the same size until key selection: the left key for the left icon, the right key for the right icon.

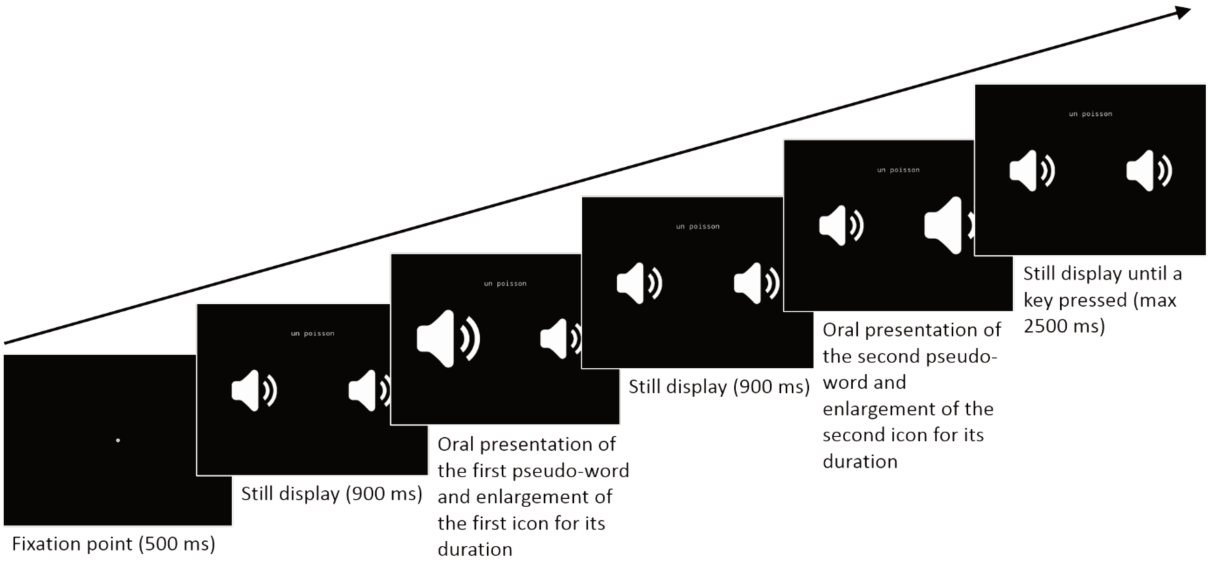


Figure 3. Representation of the succession of experimental phases in a single trial of the 1x2 protocol

2x1 (see Figure 3)

The participants made their choice by pressing the left key for the left label or the right key for the right label. Because the presentation of stimuli was simplier than other protocols, the duration of a 'still display' was longer than in other protocols.

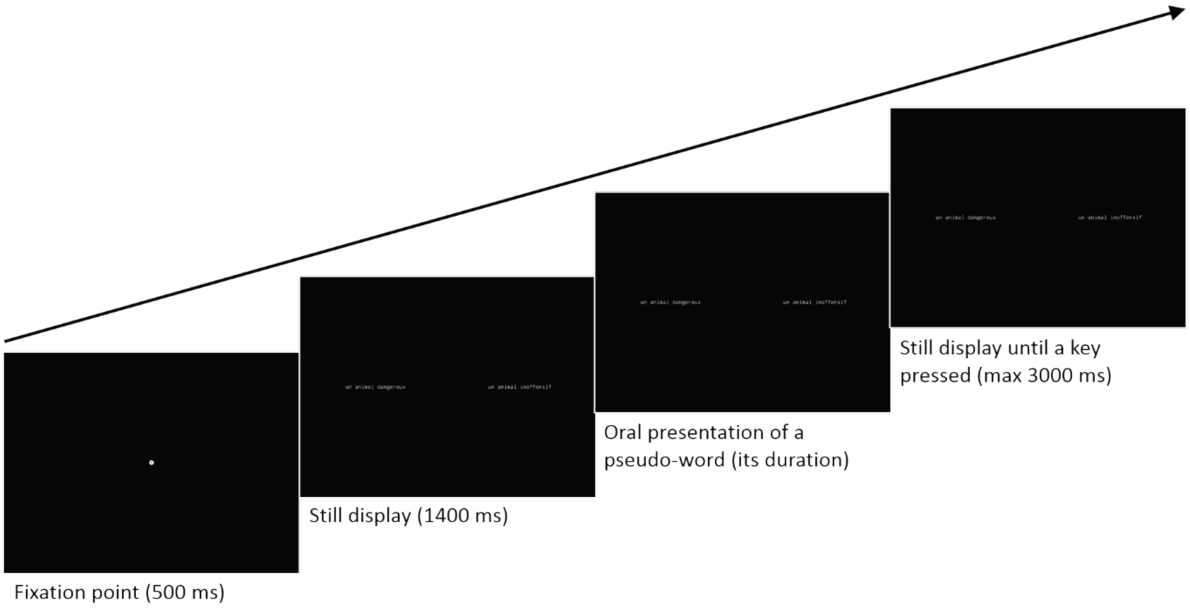


Figure 4. Representation of the succession of experimental phases in a single trial of the 2x1 protocol

1x1 (see Figure 4)

The scale was composed of 11 boxes, containing numbers from 0 to 10. Above the scale was written ‘not at all satisfied’ on the extreme left and ‘very satisfied’ on the extreme right (Mono, 18 pts). The participants made their choice by selecting a box with a mouse click.

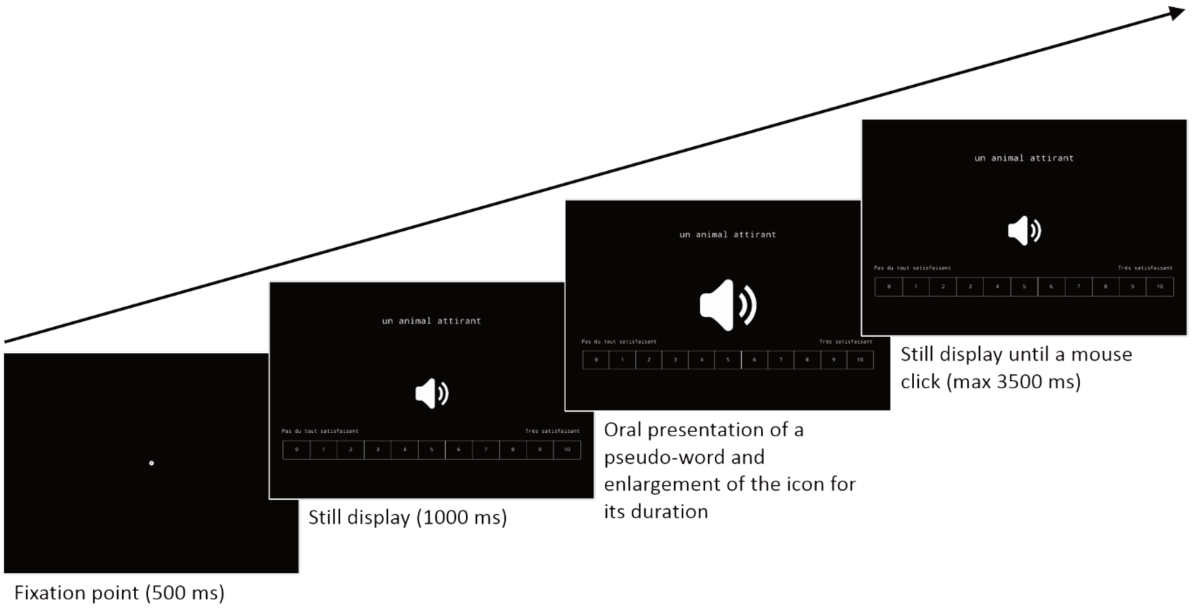


Figure 5. Representation of the succession of experimental phases in a single trial of the first part of the 1x1 protocol

1x1 part 2 (see Figure 5)

The participants made their choice by pressing one key: the left key to answer ‘no, I had not heard this word’ or the right key to answer ‘yes, I had already heard this word’.

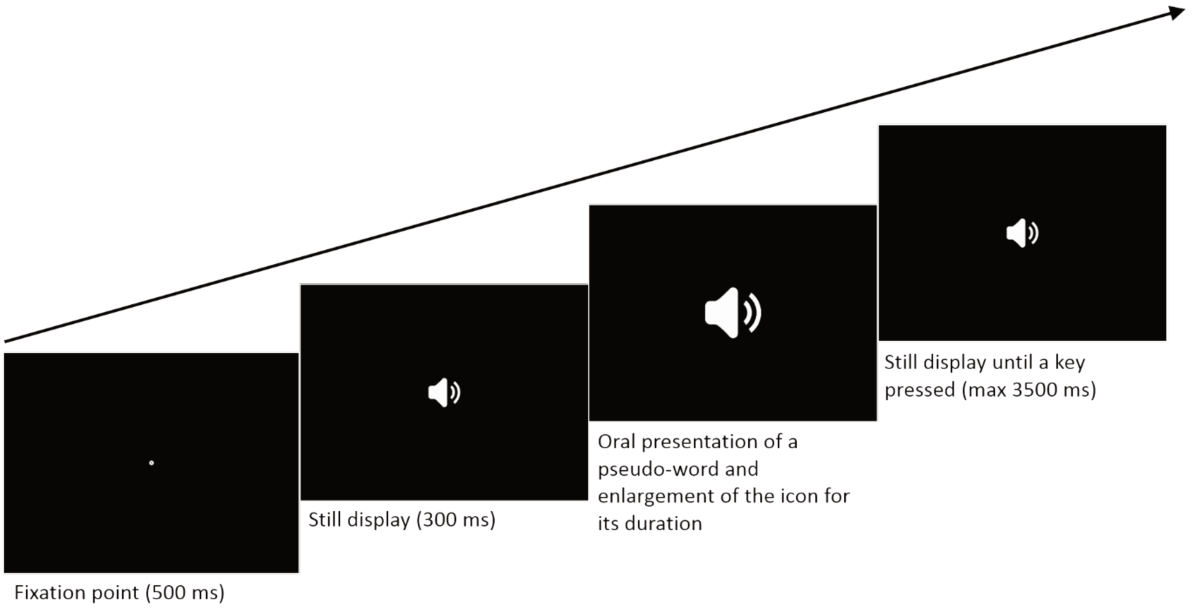


Figure 6. Representation of the succession of events in a single trial in the second part of the 1x1 protocol

Within-trial vs. between-trial contrasts

In each protocol, contrasts are present either in a trial (within-trial contrasts) or in the overall experiment (between-trial contrasts). In 2x2, both contrasts – conceptual and phonetic – are present within a trial. In 1x2, there are phonetic within-trial contrasts and conceptual between-trial contrasts. In 2x1, there are conceptual within-trial contrasts and phonetic between-trial contrasts. Finally, in 1x1, no within-trial contrast is present but they exist in a between-trial manner. Contrasting the different protocols is thus partly related to investigating how within-trial contrasts differ from between-trial contrasts, either at the phonetic or at the conceptual level.

Detecting between-trial conceptual contrasts in 1x2 and 1x1

In 1x2 and 1x1, the conceptual between-trial contrasts may potentially be detected by subjects as the experiment unfolds, since some labels are by nature in opposition (e.g. 'small' and 'large'). Hence, these protocols do not totally prevent the subjects from being influenced by these contrasts that exist in the entire experiment. Nevertheless, the 1x2 protocol does inform us about the sound symbolic associations that appear when the answer is a selection of one pseudo-word among two. Nevertheless, as with other protocols, we contrasted different concepts throughout the experiments and the trials occurred in a random order. This random alternation may have disguised the between-trial contrasts. In addition, it may have minimized the strength of the within-trial contrasts (in 2x2 and 2x1), since it lessened the training and reinforcement effects. As for phonetic between-trial contrasts, they may be less evident since pseudo-words differ on both consonants and vowels (e.g. one may hear 'ipip' and 'abab' for size contrasts).

Participants

Most participants agreed to participate in the experiment upon being invited to do so during a scientific event, the European Researchers' Night, organized by the University of Lyon on September 30, 2016. Additional data were collected in a second phase with students of the University Lumière Lyon 2. Given that the audience present during the scientific event was very diverse, we only analyzed answers of participants that fulfilled the following inclusion criteria: to be a native and monolingual speaker of French, to have no impairment; linguistic, hearing or visual, and to never have been involved in previous studies conducted by us. All the participants signed an informed consent. The groups are presented in Table 4.

Design	Nb of participants	Nb of males	Age span	Average age
2x2	41	18	17-67	26.5
1x2	64	21	16-56	23.4
2x1	48	17	17-63	25.2
1x1	36	16	18-77	31.1

Table 4. Presentation of the participants per protocol

For the 1x2 protocol, the number of participants differ significantly from those of the other protocols. In fact, 45 participants accomplished the task during the European Researchers' Night with scripts in which labels were randomly selected. However, we realized that the random selection of labels led to serious gaps in the results, and that the expected overall compensations did not occur (i.e. not as many answers for 'a small animal' with a vocalic contrast as with 'a large animal', and not as many of these answers as with a consonantal contrast). Hence, we recruited 19 supplementary participants with modified scripts containing no random selection in order to fill gaps and obtain the same number of answers in the different conditions.

Results

Contingency tables

The first approach that can be considered to analyze our data is contingency tables. In the case of 2x2, contingency tables naturally mirror answers with the four possibilities participants had. A table can be built for each instantiated hypothesis. There are therefore, for example, two tables for the conceptual contrast of size, one for the phonetic contrast between voiced consonants and voiceless consonants ([abab] and [ibib] versus [apap] and [ipip]), and one for the phonetic contrast between [i] and [a] ([ipip] and [ibib] versus [apap] and [abab]). For 1x2 and 2x1, contingency tables were constructed by simulating within-trial contrasts with between-trial contrasts. Finally, the 1x1 protocol did not consist in counts since the answers were judgements on a scale – which were averaged per answer. The contingency tables by protocol are reported in the 'Results by protocol and tested hypothesis' subsection below.

Contingency tables are classically associated with Fisher's exact test – or the (approximate) chi-square test – to assess the significance of the association, while Cramer's V is a possible measure of the size of the effect.

Generalized linear models and effect sizes

Fisher's exact test assumes independent observations, which is not the case in our experiments, since contrasts take place within contexts, i.e. a contrast of vowels is presented in a consonantal context and vice versa. Hence, this test does not permit us to properly assess sound symbolic associations, since we know from earlier studies that both consonants and vowels may have influences, and therefore contexts cannot be considered as neutral. Moreover, this test is not possible for the 1x1 protocol, in which the participants' answers are evaluations on a 0-to-10 scale, which do not lead to contingency tables. Therefore, we opted for other statistical measures that enabled direct comparisons between protocols, i.e. measures of effect size that could be applied to all of them regardless of their differences. We thus turned to binomial regression models for 2x2, 2x1 and 1x2, and to a Poisson regression for the 1x1 protocol, and in each case computed partial R^2 for the various predictors. The Poisson regression appeared to be a reasonable choice for the 1x1 protocol despite the upper boundary of the distribution of the answers, and was in particular better suited to the subjects' answers than linear regression with respect to the distribution of residuals – less heteroscedasticity and deviance from normality.

As an assessment of partial R^2 as relevant measures of effect size in generalized linear models, we compared them to Cramer's V measures related to contingency tables in 2x2, 2x1 and 1x2, and to eta squares related to linear models in 1x1 (eta squares are not available for a generalized linear model such as the Poisson regression, hence our choice to consider here linear regression despite it being less adapted than the Poisson regression), in order to verify the congruency between these approaches. Figures 6 and 7 report correlation tests for the associative tasks combined (2x2, 2x1 and 1x2) and for the judgment task (1x1), respectively.

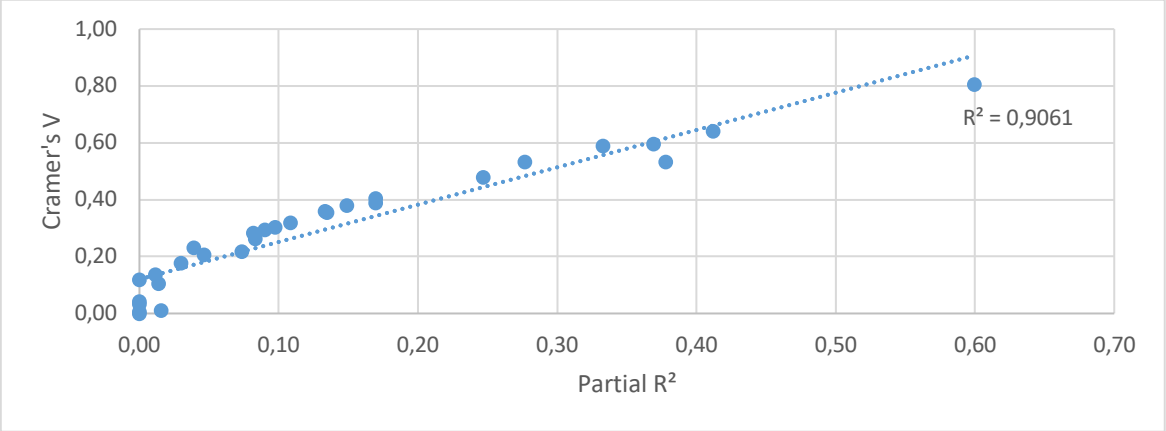
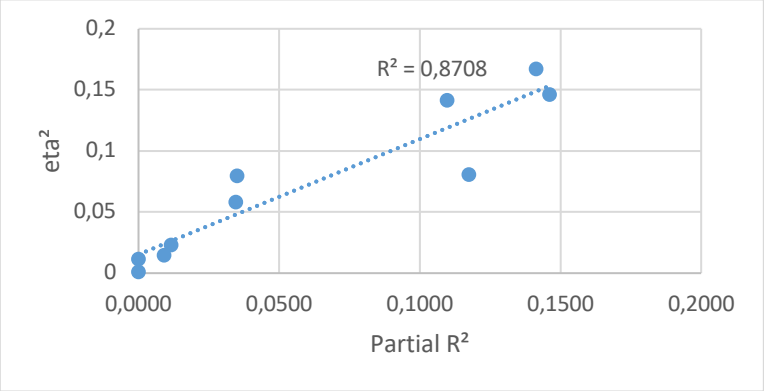


Figure 6. Assessment of the correlation between two measures of effect size – partial R^2 and Cramer's V – for the association tasks (2x2, 2x1 and 1x2)



It appears clearly that the different measures of effect size are congruent, and the small differences may be explained by the inadequacy of the simpler models (see above). Thus, we can use partial R^2 in order to have identical indices of effect size across the four protocols.

Figure 7. Assessment of the correlation between two measures of effect size – partial R^2 and eta²V – for the judgment task (1x1)

Assessing the statistical significance of effects in regression models

On the one hand, there were three conceptual categories – size, biological class and repulsiveness – that were associated with two phonetic contrasts, one between two vowels, and one between two consonants, as the result of the specific instantiation of our general hypotheses. For each conceptual category, the first hypothesis was related to the consonantal contrast, with the two vowels occurring as the possible contexts, and the second hypothesis was related to the vocalic contrast, with the two consonants occurring as the possible contexts. For example, for size, the first hypothesis opposes [p] to [b] with [i] and [a] as contexts, the second hypothesis opposes [i] to [a] with [p] and [b] as contexts. In each regression model, the context was accounted for along with the target contrast.

On the other hand, one conceptual category, dangerousness, was associated with four phonetic contrasts:

- one between two vowels, [i] versus [u], with [g] and [m] as consonantal contexts (vocalic contrast);
- one between two consonants, [p] versus [k], with only [u] as a vocalic context (place of articulation contrast);
- one between two pairs of consonants, [b] and [g] versus [p] and [k], with only [u] as a vocalic context – (voicing contrast);

- one between two pairs of consonants, [d] and [b] versus [l] and [m], with [u] and [i] as vocalic contexts – (manner contrast).

When two vowels were used as contexts, they were accounted for in the model as confounding factors, as previously for the three other conceptual categories. Additionally, pairs of consonants were opposed rather than single consonants and the model also included the consonantal feature which distinguished the consonants in the pairs. For example, when assessing plosive versus sonorant consonants with the four pseudo-words [idid], [ilil], [ubub] and [umum], we accounted for the two contextual vowels [i] and [u], and also for the two possible contextual places of articulation for the consonants (bilabial for [m] and [b], dental for [d] and [l]).

For the sake of simplicity, we only report in the following tables the results related to the target phonetic contrast for each hypothesis, and not what relates to the contextual confounding factors, except in the case of interactions between them, as explained below.

Binomial regression models were used in 1x2, 2x1 and 2x2 and predicted a label (e.g. 'a large animal' versus 'a small animal') according to the target phonetic contrast, its context of occurrence and their interaction (the context can actually be constituted of two factors in the case of dangerousness, as seen above). A type-III Anova was conducted to reveal significant predictors.

If the interaction between the target phonetic contrast and its context was significant, simple effects (contrasts between marginal means) were assessed. In the case of an absence of interaction, another binomial regression without the interaction term was conducted to reveal the main effects of the target phonetic contrast and its context (because partial R^2 could not be obtained for the main effects in the presence of an interaction term).

As for 1x1, triple interactions had to be considered since the answer was not a choice between two labels (the predicted variable in other models) but rather a judgment *according to* a label, a target phonetic contrast *and* a context. A main effect (of either the target phonetic contrast or the context) in other protocols is here a double interaction (between the label and either the target phonetic contrast or the context). A double interaction between the target phonetic contrast and the context in the other protocols is here a triple interaction (between a label, the target phonetic contrast and the context). Once again, a type-III Anova assessed the potentially significant triple interactions. Only one was significant, and related to the [i]-[u] contrast for dangerousness, which interacted with the labels and the consonantal context [g]/[m]. For the other hypotheses, the triple interaction term was dropped to assess the significance of the double interactions.

Results by protocol and tested hypothesis

For each phonetic contrast in each conceptual contrast, i.e. for each of our sound symbolic hypotheses, the effect size and statistical significance are presented in Table 5 for either the main effect being studied or its interaction with its context, when this interaction is statistically significant). As explained previously, while only the result of the tested hypothesis is reported, the effect of its context of occurrence is nonetheless always accounted for. In the case of a significant interaction between them, simple rather than main effects are relevant, which is why only the interaction is reported. The analyses of both interactions (see Discussion) do not contradict what is found with other protocols.

		1x1		2x1		1x2		2x2	
		R^2	p	R^2	p	R^2	p	R^2	p
Size	[p]-[b]	.012	*p* = .227	.073	***p* = .011**	.097	***p* = .014**	.377	***p* < .001**
	[a]-[i]	.146	***p* = .003**	.247	***p* < .001**	.333	***p* < .001**	.412	***p* < .001**
Class	[t]-[s]	< .001	*p* = .995	< .001	*p* = .803	.108	***p* = .013**	.014	*p* = .243
	[i]-[a]	.035	*p* = .085	< .001	*p* = .989	.012	*p* = .232	.134	***p* = .020**
Repulsiveness	[k]-[n]	.141	***p* = .003**	.082	***p* = .008**	.090	***p* = .018**	.170	***p* = .011**
	[i]-[a]	.035	*p* = .085	.030	*p* = .079			.599	***p* < .001**
	[i]-[a]*[k]-[n]					.089	***p* = .020**		
Danger	[u]-[i]			.046	***p* = .036**	.149	***p* = .004**	.039	*p* = .138
	[u]-[i]*[g]-[m]	.110	***p* = .010**						
	[b,g]-[p,k]	.117	***p* = .007**	< .001	*p* = .803	< .001	*p* = .675	.083	*p* = .057
	[d,b]-[l,m]	.009	*p* = .232	.015	*p* = .079	.133	***p* = .003**	.276	***p* < .001**
	[p]-[k]	< .001	*p* = .594	.169	***p* = .006**	.369	***p* = .001**	< .001	*p* = .601

Table 5. Statistical results per conceptual contrast (rows) and per protocol (columns) for the ten hypotheses under study. For each effect, an effect size (partial R^2) and a p-value are reported. If there was a significant interaction effect between the target contrast and its context, it is presented with a ''. Significant effects (p < .05) are reported in bold.*

Three main observations can be highlighted: first, two sound symbolic associations were significant in all four protocols: vowels in size contrasts ([i] associated with 'small', [a] associated with 'large') and consonants in repulsiveness contrasts ([k] associated with 'repulsive', [n] associated with 'attractive'). Second, every tested hypothesis was significant at least once across protocols. Third, and importantly, the tests for the different protocols never contradicted each other with respect to the orientation of the associations made by the subjects, and the significant associations were always in line with the hypotheses (not shown here, but see the analyses per conceptual contrast below). The four different protocols therefore point to the same direction, but, however, differ quite significantly from each other.

Comparisons between protocols based on effect sizes

In order to assess the congruency between the four protocols, Spearman's rho (r_s) correlation tests were first computed between the effect sizes of all six possible pairs of protocols (see Figure 8). This was made possible by the shared measure of effect size, i.e. the partial R^2. The results show that only one of the six correlations is significant, more specifically between 2x1 and 1x2 (p < .001, not corrected for multiple comparisons). These correlation tests demonstrate that, in general, the different protocols do not lead to similar results. Additionally, the only significant correlation effect might be explained by the fact that 2x1 and 1x2 both rest on a single within-trial contrast, either phonetic or conceptual.

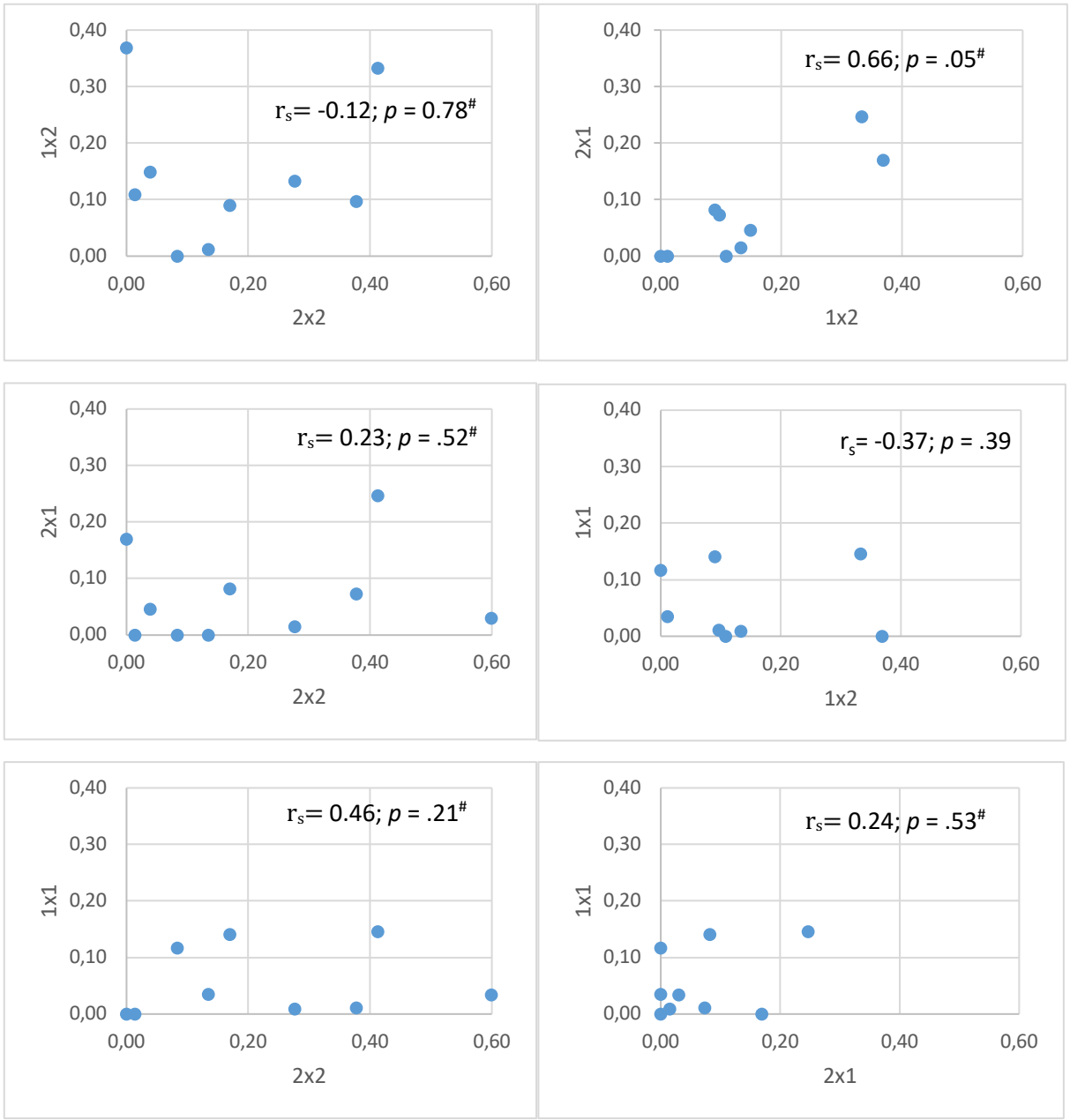


Figure 8. Spearman's rho correlation tests between effect sizes of each protocol. Each dot corresponds to a sound symbolic hypothesis. The number of dots depend on the common statistical tests between protocols, as they are presented in table X. For example, repulsiveness contrasts in 1x2 present no vocalic main effect but an interaction between vowels and consonants, which does not appear in other protocols. Hence, the related hypothesis is absent for correlations implying 1x2.
 # means that equal ranks impact on the estimation of the p-values when computing Spearman's rho.

Overall, the 2x2 protocol induced stronger effect sizes (mean: 0.21) than the 1x2 (mean: 0.14). 2x1 and 1x1 correspond to the weakest effect sizes (both means: 0.06). As a result, we can hypothesize that within-trial phonetic contrasts enhance or sometimes reveal sound symbolic associations, while this effect is further enhanced when there is an additional within-trial conceptual contrast.

Recognition task following 1x1

Several analyses were computed for the second part of the 1x1 protocol, which consisted in a recognition task. First, we tested whether recognition of a pseudo-word depends on the previous evaluation of its adequacy with a label (i.e. the higher the initial judgement of congruence, the better the recognition). Second, we tested the impact of the 'sound symbolic congruence' (according to our hypotheses) between the target pseudo-word and the label it was presented with in the first part of

the protocol, regardless of the subjects' judgments. Neither was conclusive. However, the weak number of misrecognitions may be insufficient for analyses, which may be explained by the low difficulty of the task. Only 21% of the pseudo-words that were heard during the first stage were not recognized during the second stage. This percentage of incorrect answers may seem sufficient for the analyses, but it is not given the number of participants and the allotment of these incorrect answers in specific conceptual contrasts. For example, there were only three pseudo-words out of 31 that were presented with the label 'repulsive' and that were not recognized (two [inin] and one [ikik]). This is not enough to obtain satisfying analyses and to reach conclusions about the putative impact of sound symbolism on recognition.

Analyses by conceptual contrast

In the 'Results by protocol and tested hypothesis' section, we reported statistically significant associations and their respective effect sizes, but not the orientations of these associations, which are not revealed by effect sizes and *p*-values. This section provides more details about the sound symbolic associations as they appear by conceptual contrast across the different protocols.

The effect sizes and *p*-values presented in the following tables are the same as those presented in Table 5. An additional statistical assessment of the different simple effects could have told us more about the associations, e.g. is [i] associated more with 'small' than with 'large', or more with 'small' than [a]? However, this would have led to 240 tests for these simple effects, besides the 40 *p*-values presented in Table 5 and other *p*-values calculated when there were interactions between consonants and vowels. Performing multiple tests increases the possibility of Type-I errors (false positives), but correcting the familywise error rate for such a high number of *p*-values would have likely led to many Type-II errors (false negatives) – see (Feise, 2002) for a discussion about this issue. For these reasons, we decided not to assess simple effects statistically but to report propensities based on regressions that were significant and on what we could see in the contingency tables (except for the simple effects in the case of the two significant interactions reported in Table 5, which are analyzed more thoroughly in Supplementary Information). One needs to remember here that we have, however, applied corrections for main effects and interactions. This explains why for the *p*-values reported below, corrections are sometimes mentioned (for interactions and main effects).

Size

In 2x2, 1x2 and 2x1 for size contrasts, the effects of vowels and consonants are clear: there are associations between [a], [b] and 'large' and between [i], [p] and 'small' (cf. Table 6). In the 1x1 protocol, the effect of the consonants does not appear and the effect of vowels is weaker than those in other protocols. In 2x1 and 1x1, patterns of the responses are similar: [p] and [i] are 'small'; [a] is more 'large' than 'small', but it is not as clear-cut as in the other protocols; [b] is neither 'large' nor 'small'.

	Vowels				Consonants			
		Large	Small			Large	Small	
2x2	**[a]**	9	6	***p* < .001**	**[b]**	12	8	***p* < .001**
	[i]	0	27	**R^2 = .412**	**[p]**	1	19	**R^2 = .377**
1x2	**[a]**	22	4	***p* < .001**	**[b]**	18	7	***p* = .014**
	[i]	7	25	**R^2 = .333**	**[p]**	14	23	**R^2 = .097**
2x1	**[a]**	30	17	***p* < .001**	**[b]**	24	23	***p* = .011**
	[i]	7	40	**R^2 = .247**	**[p]**	13	34	**R^2 = .073**

1x1	**[a]**	6,41	5,06	***p* = .003**	**[b]**	5,53	5,38	*p* = .227
	[i]	3,82	6,73	**R^2 = .146**	**[p]**	4,71	6,4	R^2 = .012

Table 6. Contingency tables per vowel and consonant contrast for contrasts of size and for the four protocols, with their respective p-values and partial R^2. The 1x1 tables show averaged judgments on a 0-to-10 scale. Significant effects (p < .05) are reported in bold.

Biological class

The results about biological classes (cf. Table 7) are difficult to interpret. There is an effect of vowels in 2x2 and an effect of consonants in 1x2. In 2x1 and 1x1, there is a bias in favor of 'bird': generally speaking, they are more often chosen or judged as more fitting with the presented pseudo-words. This bias cannot occur in 1x2, since there are as many answers for 'bird' as for 'fish', while it could have occurred in 2x2 and did not. This preferential bias may complicate occurrences of sound symbolic associations (with fewer answers, less associations may be revealed). However, the absence of a vocalic effect in 1x2 does not support this idea that the preferential bias is the reason why sound symbolic associations do not appear more clearly. Moreover, 'fish' and 'bird' do not intrinsically oppose each other, and we could therefore have expected more sound symbolic associations in 2x2 and 2x1, where the conceptual contrast is explicit to subjects, than in 1x2 and 1x1; however, this was not the case.

		Vowels				**Consonants**		
		Bird	**Fish**			**Bird**	**Fish**	
2x2	**[a]**	4	10	***p* = .020**	**[s]**	11	12	*p* = .243
	[i]	17	7	**R^2 = .134**	**[t]**	12	7	R^2 = .014
1x2	**[a]**	8	13	*p* = .232	**[s]**	13	23	***p* = .013**
	[i]	23	18	R^2 = .012	**[t]**	16	6	**R^2 = .108**
2x1	**[a]**	32	12	*p* = .989	**[s]**	32	13	*p* = .803
	[i]	34	13	R^2 < .001	**[t]**	34	12	R^2 < .001
1x1	**[a]**	4,76	3,82	*p* = .085	**[s]**	5,82	3,63	*p* = .995
	[i]	7,44	3,81	R^2 = .035	**[t]**	6,44	4	R^2 < .001

Table 7. Contingency tables per vowel and consonant contrast for contrasts of biological class and for the four protocols, with, their respective p-values and partial R^2. The 1x1 tables show averaged judgments on a 0-to-10 scale. Significant effects (p < .05) are reported in bold.

Repulsiveness

In repulsiveness contrasts (cf. Table 8), there is a clear effect of consonants, that is present in all protocols, irrespective of consonants being contrasted within trials (2x2, 1x2) or not (2x1 and 1x1). There is also an interaction between consonants and vowels in vocalic contrasts in 1x2, which is reported in Figure 9 (simple effects are presented in Supplementary Information). Besides, an effect of vowels appears in vocalic contrasts in 2x2. Hence, overall, consonants have a stronger impact than vowels on choices regarding repulsiveness. Moreover, the impact of the latter may depend on the former, as in 1x2: the vocalic effect is stronger with [k] than with [n]. The same interaction is nearly significant in 2x1 after

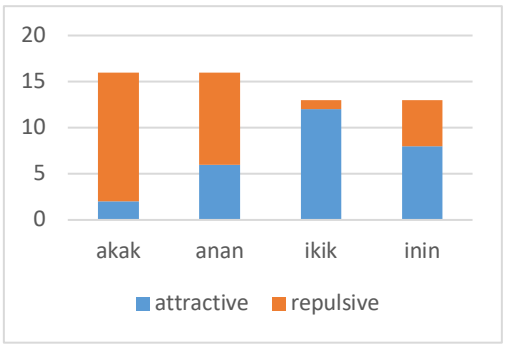


Figure 9. Repartition of answers for the contrast of repulsiveness in 1x2, in which an interaction effect appears: the vocalic impact is stronger with the consonant [k] in comparison with [n].

correction (p = .06). It thus seems that when a pseudo-word is imposed, i.e. no choice is given between two pseudo-words, the participants lean *exclusively* on consonants for choosing a label (2x1) or making a judgment (1x1). When a label is imposed, i.e. no choice is given between two labels, and a choice of pseudo-word has to be made, the participants lean *more* on consonants than on vowels (1x2). Finally, in 2x2, participants rely on the presented vocalic or consonantal contrasts, whatever the context for these contrasts.

	Vowels				Consonants			
		Attractive	Repulsive			Attractive	Repulsive	
2x2	**[a]**	2	20	***p* < .001**	**[k]**	8	15	***p* = .011**
	[i]	18	1	**R^2 = .599**	**[n]**	12	3	**R^2 = .170**
1x2	**[a]**	*8*	*24*	***p* = .020**	**[k]**	9	20	***p* = .018**
	[i]	*20*	*6*	**R^2 = .089**	**[n]**	21	12	**R^2 = .090**
2x1	**[a]**	14	32	*p* = .079	**[k]**	12	35	***p* = .008**
	[i]	23	23	R^2 = .030	**[n]**	25	20	**R^2 = .082**
1x1	**[a]**	4.56	5.41	*p* = .085	**[k]**	4.33	6.06	***p* = .003**
	[i]	6.22	4.75	R^2 = .035	**[n]**	6.44	4.06	**R^2 = .141**

Table 8. Contingency tables per vowel and consonant contrasts for contrasts of repulsiveness and for the four protocols, with, their respective p-values and partial R^2. The 1x1 tables are constituted of averaged judgments on a 0-to-10 scale. Significant effects (p < .05) are reported in bold.

Dangerousness

Vowels:

The vocalic effect (cf. Table 9) does appear in 2x1 and 1x2, in 1x1 in interaction with the consonants used as context, but not in 2x2. The absence of significant effect in this latter condition may be due to an intrinsic effect of the consonantal context made of [g] and [m]. This is suggested first by an effect of the consonantal context in 2x2 (p = .02, R^2: 0.13) and in 2x1 (p = .03 without correction, R^2: 0.05), although the *p*-values are not (and cannot easily be) corrected. Moreover, the significant effect in 1x1 is an interaction between vowels and contextual consonants, as presented in Figure 10 (simple effects are presented in Supplementary Information). Overall, one may thus hypothesize that an effect of consonants may mask here the effect of vowels.

	Vowels			
		Dangerous	Harmless	
2x2	**[i]**	5	15	*p* = .138
	[u]	11	10	R^2 = .039
1x2	**[i]**	5	17	**p = .004**
	[u]	25	13	**R^2 = .149**
2x1	**[i]**	7	40	**p = .036**
	[u]	16	30	**R^2 = .046**
1x1	**[i]**	*3.87*	*7.38*	**p = *.010***
	[u]	*5.22*	*5.79*	***R^2 = .110***

Table 9. Contingency tables per vocalic contrasts for contrasts of dangerousness and for the four protocols, with, their respective p-values and partial R^2. The 1x1 tables are constituted of averaged judgments on a 0-to-10 scale. Significant effects ($p < .05$) are reported in bold. The significant interaction is in italic.

More specifically, the interaction that appears between consonants and vowels in vocalic contrasts is due to opposed preferential associations, on the one hand between the pseudo-word [ugug] and 'dangerous', and on the other hand between the pseudo-words [igig], [imim] and [umum] and 'harmless'. The complementary pairs (e.g. [ugug] with 'harmless') all result in weaker judgements. Hence, the impact of a vowel or a consonant depends on the consonantal or the vocalic context, respectively, in which it is presented. This may explain the absence of effects in 1x1 for place of articulation and for manner (see below): in a 1x1 context, an interaction between consonants and vowels is indeed critical.

Place and voicing:

There are strong effects of the place of articulation (cf. Table 10) – involving front versus back consonants – in 2x1 and 1x2. The fact that it does not appear in 2x2 may be due to the fact that only half of the participants where shown the contrast [upup]-[ukuk] – the other half was shown [ukuk] in a voicing contrast with [ugug]. Judgments for [ukuk] with 'harmless' animals in 1x1 are on average quite high (7.22), which is surprising since it is not associated with 'harmless' in 1x2 (14%), 2x1 (15%) and in 2x2 (24%) – 25% being chance level. However, apart from being a back consonant, [k] is also voiceless and voiceless consonants ([upup) and [ukuk]) are judged as better suited to 'harmless' animals (6.25) than to 'dangerous' ones (3.59) in voicing contrasts in 1x1. Half of the answers contained in the 'voicing' contingency table (cf. Table 10) are the same answers as those contained in the 'place' contingency table ([upup] and [ukuk]). Here appears the limit of a 1x1 protocol with only a few segments tested. It is surprising that the effect between danger and voicing only appears in 1x1, since overall 1x1 is associated with weaker effect sizes and less significant effects than other protocols.

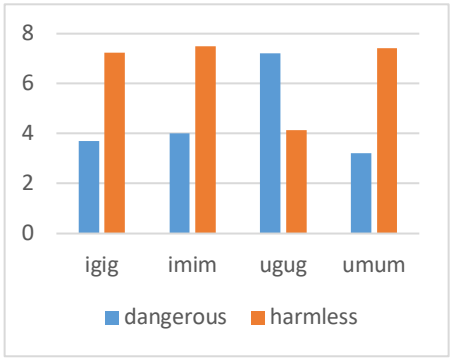


Figure 10. Mean judgments in vocalic contrasts in 1x1, in which an interaction effect appears: the co-occurrence of [g] and [u] leads to a pattern of judgments that is opposite to the other C-V co-occurrences.

		Place				Voicing		
		Dangerous	Harmless			Dangerous	Harmless	
2x2	**[p]**	4	5	p = .601	**[b] & [g]**	13	3	p = .057
	[k]	7	5	R^2 < .001	**[p] & [k]**	12	12	R^2 = .083
1x2	**[p]**	1	10	***p* = .001**	**[b] & [g]**	18	16	p = .675
	[k]	14	4	**R^2 = .369**	**[p] & [k]**	13	15	R^2 < .001
2x1	**[p]**	6	17	***p* = .006**	**[b] & [g]**	21	25	p = .803
	[k]	17	7	**R^2 = .169**	**[p] & [k]**	23	24	R^2 < .001
1x1	**[p]**	2.38	5.00	p = .594	**[b] & [g]**	**5.94**	**5.31**	***p* = .007**
	[k]	4.67	7.22	R^2 < .001	**[p] & [k]**	**3.59**	**6.25**	**R^2 = .117**

Table 10. Contingency tables per place and voicing contrasts for contrasts of dangerousness and for the four protocols, with, their respective p-values and partial R^2. The 1x1 tables are constituted of averaged judgments on a 0-to-10 scale. Significant effects ($p < .05$) are reported in bold.

Manner:

The phonetic contrast seems essential in sound symbolic associations involving manner, since the effects appear only in 2x2 and 1x2 (cf. Table 11). There seems to be preferential biases for 'harmless' animals in 2x1 and 1x1, just as there was a preferential bias for 'bird' when investigating biological class.

		Manner		
		Dangerous	**Harmless**	
2x2	**[b] & [d]**	**17**	**10**	***p* < .001**
	[m] & [l]	**3**	**30**	**R^2 = .276**
1x2	**[b] & [d]**	**24**	**10**	***p* = .003**
	[m] & [l]	**13**	**28**	**R^2 = .133**
2x1	**[b] & [d]**	15	29	*p* = .079
	[m] & [l]	10	37	R^2 = .015
1x1	**[b] & [d]**	4.00	6.53	*p* = .232
	[m] & [l]	2.89	6.56	R^2 = .009

Table 11. Contingency tables per manner contrasts in size contrasts, depending on protocols, and their respective p-values and R^2. The 1x1 tables are constituted of averaged judgments on a scale between 0 and 10. Significant effects (p < .05) are reported in bold.

Response times

We had no hypotheses about response times according to the four different protocols. They may nevertheless be informative (see table 12). One might have anticipated that the protocol leading to the longest response times would be 2x2 because it involves dealing with two contrasts, which must both be extracted processed before producing two answers rather than 1. This was, however, not the case. At the same time, it is not surprising that 1x1 is actually the 'slowest' protocol, since choosing an answer among 11 possible ones on a scale was likely demanding. The two protocols leading to the shortest response times were the ones that presented a within-trial phonetic contrast (1x2 and 2x2). They were also the ones that presented the highest effect sizes and the largest number of significant associations.

	1x2	**2x2**	**2x1**	**1x1**
Mean response times	768 ms	877 ms	1164 ms	1971 ms
Standard deviation of response times	529 ms	439 ms	526 ms	688 ms
Number of significant hypotheses	8	6	5	4
Mean effect size across hypotheses	.14	.21	.06	.06

Table 12. Response times, number of significant associations and average effect sizes per protocol.

These data may highlight a correlation between sound symbolic association patterns and response times. However, the differences between protocols in how trials were presented may explain the differences in response times. The recording of response times started at the end of the presentation of a trial, whose duration was similar across protocols. Nevertheless, when there was a within-trial phonetic contrast, one pseudo-word was heard before the second. It is possible that some participants began to make their choice after the oral presentation of the first pseudo-word, which added supplementary time for the decision (1600 ms approximately), that is between the end of the first pseudo-word and the end of the presentation of the trial. To assess this hypothesis, we checked how many times the first pseudo-word and the second were chosen in 2x2 and 1x2. The first pseudo-word

was chosen 45% of the time in 2x2 and 50% in 1x2. Hence, it seems that faster response times in these protocols are not explained by an order bias.

Discussion

General observations

Our experiment and its four protocols provide information about the impact of both phonetic and conceptual contrasts. One should, however, be reminded that the following arguments are based on a limited number of consonants and vowels, and that larger-scale assessments could potentially lead to partly different conclusions.

The first thing that stands out is the heterogeneity of the results, in terms of both statistical significance and effect size of the associations. Sound symbolic associations involving size, repulsiveness, dangerousness and biological class, were indeed differently highlighted across the different protocols. According to effect sizes, on average, 2x2 leads to the strongest associations, followed by 1x2, and then, at the same level, by 2x1 and 1x1. On average, effect sizes are low compared to what is found in some other studies, but they are not obtained with repetitive tasks, for example when several spiky and round shapes are associated multiple times with different pseudo-words. Indeed, in such tasks, learning throughout the task, or at least consistent behaviors in subjects, may strengthen the associations.

Second, the consistency across paradigms differs in accordance with the conceptual contrasts. When it comes to associating or judging labels about size and repulsiveness, answers are quite consistent (with stronger impacts of vowels for size and consonants for repulsiveness). On the contrary, associations and judgements for biological class are much less consistent, and it is difficult to draw firm conclusions. We observe in particular a preferential bias for ‘bird’ in 2x1 and 1x1: ‘bird’ is chosen more often than ‘fish’ in the former, and judgments involving ‘bird’ are higher on the Likert scale than those involving ‘fish’ in the latter. When ‘fish’ is the only concept to be presented – preventing the preferential bias to occur – and a choice has to be made between two pseudo-words (1x2), an effect of consonants appears, with [s] being preferred over [t]. More broadly speaking, a single phonetic contrast (1x2) prevents such preferential biases (1x2), and two contrasts (2x2) may palliate them, since there is also no preferential bias in the case of biological class in 2x2. Thus, phonetic contrasts may be more appropriate with materials that are unbalanced in terms of preferential choice at the conceptual level. Last, danger contrasts are also hard to interpret since there are four different phonetic contrasts (vowels, voicing, place and manner), using different consonantal or vocalic contexts. There are nonetheless some interesting results, such as the necessity to contrast the phonetic feature of manner in order to reveal associations with a ‘harmless’ or ‘dangerous’ animal – since a significant association only appears in 2x2 and 1x2.

The 1x1 protocol, whatever the (between-trial) conceptual contrast, may confirm some associations that are presented in other protocols. Its limited sensitivity suggests, however, dedicating this protocol to strong sound symbolic associations, such as size and repulsiveness.

Overall, the discrepancy between protocols we report here points to the necessity of taking into account the paradigm according to which some associations are revealed in the literature. Indeed, the presence or absence of phonetic and conceptual contrasts within a trial may provoke differences in the experimental emergence of sound symbolic associations.

Differences in effect size induced by differences in the presentation of contrasts

Mean values of effect size suggest greater influence of a phonemic contrast over a conceptual one, since 2x2 and 1x2 present higher means in comparison with 2x1 and 1x1. However, one can argue that the higher means and higher numbers of significant associations in 2x2 and 1x2 are induced by the presence of both contrasts in both protocols, and not only in 2x2. Indeed, in 1x2, despite the focus on a phonemic contrast, the conceptual one is in some way also present: 'a small animal' is indeed inherently linked to 'a large animal'. As a result, we must be cautious when concluding as to the causes of the higher amount of significant associations and of their higher average strength in these protocols. This may indeed result either only from the phonetic within-trial contrast, or from the presence of both contrasts, even if the conceptual one is only 'derivatively' present (at least for size, dangerousness and repulsiveness). Along similar lines, 1x1 – where there was no within-trial conceptual contrast but where this contrast could nevertheless be easily intuited (except for biological class) – possibly differed from 1x2 – where there was a within-trial phonetic contrast – only on the basis of the different cognitive operations required: a judgment and an association, respectively. The weaker effects in 1x1 may thus be due to the judgment task itself, and not to the absence of contrasts.

The four protocols investigated in our experiment thus exhibit different sensitivities to sound symbolism. This being said, the congruence observed in the results – no protocol goes against the others in terms of what linguistic stimuli are associated to the various concepts – suggests that differences in which associations appear to be significant derive from differences in attention paid to the stimuli and to their features because of the different contrastive presentations, not from fundamentally different cognitive processes.

The 1x1 is the least 'efficient' paradigm and one potential explanation for its weakness is the possibility offered to participants to remain neutral in front of an associations. Indeed, the scale to express judgments ranged from 0 to 10, and 5 therefore indicated a neutral judgment, if not an absence of judgment. However, the repartition of answers reveals that 5 was not participants' default choice (9% of all answers) (see Figure 11). The most frequent answers were 2 (15%), 3 (11%), 7 (13%) and 8 (13%), while the extreme answers, 0 and 10, were chosen the least. Therefore, the 'lack of efficiency' of 1x1 may not be due to participants refusing to take side in their answers.

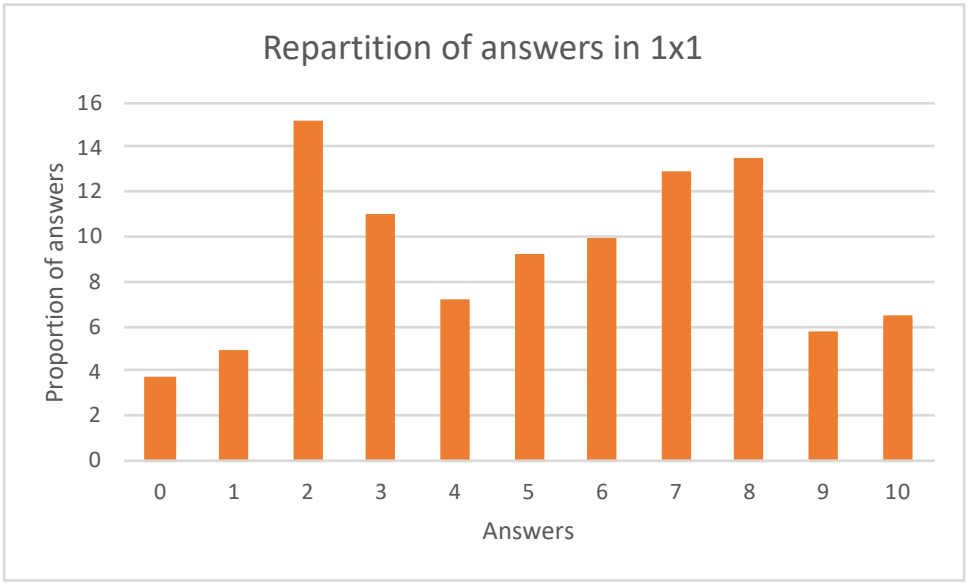


Figure 11. Proportion of answers per response in the judgment task (1x1).

Vowels and consonants

Whether sound symbolism is mostly driven by consonants or vowels is controversial. In the relevant literature, some authors conclude in favor of the latter (Tarte, 1974; Knoeferle et al. 2017) and others in favor of the former (Nielsen & Rendall, 2011; Aveyard, 2012; Fort et al. 2015). Our results shed new light on this open issue, since we observe that dominance of one category over the other is protocol-dependent. For instance, in the case of size contrasts (see Table 13), vowels have overall more influence than consonants (except for 2x2, for which influences are approximately equal). Such a statement is however dependent on our specific choice of consonants and vowels. Hence, relying for example on voiced or voiceless palatal (e.g. [k]-[g]) rather than on voiced or voiceless bilabial ([p]-[b]) consonants could perhaps have led to different consonantal effects.

In parallel, our results revealed potentially different effects depending on paradigms. Indeed, two interactions appeared between consonants and vowels in our results, specifically in contrasts of dangerousness in 1x1, and in contrasts of repulsiveness in 1x2. In both cases, the effect of the vowels under study differed according to the consonantal context, but one could expect the opposite pattern to also occur, i.e. the influence of a vocalic context on the effects of consonants. Between and within-trials contrasts may therefore reveal different patterns of interactions between consonants and vowels. This, however, demands further investigation.

		1x1	2x1	1x2	2x2
Size	[p]-[b]	.012	.073	.097	.377
	[a]-[i]	.146	.247	.333	.412

Table 13. Effect sizes (R^2) in size contrasts for vowel and consonant contrasts.

Conclusion

In order to better understand and assess previous results in the literature on sound symbolism, we have carried a comparative investigation of different experimental settings involving various types of association tasks and a judgment task. Although it appears that the results from these different protocols never contradict each other and overall support hypotheses on sound symbolism found in the literature, considerable differences are observed in terms of significant effects, i.e. different protocols highlight different hypotheses. The 2x2 association task may be here considered as the most 'efficient', in the sense that it confirmed the highest number of hypotheses. This may be related to the explicit presentation of both phonetic and conceptual contrasts within each trial, and at the cognitive level to higher-level strategies from the subjects. The 1x2 association task, with a within-trial phonetic contrast only, came second, suggesting that explicit phonetic contrasts are pushing subjects toward making sound symbolic associations. For three out of four conceptual categories (size, dangerousness, repulsiveness), each pole of the conceptual contrast likely activates the other one in the subjects' minds (e.g. 'small' activates 'large', 'dangerous' activate 'harmless'), and the influence of this implicit conceptual contrast thus cannot be ruled out. However, one association about biological class is significant in 1x2 and, given this protocol, cannot be explained by the implicit activation of the opposite association. It seems thus possible to highlight sound symbolic associations without conceptual contrast. On the contrary, the 2x1 association task confirms few hypotheses and suggests that some preferential biases regarding the concepts (e.g. 'bird' is more chosen over 'fish') may mask some associations. Finally, the 1x1 judgment task leads mostly to non-significant effects.

Overall, on the basis of our results, we recommend caution when comparing in depths the results of studies based on different protocols. We also argue that the divergences between studies as for the sound symbolism of consonants and vowels may mostly stem from differences in protocols. Our results

finally point at the underlying cognitive mechanisms possibly explaining the differences between protocols, although more work is definitely needed to better understand the respective and possibly complementary roles of conceptual and phonetic contrasts.

Acknowledgments

The authors are grateful to the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program Investissements d'Avenir (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). More specifically, the first author benefited from a 3-year doctoral scholarship from the LABEX ASLAN. There was no specific funding for the collection, analysis and interpretation of data, nor for the writing of the report. The sponsor also did not play any role in the decision to submit the article for publication.

The authors also thank colleagues for their help to collect data during the European Researchers' Night on September 30, 2016.

Declaration of interest

The authors do not have any competing interest to declare.

Appendix A. Pseudo-words

		Sample 1		Sample 2 Recognition test (1x1 part 2)	
		Duration (ms)	f0 (Hz)	Duration (ms)	f0 (Hz)
PW used in associative and judgement tasks	[abab]	554	114	439	119
	[apap]	700	127	519	125
	[ibib]	683	118	494	141
	[ipip]	518	120	582	174
	[asas]	661	125	584	133
	[atat]	651	127	531	120
	[isis]	658	123	620	192
	[itit]	642	128	590	147
	[akak]	664	120	491	158
	[anan]	617	117	491	116
	[inin]	674	125	566	123
	[ikik]	692	118	659	167
	[ibib]	683	118	494	141
	[idid]	551	117	522	127
	[igig]	736	118	564	132
	[ilil]	598	114	508	134
	[ipip]	518	120	582	174
	[imim]	641	125	669	139
	[ubub]	609	119	539	134
	[ugug]	610	114	640	135
	[ukuk]	555	126	742	237
	[umum]	646	123	517	131
	[upup]	698	160	604	173
PW used in the	[yzyz]	622	115		
	[usus]	613	123		
	[ypyp]	658	126		

training part	[adad]	611	115		
	[agag]	563	119		
	[udud]	651	119		
	[ifif]	675	128		
	[aʁaʁ]	637	119		
PW added in the recognition task	[afaf]			520	135
	[uʃuʃ]			567	233
	[yʁyʁ]			492	149
	[yvyv]			561	134
mean		634	122	562	150
SD (Pearson)		0,054	8.11	0.067	30.50

Table 1. F0 and duration times of each pseudo-word according to two different recordings of samples. The first one was used for associative and judgement tasks. The second one was used for the recognition test following the 1x1 protocol and does not contain training trials; four pseudo-words were also added in this recognition task, which explains the presence of grey cells. PW stands for pseudo-words.

Appendix B. Repartition of pseudo-words

		1x1		1x2		2x1		2x2	
		V1	V2	V1	V2	V1	V2	V1	V2
Size	[abab]	x		x	x	x		x	x
	[apap]		x	x	x		x	x	x
	[ibib]		x	x	x		x	x	x
	[ipip]	x		x	x	x		x	x
Class	[asas]	x		x	x	x		x	x
	[atat]		x	x	x		x	x	x
	[isis]		x	x	x		x	x	x
	[itit]	x		x	x	x		x	x
Repulsiveness	[akak]	x		x	x	x		x	x
	[anan]		x	x	x		x	x	x
	[inin]	x		x	x	x		x	x
	[ikik]		x	x	x		x	x	x
Dangerousness	[ikik]	x				x			
	[ibib]	x				x			
	[idid]		x	x	x		x	x	x
	[igig]		x		x		x		x
	[ilil]	x		x	x	x		x	x
	[ipip]		x				x		
	[imim]		x	x			x	x	
	[ubub]		x	x	x		x	x	x
	[ugug]	x		x	x	x		x	x
	[ukuk]		x	x	x		x	x	x
	[umum]	x		x	x	x		x	x
	[upup]	x		x	x	x		x	x
Number of PW	21	12	12	20	20	12	12	20	20

Table 2. Distribution of pseudo-words in accordance with protocols and versions. PW stands for pseudo-words, V1 for version 1, V2 for version 2.

Appendix C. Contingency tables

Aggregation of the results within each protocol

Contrasts of size are used here for the sake of illustration. Four pseudo-words were used ([abab], [apap], [ibib], [ipip]). In 2x2 (see Figure 1), answers were combined according to the phonetic contrast that was presented (for the [i]-[a] contrast, [abab] and [apap] were combined, as well as [ibib] and [ipip]; for the voiced-voiceless consonant contrast, [abab] and [ibib] were combined, as well as [apap] and [ipip]).

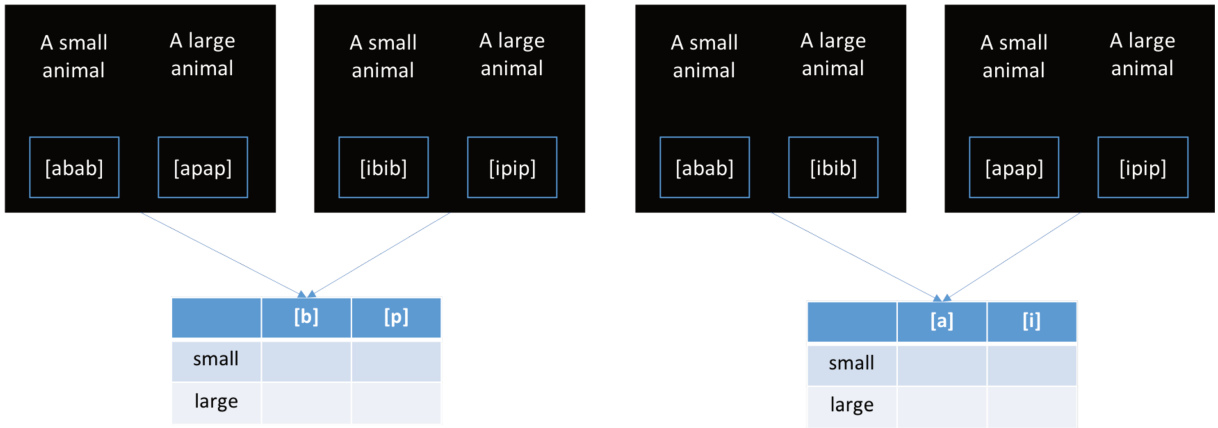


Figure 1. Schematic representations of trials for the 2x2 protocol and resulting contingency tables. The two left representations contain consonantal contrasts; the two right ones contain vocalic contrasts.

In 1x2, the same combinations occurred but there were as many answers for 'small' as for 'large' (see Figure 2).

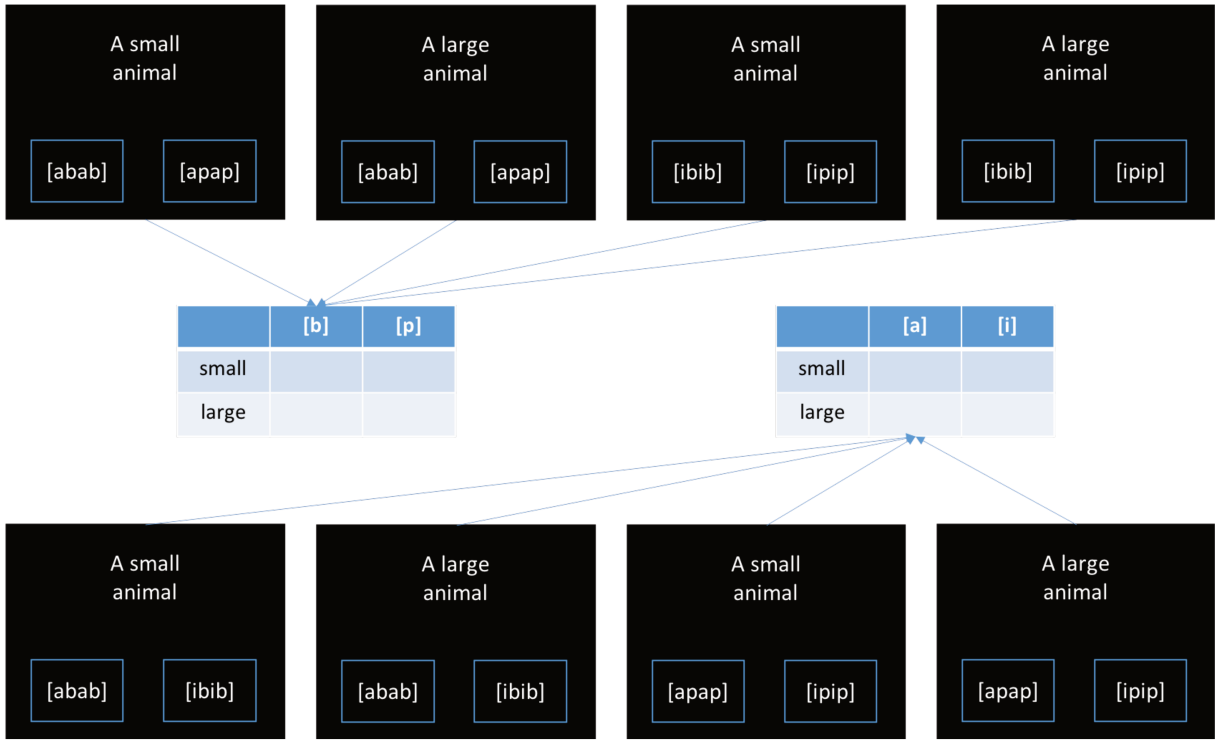


Figure 2. Schematic representations of trials for the 1x2 protocol and resulting contingency tables. Representations at the top contain consonantal contrasts; the one at the bottom contain vocalic contrasts.

In 2x1, we simulated phonetic within-trial contrasts with contrasts between trials in order to produce comparable tables (see Figure 3). We obtained as many answers for the two vowels and the two consonants.

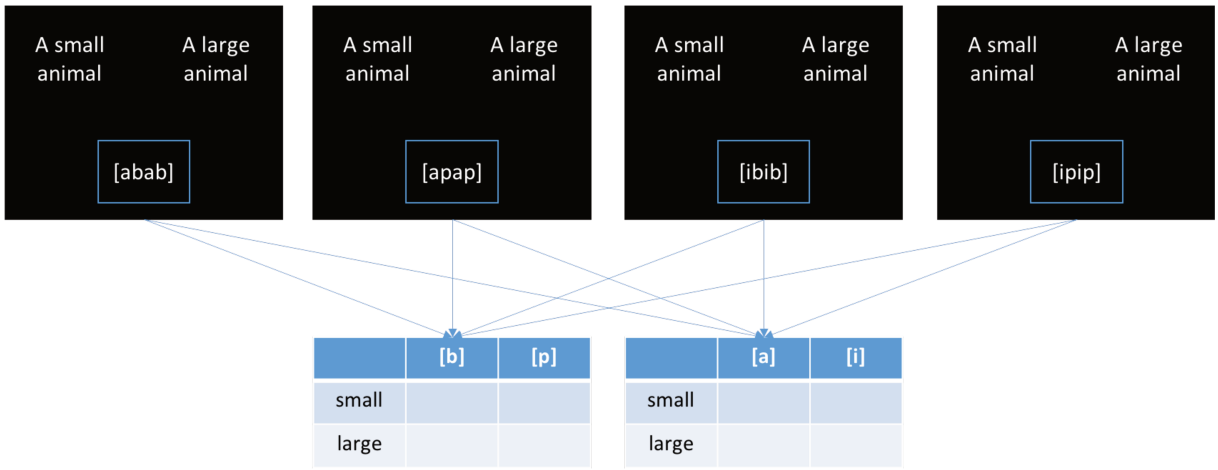


Figure 3. Schematic representations of trials for the 2x1 protocol and resulting contingency tables. Each result of each representation is added to the two contingency tables, according to the consonant and the vowel contained in the pseudo-word.

Finally, for the 1x1 protocol, there was the same number of answers for each combination (each answer on a scale between 0 and 10), and the means of these answers were (see Figure 4).

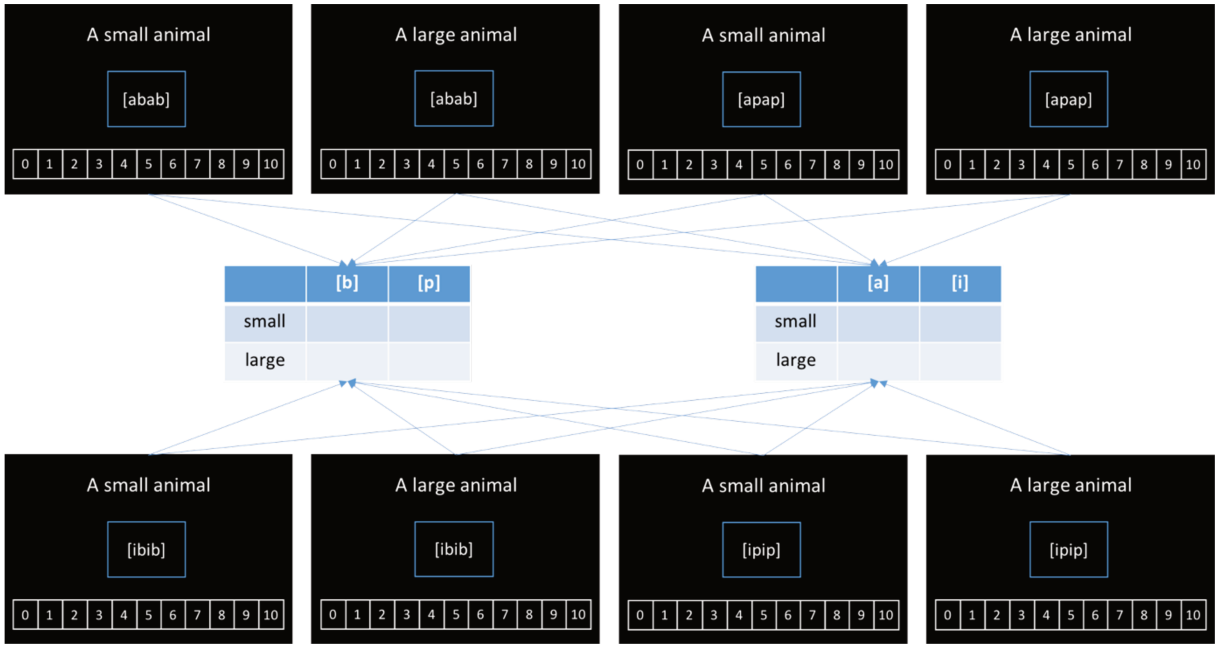


Figure 4. Schematic representations of trials for the 1x1 protocol and resulting tables containing means. Each result of each representation is added to the two contingency tables, according to the consonant and the vowel contained in the pseudo-word.

Appendix D. Simple interaction effects

Simple effects for the two significant interactions reported in the "Results by protocol and tested hypothesis" section are presented in this section. Holm-Bonferroni corrections would have led to less Type-I errors but increase the probability of Type-II errors. Since we judged more interesting here to risk overestimating effects than to potentially miss some, *p*-values have not been corrected.

1x2

When assessing the target vocalic contrast related to repulsiveness, an interaction effect between consonants and vowels was found (partial R^2: 0.09, p = .02). Results show that the vocalic effect is stronger in the context of [k], compared to [n], since [akak] and [ikik] differ significantly, contrary to [anan] and [inin]. Since there is no significant difference between [akak] and [anan], nor between [inin] and [ikik], this vocalic contrast also likely underlies the significant differences between [akak] and [inin], and between [anan] and [ikik].

Contrast		odds.ratio	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
[akak]	[anan]	4.20	3.845	Inf	0.698	25.26	1.568	.117
[akak]	[ikik]	84.00	108.06	Inf	6.750	1045.31	3.444	**< .001**
[akak]	[inin]	11.20	10.60	Inf	1.751	71.64	2.552	**.011**
[anan]	[ikik]	20.00	23.24	Inf	2.051	195.00	2.578	**.010**
[anan]	[inin]	2.67	2.05	Inf	0.591	12.04	1.275	.202
[ikik]	[inin]	0.13	0.16	Inf	0.013	1.37	-1.698	.089

Table 3. Simple interaction effects found when assessing the [i-a] vocalic contrast related to repulsiveness with the 1x2 protocol. P-values smaller than 0.05 are in bold.

1x1

When assessing the target vocalic contrast related to dangerousness, a triple interaction between target vowels, contextual consonants and labels was found (partial R^2: 0.11, p = .004). Judgments differ significantly for each pseudo-word according to the label it was presented with: the pseudo-words [igig] (p = .009), [imim] (p = .008) and [umum] (p < .001) are judged as more fitting with 'harmless' rather than 'dangerous' animals; [ugug] is judged to fit more with 'dangerous'' rather than 'harmless' animals (p = .02). Hence, the combination between [g] and [u] induces stronger associative judgements with 'dangerous' animals, which differs from the other conditions. This being said, the global pattern of associations is difficult to interpret, and providing a full picture of it is partly beyond the target of this article.

Contrast		ratio	SE	df	asymp.LCL	asymp.UCL	z.ratio	p.value
D*[igig]	H*[igig]	0.57	0.122	Inf	0.376	0.87	-2.625	**.009**
D*[igig]	D*[imim]	0.94	0.222	Inf	0.595	1.50	-0.250	.802
D*[igig]	H*[imim]	0.56	0.118	Inf	0.366	0.84	-2.779	**.005**
D*[igig]	D*[ugug]	0.57	0.120	Inf	0.380	0.86	-2.657	**.008**
D*[igig]	H*[ugug]	0.92	0.221	Inf	0.572	1.47	-0.361	.718
D*[igig]	D*[umum]	1.12	0.266	Inf	0.700	1.78	0.463	.643
D*[igig]	H*[umum]	0.56	0.122	Inf	0.365	0.86	-2.673	**.008**
H*[igig]	D*[imim]	1.65	0.331	Inf	1.114	2.44	2.499	**.012**
H*[igig]	H*[imim]	0.97	0.168	Inf	0.692	1.36	-0.173	.863
H*[igig]	D*[ugug]	1.00	0.170	Inf	0.720	1.40	0.020	.984
H*[igig]	H*[ugug]	1.60	0.332	Inf	1.069	2.41	2.281	**.022**
H*[igig]	D*[umum]	1.95	0.398	Inf	1.311	2.91	3.289	**.001**
H*[igig]	H*[umum]	0.98	0.175	Inf	0.689	1.39	-0.120	.905
D*[imim]	H*[imim]	0.59	0.117	Inf	0.398	0.87	-2.663	**.008**
D*[imim]	D*[ugug]	0.61	0.119	Inf	0.414	0.89	-2.535	**.011**
D*[imim]	H*[ugug]	0.97	0.223	Inf	0.620	1.53	-0.123	.902
D*[imim]	D*[umum]	1.18	0.268	Inf	0.760	1.85	0.746	.455
D*[imim]	H*[umum]	0.59	0.122	Inf	0.397	0.89	-2.550	**.011**
H*[imim]	D*[ugug]	1.03	0.174	Inf	0.744	1.44	0.198	.843
H*[imim]	H*[ugug]	1.65	0.341	Inf	1.103	2.48	2.438	**.015**

H*[imim]	D*[umum]	2.01	0.408	Inf	1.354	2.99	3.455	**.001**
H*[imim]	H*[umum]	1.01	0.179	Inf	0.712	1.43	0.047	.962
D*[ugug]	H*[ugug]	1.60	0.325	Inf	1.074	2.38	2.309	**.021**
D*[ugug]	D*[umum]	1.95	0.389	Inf	1.317	2.88	3.340	**< .001**
D*[ugug]	H*[umum]	0.98	0.170	Inf	0.693	1.37	-0.142	.887
H*[ugug]	D*[umum]	1.22	0.283	Inf	0.772	1.92	0.848	.396
H*[ugug]	H*[umum]	0.61	0.129	Inf	0.403	0.92	-2.336	**.020**
D*[umum]	H*[umum]	0.50	0.104	Inf	0.333	0.75	-3.323	**< .001**

Table 4. Simple interaction effects in the 'triple' interaction found when assessing the [i-u] vocalic contrast related to dangerousness with the 1x1 protocol. D stands for 'dangerous', H for 'harmless'. P-values smaller than 0.05 are in bold.

References

Aveyard, M. E. (2012). Some consonants sound curvy: Effects of sound symbolism on object recognition. *Memory & Cognition*, *40*(1), 83–92. https://doi.org/10.3758/s13421-011-0139-3

Berlin, B. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 76–93). New York: Cambridge University Press.

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172. https://doi.org/10.1016/j.cognition.2012.09.007

Chen, Y.-C., Huang, P.-C., Woods, A., & Spence, C. (2016). When "Bouba" equals "Kiki": Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, *6*(May), 26681. https://doi.org/10.1038/srep26681

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding Emotions in Speech with the Size Code. A Perceptual Investigation. *Phonetica*, *65*(4), 210–230.

Cuskley, C., Simmer, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research*. https://doi.org/DOI 10.1007/s00426-015-0709-2

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*(3), 259–268. https://doi.org/10.1111/j.2044-8295.1961.tb00788.x

De Carolis, L., Marsico, E., Arnaud, V., & Coupé, C. (2018). Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm. *PLoS ONE*, *13*(12), e0208874. https://doi.org/10.1371/journal.pone.0208874

De Carolis, L., Marsico, E., & Coupé, C. (2017). Evolutionary roots of sound symbolism. Association tasks of animal properties with phonetic features. *Language and Communication*, *54*, 21–35. https://doi.org/10.1016/j.langcom.2016.10.003

Feise, R. J. (2002). Do multiple outcome measures require p-value adjustment? *BMC Medical Research Methodology*, *2*(July 2002), 1–4. https://doi.org/10.1186/1471-2288-2-8

Fónagy, I. (1961). Communication in poetry. *Word*, *17*, 194–218.

Fónagy, I. (1983). *La vive voix. Essais de psycho-phonétique*. Paris: Payot.

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are More Important than Vowels in the Bouba-kiki Effect. *Language and Speech*, *58*(2), 247–266. https://doi.org/10.1177/0023830914534951

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-05965-y

Köhler, W. (1947). *Gestalt Psychology (2nd Ed.)*. New York: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. https://doi.org/10.1016/j.cognition.2009.08.016

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

Nielsen, A. K. S., & Rendall, D. (2011). The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *65*(2), 115–124. https://doi.org/10.1037/a0022268

Nielsen, A. K. S., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*(2012), 115–125. https://doi.org/10.1515/langcog-2012-0007

Nielsen, A. K. S., & Rendall, D. (2013). Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *67*(2), 153–163. https://doi.org/10.1037/a0030553

Nobile, L. (2015). Phonemes as images. An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In M. K. Hiraga, W. J. Herlofsky, K. Shinoara, & K. Akita (Eds.), *Iconicity: East meets West* (pp. 71–91). Amsterdam: John Benjamins. https://doi.org/10.1075/ill.14.04nob

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

Perfors, A. (2004). What's in a Name ? The effect of sound symbolism on perception of facial attractiveness. *Proceedings of CogSci*, *2*, 2139–2139.

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

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

Sidhu, D. M., & Pexman, P. M. (2017). A Prime Example of the Maluma/Takete Effect? Testing for Sound Symbolic Priming. *Cognitive Science*, *41*(7), 1958–1987. https://doi.org/10.1111/cogs.12438

Tarte, R. D. (1974). Phonetic symbolism in adult native speakers of czech. *Language and Speech*, *17*(1), 87–94. https://doi.org/10.1177/002383097401700109

Turoman, N., & Styles, S. J. (2017). Glyph guessing for 'oo' and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society Open Science*, *4*(9).

https://doi.org/10.1098/rsos.170882

Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M. (2017). Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. *Psychological Research*, *81*(4), 827–839. https://doi.org/10.1007/s00426-016-0778-x

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language*, *93*(1), 10–19. https://doi.org/10.1016/j.bandl.2004.07.006

3. Third study: Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm

Léa De Carolis[1], Egidio Marsico[1], Vincent Arnaud[2] & Christophe Coupé[3]

[1] Laboratoire Dynamique du Langage, CNRS & Université de Lyon, Lyon, France

[2]Département des arts et lettres, Université du Québec à Chicoutimi, Chicoutimi, Canada

[3]Department of Linguistics, The University of Hong Kong, Hong Kong SAR, China

Article published in *PLoS ONE*, 13:12, e0208874, in 2018

PLOS ONE

RESEARCH ARTICLE

Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm

Léa De Carolis[1], Egidio Marsico[1], Vincent Arnaud[2], Christophe Coupé[1,3]*

1 Laboratoire Dynamique du Langage, CNRS & Université de Lyon, Lyon, France, **2** Département des arts et lettres, Université du Québec à Chicoutimi, Chicoutimi, Canada, **3** Department of Linguistics, The University of Hong Kong, Hong Kong SAR, China

* Christophe.Coupe@hku.hk

Citation: De Carolis L, Marsico E, Arnaud V, Coupé C (2018) Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm. PLoS ONE 13(12): e0208874. https://doi.org/10.1371/journal.pone.0208874

Editor: Veronica Whitford, University of Texas at El Paso, UNITED STATES

Received: August 25, 2017

Accepted: November 27, 2018

Published: December 21, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: For their financial support, the authors are grateful to the Université du Québec à Chicoutimi (UQAC), as well as to the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon within the program Investissements d'Avenir (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). Université du Québec à Chicoutimi: https://www.uqac.ca/. Labex

Abstract

Classically, in the bouba-kiki association task, a subject is asked to find the best association between one of two shapes–a round one and a spiky one–and one of two pseudowords–bouba and kiki. Numerous studies report that spiky shapes are associated with kiki, and round shapes with bouba. This task is likely the most prevalent in the study of non-conventional relationships between linguistic forms and meanings, also known as sound symbolism. However, associative tasks are explicit in the sense that they highlight phonetic and visual contrasts and require subjects to establish a crossmodal link between stimuli of different natures. Additionally, recent studies have raised the question whether visual resemblances between the target shapes and the letters explain the pattern of association, at least in literate subjects. In this paper, we report a more implicit testing paradigm of the bouba-kiki effect with the use of a lexical decision task with character strings presented in round or spiky frames. Pseudowords and words are, furthermore, displayed with either an angular or a curvy font to investigate possible graphemic bias. Innovative analyses of response times are performed with GAMLSS models, which offer a large range of possible distributions of error terms, and a generalized Gama distribution is found to be the most appropriate. No sound symbolic effect appears to be significant, but an interaction effect is in particular observed between spiky shapes and angular letters leading to faster response times. We discuss these results with respect to the visual saliency of angular shapes, priming, brain activation, synaesthesia and ideasthesia.

Introduction

Sound symbolism refers to the broad hypothesis that some phonetic units intrinsically carry semantic content. One of the best-known experimental evidences in favor of sound symbolism emergence is the so-called bouba-kiki effect. It consists in the presentation of two shapes, a

ASLAN: http://aslan.universite-lyon.fr/ ANR: http://www.agence-nationale-recherche.fr/.

Competing interests: The authors have declared that no competing interests exist.

curvy and a spiky one, and of two pseudowords, 'bouba' and 'kiki' (or 'maluma' and 'takete' in Köhler [1]'s original experiment). The subject has to select their preferred association between a shape and a pseudoword during a forced-choice task. Many studies show the same pattern of response: 'bouba' is more often associated with round shapes, while 'kiki' is more often associated with spiky shapes. This has been demonstrated with people from different countries and speaking different languages, using different kinds of phonetic and visual materials [2–6]. The effect is also discussed in infants [7]. Rogers and Ross reported no preferential association in the Songe people of New-Guinea [8]. Their study, however, lacks a precise description of the protocol and information such as the number of persons surveyed. Overall, the results suggest that these sound symbolic associations are a near-universal trend in human populations.

In 2005, Westbury [9] shifted from the classical explicit association task to a more implicit paradigm assessing sound symbolism: a lexical decision task where written forms were presented in either spiky or rounded frames. The general purpose of the present study is to extend this experiment. In the section below, four core components of the approach are discussed in the light of recent studies: (1) which phonemes and features get associated with visual shapes; (2) the role played by contrasts in association tasks; (3) the transparency of the tasks used in the field of sound symbolism, and its influence on the strength of associations; (4) the possible influence of the graphemic shape of letters when written forms are part of the experimental setting.

Rationale

Sound symbolism of consonants and vowels

A number of psycholinguistic studies have refined the phonetic properties involved in sound symbolism. An early question has been the relative weight of consonants and vowels in subjects' preferred associations. In the 1970s, Tarte [5] argued for a greater influence of vowels while testing a small number of pseudowords and phonemes. A precise assessment of the implication of vocalic features was, furthermore, provided by Knoeferle, Li, Maggioni & Spence [10]. Departing from Tarte's statement, however, recent studies have overall ascertained the predominant role of consonants [6,11,12]. Although with few subjects, Ahlner and Zlatev [13] have also argued that vowels and consonants have distinct but complementary roles. More precisely, in their study, the vowel and the consonant of a CVCV pseudoword could be either congruent or incongruent with respect to what they associated with at the sound symbolic level. In the congruent case, no statistically significant difference was observed between associations involving pseudowords differing on both consonant and vowel (e.g. /titi/ vs /lulu/) and associations involving pseudowords differing on either of them (e.g. /kiki/ vs /nini/ and /lili/ vs /lulu/). In the incongruent case (eg., /tutu/ vs /lili/), the associations were primarily explained by the consonant, although the associative bias did not reach statistical significance.

It is difficult to know precisely what features of consonants trigger associations. Generally speaking, plosives have been shown to associate with spiky shapes, and sonorants with rounded shapes. There are, however, methodological issues, namely the choice of various subsets of consonants in the broad categories of plosives and continuants, and their potentially unbalanced contrasts in the construction of experimental pseudowords. Both can mask which phonetic contrasts subjects rely on in their answers. It is thus hard to disentangle the role of each consonantal feature, among others manner, place of articulation and voicing–the three main features of consonants. For example, Nielsen and Rendall [6,14] contrasted voiceless plosives ([p, t, k]) with sonorants ([l, m, n]), making it hard to judge whether voicing alone, manner alone or both of them have significant effects. As suggested more recently by Nobile [15],

Table 1. Nobile [15]'s results of sound symbolic associations between consonant features and visual shapes.

		Visual shape	
		Spiky	Round
Phonetic features	**Voicing**	Voiceless obstruents (plosives & fricatives)	Voiced obstruents (plosives & fricatives)
	Manner	Plosives (voiced & voiceless)	Fricatives (voiced & voiceless)
	Manner	Oral (fricatives)	Nasal (sonorants)
	Place	Palato-velar (plosives & fricatives)	Alveo-dental (plosives & fricatives)

https://doi.org/10.1371/journal.pone.0208874.t001

manner of articulation and voicing may in fact independently influence subjects' patterns of answers (see Table 1 below). This author tested phonetic contrasts along a number of articulatory dimensions with a 2×2 association task (two visual stimuli and two pseudowords). Confounding phonetic dimensions were, however, present in some of the contrasts, for example voicing and place of articulation in the contrast between plosives and fricatives.

Contrasts in association tasks

With respect to the former question of the phonetic features involved in sound symbolism, the contrastive or non-contrastive nature of the proposed task may have an influence on the associations favored by subjects. If a contrast between two sound forms is presented to the subject in a 1×2 (one visual stimulus and two pseudowords) or in a 2×2 association task, a comparison may take place *between* contrasted sounds or phonetic features in order to choose the more appropriate *with respect to the other*, *much along the lines distinctive features differentiate between phonemes*. As an example, let's assume that when presented with /d/ and /t/ in a 1×2 bouba-kiki association task, subjects associate preferentially /d/ with round shapes, and /t/ with spiky ones. Let's also assume that when presented with /d/ and /n/, they associate preferentially /n/ with round shapes, and /d/ with spiky shapes. Comparing both results, what is associated to /d/ depends on the contrast created between it and another consonant. A logical deduction would then be that /t/, /d/ and /n/ can be placed along a continuum, with /t/ and /n/ at the extremities and /d/ between them.

What if /t/, /d/ and /n/ are now presented independently, i.e. without contrast, in a 2×1 (two visual stimuli and one pseudoword) association task? It may be here risky to straightforwardly anticipate the results from the previous ones: any of these three segments may turn to associate preferentially with either round shapes or spiky shapes, or show no significant pattern of association. Indeed, what is tested here is now phonetically intrinsic relationships between sound forms and shapes, not relative ones. One could, however, suggest from the previous results that /t/ will be more associated with spiky shapes, and /n/ with round shapes—/d/ is more elusive.

Relating to Nobile's results on the independent effects of voicing and manner, it is unsurprising that voiced plosives like /d/ are harder to assess. On the one hand, plosives are associated with spiky shapes while sonorants are associated with round ones; on the other hand, voiced consonants are associated with round and voiceless ones with spiky. If a contrastive presentation of two pseudowords sheds light on one of these two characteristics, it can be predicted that /d/ will be more associated with round shapes when presented along with a voiceless plosive such as /t/, and more associated with spiky shapes when presented with a voiced sonorant such as /n/. But it remains difficult to predict what will happen when /d/ is presented

alone in 1×1 or 2×1 association tasks, since this will likely depend on the relative associative strengths of the competing features–voicing and manner (letting aside further possible interactions with vowels). Additionally, the fine graphic details of the used figures will also play a significant role. All in all, the precise nature of the task must be factored in in the analysis of the results.

Explicit versus implicit tasks to assess sound symbolism

Nielsen and Rendall [14] have argued that the strength of the bouba-kiki effect is overestimated because of the 'transparency' of the testing protocols. Indeed, associative tasks point to sound symbolism when requesting the subjects to establish a link between stimuli of different natures. As previously said, they also point to phonetic and/or visual contrasts when asking explicitly to choose between two stimuli of the same nature. Transparent presentations of contrasts may lead to metacognitive strategies masking more low-level processes and increasing effect sizes.

The previous consideration suggests why Nielsen and Rendall got smaller effect sizes for sound symbolic associations brought to light in their implicit experimental protocol. It consisted in learning pairs of shapes (rounded or spiky) and pseudowords (composed of either voiceless plosives [p, t, k], or sonorants [l, m, n]). In the first part of the experiment, half of the participants learnt 'congruent' pairs (with an assessment of congruence coming from earlier studies), the other half 'incongruent' pairs. Then, in the second part of the experiment, other pairs were presented and subjects had to decide whether these pairs were correct according to the rules they had previously learned. The recall performance was better in the congruent condition (53.3% vs 50.4%), which suggests that the congruent pairs were easier to learn and to remember.

A number of further studies have aimed at assessing sound symbolism in a more implicit way than 'classical' judgment or association tasks.

In a first study, Aveyard [11] asked participants to decide which of two images best associated with a pseudoword presented orally. A feedback was provided after each response, stating whether the association was correct or incorrect. Stimuli were presented repeatedly and the associations to be learnt were consistent throughout the experiment, but half of them were congruent at the sound symbolic level, and the other half was not. Participants could therefore not generalize sound symbolic rules for the whole set of associations. Given this, a relatively better learning performance was observed when rules were congruent (57% vs 50%).

In a subsequent study, which also consisted in a choice between two shapes for a pseudoword presented orally, subjects had to implicitly detect which shape was consistently associated with a given pseudoword [16]. This shape, e.g. a round shape, was either associated with a second shape of opposite nature, e.g. a spiky shape, or a distractor, e.g. a different round shape. Neither explicit rules nor feedbacks were given. Four learning blocks followed one another, and a quicker improvement for congruent associations was observed (from 55% vs 52% for congruent and incongruent associations in the first block to 68% vs 58% in the second block, as extracted from the figures of the article), although performance was similar at the end for congruent and incongruent pairings (70% vs 71% in the last block).

In an another study, Sidhu and Pexman [17] demonstrated the impact of the supraliminal priming of a pseudoword on the categorization of ambiguous shapes. In the written condition of the task, shapes were more categorized as round when preceded by a pseudoword composed of 'round' phonemes including consonants /b, m/ rather than of 'sharp' phonemes including consonants /t, k/ (57% vs 50%) (p. 1971–1973). This result was replicated in the oral condition (53% vs 43%).

In these three recent experiments, the effect sizes suggest altogether that the sound symbolic associations were much weaker compared to what is commonly observed in the classical explicit association tasks. However, it can be argued that the existence of metacognitive strategies cannot be ruled out. In Sidhu and Pexman's study in particular, pseudowords were consciously perceived and the justification given for their presence–they were described as not relevant to the current task but to later ones–could easily be questioned by participants.

As already mentioned, Westbury [9] conducted a study with English speakers using a protocol that can be considered as significantly more implicit than the previous ones. A lexical decision task was conducted with written words and pseudowords composed of either or both plosives or sonorants (Westbury actually uses the terms stops and continuants, following a phonological distinction rather than the phonetic distinction we adopt in this paper; both descriptions are valid, as explained in [18]), presented in a spiky or a rounded frame. Response time for pseudowords composed of plosives were significantly faster when presented in spiky frames, and conversely, responses of pseudowords composed of sonorants were significantly faster when presented in rounded frames. In a second task, letters and numbers were tested in the same manner. Decisions on these items did not require the same semantic access, and hence allowed to evaluate lower-level cognitive processes. In both experiments, results were consistent with sound symbolic expectations, i.e. an interaction was observed between the shape of frames and the type of consonants. However, effects were only marginally significant, suggesting once again that the less transparent a protocol, the weaker the sound symbolic associations.

Influence of the shape of letters in tasks on sound symbolism

In all studies focusing on the sound symbolism of graphic shapes but relying on written forms, a potential confound exists given possible intra-modal visual associations involving the graphemic shapes of the written forms. This issue has been noted in some of the aforementioned studies. Westbury [9] assessed the influence of the shape of letters and numbers in his second task, distinguishing angular characters from curvy ones. He noticed no interaction between graphemic features and frames, which led him to conclude that there was a '*lack of evidence to support the orthographic form hypothesis*' (p. 16), although it can be argued that the second task was in many ways different from the first lexical decision task. As for Nielsen and Rendall [14], they neutralized the issue by creating 'mixed orthographic representations', with lowercase and uppercase letters, that 'did not consistently align with either possible matching rule' (p. 119).

The fact that a lot of studies actually presented the linguistic material orally [5,6,11–13,16,17] is obviously a strong point in favor of sound symbolism. One could argue that acoustic stimuli activate written representations in the subjects' minds, at least in the minds of the competent readers, usually university students, that form the bulk of participants in experimental psycholinguistics. This cannot however be the whole story, given Bremner et al. [2]'s results with a "bouba-kiki" task in a population without written tradition, the Himbas of Namibia. While it makes perfect sense to prevent a writing bias when investigating sound symbolism, doing so however restricts our understanding of possibly intertwined processes, implying both sound symbolic associations and purely visual ones.

Cuskley, Simner and Kirby [19] attempted to explain the bouba-kiki phenomenon in terms of graphemic bias rather than, or in addition to, sound symbolism. With written pseudowords, they found that angular letters (*k*, *t*, *z* or *v*) associate with spiky shapes, while curvy letters (*g*, *d*, *s*, *f*) associate with rounded shapes. Interestingly, this effect persisted with oral pseudowords, which could possibly suggest that hearing a pseudoword automatically activate the mental

representation of its written form, although other explanations are possible, as mentioned in the next two paragraphs. The authors additionally found an interaction between voicing and shape (voiced consonants with round shapes, and voiceless consonants with spiky shapes), in the oral condition only. These results do not challenge preceding results, but highlight that it is indeed not easy to separate sound symbolic associations from purely visual ones when using written material.

Along the same line, Westbury [9] noted that disentangling concurrent effects–purely graphic and sound symbolic—is difficult. This starts with the difficulty of judging whether a graphemic symbol is more angular or curvy–for example, "f" is considered as angular by Westbury and as curvy by Cuskley, et al. [19]). Additionally, associations between i) curvy letters and round shapes, and ii) angular letters and spiky shapes may also reflect intricate phonetic properties of the corresponding sounds. /d/, /g/, /s/ and /f/ may be related to round shapes because voiced plosives and voiceless fricatives are. Conversely, /t/, /k/, /z/ and /v/ may be related to spiky shapes because voiceless plosives and voiced fricatives are. This is actually supported by Fort et al. [12]'s results in a purely auditory task.

Finally, phonetic features may partly decide of the graphemic forms of letters, as discussed by Turoman and Styles [20]. These authors obtained better-than-chance performance in a task that consisted in guessing which glyph among two was referring to the sound /i/ or /u/ in multiple written traditions. This suggests that the shapes of letters may be historically motivated by the sound they refer to, which would then be another instance of sound symbolism.

If they exist, intra-modal visual interactions may appear in addition to sound symbolic effects. The question is then raised of the respective effect sizes of these effects. The near-absence of significant sound symbolic effects in Cuskley et al.'s statistical models could be explained by the intricacies of an unbalanced experimental material and specific choices of consonants, e.g. choosing only fricatives [s, z, f, v] for continuants, while other studies mostly consider sonorants like [l, m, n].

Because of such difficulties, a more encompassing test of various associative effects is needed, which explicitly allows for effects that add to each other, or compete with each other.

Method

Ethics statement

This research has been approved by the ethical committee "Comité de Protection des Personnes SUD-EST IV" (Lyon, France) with the reference number L15-210. All participants gave written informed consent to participate in the experiment.

Overview

Our objectives were i) to assess sound symbolism in a non-transparent task to address the issue of possible metacognitive strategies and oversized effects, ii) to pay specific attention to the involved phonetic dimensions in order to better assess their respective roles, and iii) to explicitly tackle the possible effects of written forms. We thus chose to extend rather than to replicate Westbury [9]'s experiment by adding a third independent variable to his original design: the shape of letters, using either a curvy font, *Gabriola*, or an angular font, *Agency FB*, for the display of words and pseudowords.

Furthermore, we applied some modifications to Westbury's experimental setting. First, aiming to better disentangle the phonetic dimensions at play, we dissociated voiced and voiceless plosives, on the basis of Nobile [15]'s findings (see Table 1). Second, although they have been used in Westbury's study and in a few others [21], we did not create mixed pseudowords

(i.e. composed of both plosives and sonorants) because of our lack of expectation in this case with respect to the two unmixed conditions.

We therefore investigated the effects of three parameters with a 3×2×2 plan: the category of consonants (voiced plosives, voiceless plosives and sonorants), the type of frames (spiky or round) and the font (angular or curvy). In our analyses, we allowed for the possibility of additive effects, either superimposing or competing, and we considered pseudowords and words independently, having in mind the well-known result that expert readers do not process the former the same way as the latter [22].

Throughout the paper, the *p*-values report the 'exact level of significance, calculated from the data after the experiment' [23] and no arbitrary level (such as 0.05) in hypothesis testing is indicated.

Hypotheses

Based on Westbury [9]'s experiment, we could expect an interaction between the shapes of frames and the category of consonants. More precisely, faster response times were expected in congruent situations than in incongruent situations, as specified in part (a) of Table 2. Based on Nobile's findings, it was more difficult to formulate predictions in the case of voiced plosives, as they could be associated both with spiky frames as plosives and with round shapes as voiced consonants.

Congruent associations are expected to induce faster reaction times than incongruent associations for each 2×2 interaction of parameters under study.

Given Cuskley et al.'s results, we could further expect an interaction between the type of frames and the font, with again faster response times in congruent situations than in incongruent situations (see (b) in Table 2).

The hypothesis of an interaction between font and phonetic composition could finally be made, considering sound symbolic associations with letter shapes in a similar fashion as with frames. Once again, faster response times were expected in congruent situations than in incongruent situations (see (c) in Table 2). As explained above, response times in the case of voiced plosives were difficult to predict since these consonants could be congruent with spiky frames because they were plosives, or congruent with rounded frames because they were voiced. Which association would be stronger could not be anticipated, and we thus did not have specific hypotheses.

Table 2. Congruent and incongruent associations of visual and phonetic stimuli according to previous studies.

	Interaction of parameters under study	Type of association of stimuli	
(a)	**Type of frames × Category of consonants** Sound symbolic interaction	**Congruent**	spiky frames & voiceless plosives
			round frames & sonorants
		Incongruent	round frames & voiceless plosives
			spiky frames & sonorants
(b)	**Type of frames × Font** Visuo-visual interaction	**Congruent**	spiky frames & angular font
			round frames & curvy font
		Incongruent	round frames & angular font
			spiky frames & curvy font
(c)	**Category of consonants × Font** Sound symbolic interaction	**Congruent**	angular font & voiceless plosives
			curvy font & sonorants
		Incongruent	curvy font & voiceless plosives angular font & sonorants

https://doi.org/10.1371/journal.pone.0208874.t002

This experimental design thus allowed us to test three hypotheses of interaction. Given the previous studies and assuming the existence of simultaneous effects, we postulated that the three preceding interactions could be significant. Finally, we did not have expectation about a potential triple interaction *Type of frames* × *Category of consonants* × *Font*.

Participants

21 female and 20 male students from universities in Lyon, aged from 18 to 30 years (average 22.2 years), were recruited (N = 41). All were monolingual native French speakers and had a normal vision or corrected to normal, with no history of speech or hearing disorders reported at the time of experiment. Six of them were left-handed.

Material

Words and pseudowords. We defined a number of criteria to select words and create pseudowords. All strings (i.e. both words and pseudowords) contained:

- three, four or five letters;
- three or four phonemes;
- three possible syllabic structures: CVC, CVCV or VCVC (C stands for consonant, V for vowel).

Specific constraints were applied to the choice of vowels, as detailed in S1 Protocol.

We collected 233 words corresponding to our criteria, with associated information in the *Lexique 3.81* database [24]:

- written and oral frequencies (respectively in books and in movies);
- number of letters and phonemes;
- syllabic structures.

We further extracted the categories of consonants: word made of plosives, of sonorants or mixed.

In parallel, we generated 974 pseudowords. Apart from frequencies, similar information as for words was compiled.

For both words and pseudowords, orthographic and phonological neighbors were figured out on the basis of Luce and Pisoni [25]'s method by deleting, adding or substituting one phoneme / letter (for phonological / orthographic neighbors) in any position. Once neighbors were found, neighborhood frequencies were computed.

On the basis of the preceding corpora, a genetic-algorithm-based program named *Bali* [26] was used to generate lists of words and pseudowords that were as balanced as possible with respect to confounding variables (number of letters, of phonemes, of phonological and orthographic neighbors, frequencies of these neighbors etc.), and as internally diverse as possible. This was in order to produce a well-balanced corpus and a variety of combinations for later regression analyses.

Lists of pseudowords were first generated, then lists of words were created with lists of pseudowords as counterparts in the balancing optimization process.

For pseudowords, four lists were created: one with voiced plosives, one with voiceless plosives, and two with sonorants–in order to have as many pseudowords composed of sonorants as of plosives. For words, four lists were also created: one with plosive-only words, one with sonorant-only words, and two with mixed words.

Table 3. Number of words and pseudowords for each category of consonants in the experimental material.

Words			Pseudowords		
Mixed	Sonorants	Plosives	Sonorants	Plosives	
				Voiced	Voiceless
64	32	32	64	32	32

https://doi.org/10.1371/journal.pone.0208874.t003

We obtained a total of 256 character strings, divided into 128 words and 128 pseudowords. Half (64) of the pseudowords were composed of sonorants ([l, m, n]), half of plosives. Furthermore, half of the latter (32) were composed of voiceless plosives ([p, t, k]), and half (32) of voiced plosives ([b, d, g]). In the group of words, 32 were composed of sonorants, 32 of plosives (voiced or voiceless) and 64 were mixed words (containing both sonorants and plosives, whatever the voicing) (see Table 3, and see S1, S2, S3 and S4 Tables for the actual lists of items and their properties).

There were five pairs of homophones among the 128 pseudowords (imale/immal; nalle/nal; lummu/lumue; lul/lulle; nanu/nannu), and one among the 128 words (laque/lac), with no occurrence of two homophones in the same list.

Frames. In Westbury's original experiment, frames were presented as white objects on a black background. Yet, to avoid eye strain due to the presentation of character strings on a white background (in the center of the frame), we decided to keep only the contours of frames, presented in white on a black background (see Fig 1).

We selected 16 of the 40 frames used in Westbury's experiment–eight spiky and eight rounded–in order to focus on those that seemed most relevant to assess sound symbolic effects. To this end, we chose shapes that were as asymmetric, unambiguous in terms of roundedness or spikiness, and non-reminiscent of existing or imaginary entities (like ghosts), as possible (see Fig 2).

Fonts. *Agency FB* was chosen as our 'angular' font due to its right-angled letters. *Gabriola* offered rounded letters without right-angled corners, and was therefore chosen as our 'curvy' font (see Fig 3). No formal test was, however, performed, or judgment task conducted, as for the angularity or curviness of the letters. Fig 4 offers two examples of written forms displayed in a frame, as presented to participants.

Procedure

The open-source software *OpenSesame* [27] was used to generate the experiment and collect answers and response times, more specifically with the *psycho* back-end, which relies on *PsychoPy* and offers good temporal resolution for display and response time.

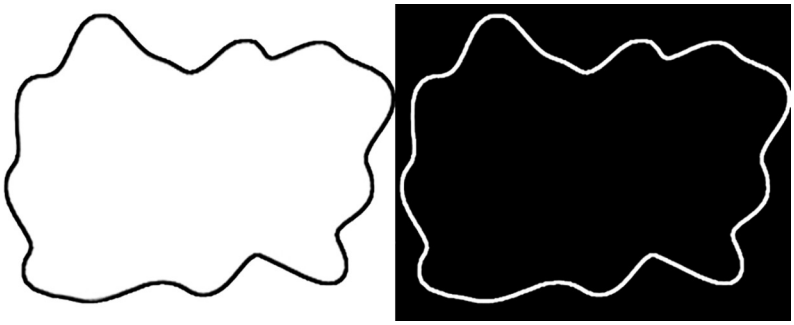


Fig 1. Original frame used in Westbury [9]'s experiment (left) and corresponding edited frame in our experiment (right).

https://doi.org/10.1371/journal.pone.0208874.g001

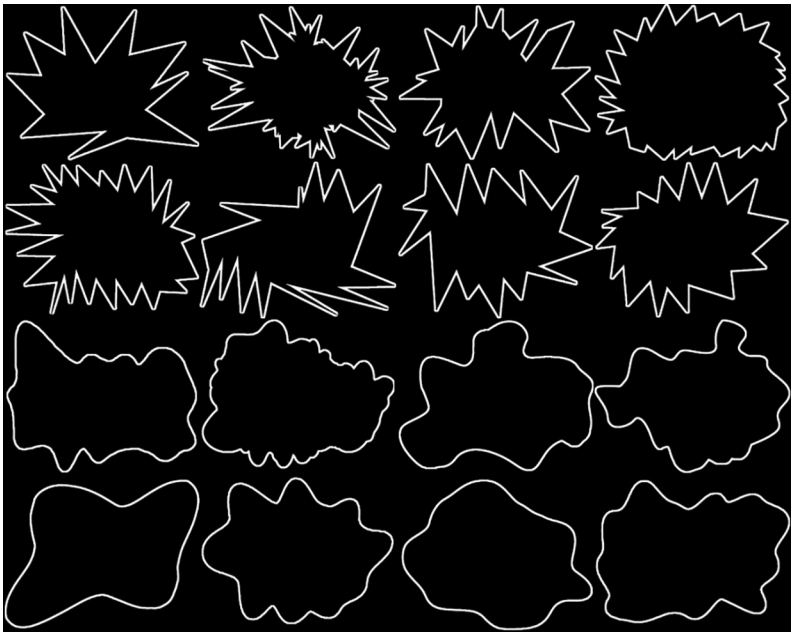


Fig 2. The 16 frames used in the experiment. The top eight frames are considered spiky, the lower eight rounded.

https://doi.org/10.1371/journal.pone.0208874.g002

Subjects entered their choice—word or not—with two keys (see S1 Protocol). They were asked to answer accurately and as fast as possible.

A fixation point was presented for 200 ms, then the frame appeared for a varying duration (between one and three seconds) before the character string appeared in its center (this corresponds to the stimulus-onset asynchrony or SOA). The string and the frame were displayed until the answer of the participant, otherwise they disappeared after 2 seconds. Then, a mask composed of a succession of three images of a Gaussian noise was presented for 99 ms (33 ms for each image) to avoid any retinal persistence.

The experiment began with a training phase in which height trials were presented (four words and four pseudowords). These practice stimuli were not presented again in the main experiment. After the training, the percentage of success was displayed on screen to both the participant and the experimenter, which allowed the latter to ensure the former understood the instructions and used the right keys. The experiment was then divided in four blocks of 64 trials interspersed by breaks whose duration was determined by the participants themselves.

abcdefghijklmnopqrstuvwxyz

abcdefghijklmnopqrstuvwxyz

Fig 3. The *Agency FB* and *Gabriola* fonts used in the experiment. The *Agency FB* font (bottom) is the angular font, and the *Gabriola* font (top) the curvy font.

https://doi.org/10.1371/journal.pone.0208874.g003

Fig 4. Examples of trials with two pseudowords. On the left, the pseudoword is presented in a round frame with the angular font, on the right in a spiky frame with the curvy font (Gabriola).

https://doi.org/10.1371/journal.pone.0208874.g004

The matchings (between the phonetic categories related to the character strings, the type of frame and the font) were generated in a pseudorandom way for each subject. Half of the pseudowords were presented in a spiky frame, and half in a rounded frame. Half of the pseudowords in each of these two conditions were displayed with the *Gabriola* font, and half with the *Agency FB* font. Finally, the category of consonants of the pseudoword was taken into account: each match between a type of frame and a font (for example, spiky and angular) was presented with 16 sonorants, eight voiceless plosives and eight voiced plosives (see Table 4).

The order of presentation of the stimuli was constrained to avoid repetition effects (see S1 Protocol).

Results

For all statistical analyses, we used the R project [28] with especially the package `ggplot2` for graphics [29], `reshape` and `plyr` [30] for data manipulation, and `lme4` [31] and `gamlss` [32] for generalized mixed modelling.

Success rate

Following Westbury [9], we chose to a priori eliminate subjects who had more than 20% of erroneous answers. The highest error rate was 12.9%, hence all 41 subjects were taken into account.

Presentation of the response times

Only correct responses were selected for further analysis. For these responses, the average response time was then 848 ms (sd = 243ms) for pseudowords, and 810 ms (sd = 246 ms) for words. There was no trimming of the data due to the skewness of the distribution of response times, both for words and pseudowords (see S1 Analysis). The datasets for pseudowords and

Table 4. Distribution of experimental stimuli with respect to type of frame, font and category of consonants.

Spiky frame						Round frame					
Gabriola			*Agency FB*			*Gabriola*			*Agency FB*		
Sonorants	Plosives		Sonorants	Plosives		Sonorants	Plosives		Sonorants	Plosives	
	Vd	Vs		Vd	Vs		Vd	Vs		Vd	Vs
16	8	8	16	8	8	16	8	8	16	8	8

Vd stands for Voiced, Vs for Voiceless.

https://doi.org/10.1371/journal.pone.0208874.t004

words are available as supplementary material (see S1 Dataset and S2 Dataset respectively, as well as S1 Structure for a detailed description of their content).

Analysis of response times for pseudowords

Regarding explanatory variables, we included the font, the type of frame and the category of consonant, as well as their three two-by-two interactions and their triple interaction. We also included the trial position and the response time of the preceding trial, as justified in S1 Analysis.

The fixed effects were therefore:

- **Font** (Angular or Curvy)
- **Type of Frame** (Spiky or Round)
- **Category of consonants** (Voiceless Plosives, Voiced Plosives or Sonorants)
- **Type of Frame × Font**
- **Category of consonants × Font**
- **Type of Frame × Category of consonants**
- **Type of Frame × Font × Category of consonants**
- **Trial position**
- **Preceding Response Time**

We considered three random effects to account for the non-independence of our response times and to avoid any type of pseudo-replication [33]:

- **Subject** (for our 41 participants)
- **Stimulus** (for the 128 pseudowords)
- **Frame** (for our 16 frames)

Other variables such as the number of letters, the syllabic structure etc. could have been included as predictors in the model too, thus adding an a posteriori control to the a priori control (see Section Words and Pseudowords). However, there were then high levels of

Table 5. Likelihood ratio tests for the fixed predictors of response times for pseudowords in a Generalized Gamma gamlss model.

	Df	AIC	LRT	Pr(Chi)
Full model		64,715		
Type of Frame : Cat. Of Consonant	2	67,711	0.75	0.689
Type of Frame : Font	1	64,719	6.84	0.009
Cat. Of Consonant : Font	2	64,714	3.02	0.221
Trial position	1	64,820	107.84	< 0.001
Preceding response time	1	64,816	103.20	< 0.001
Subject (random)	40.5	66,524	1,890.92	< 0.001
Stimulus (random)	114.4	65,419	933.42	< 0.001
Frame (random)	1.1	64,714	2.17	0.161

Df stands for 'degrees of freedom', AIC for 'Aikake Information Criterion', and LRT for 'Likelihood ratio tests'.

https://doi.org/10.1371/journal.pone.0208874.t005

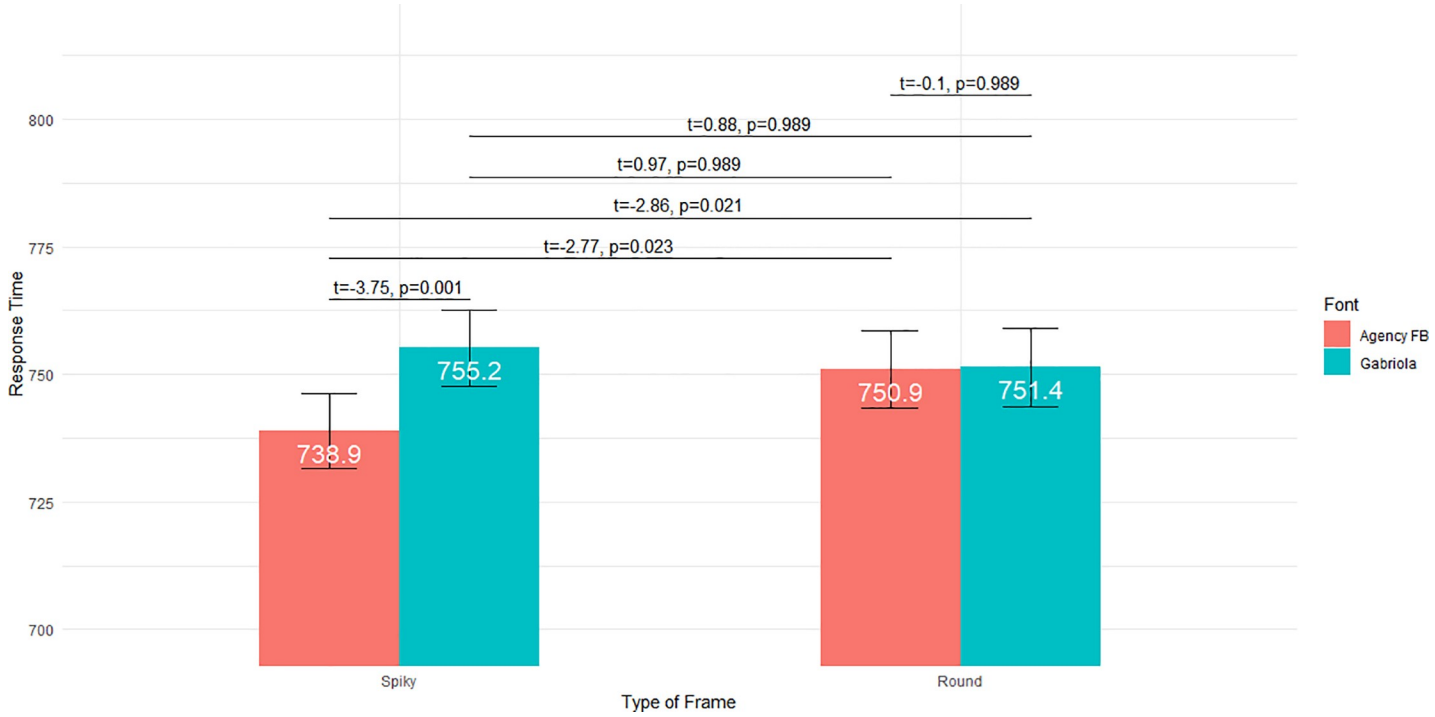


Fig 5. Interaction between Type of Frame and Font for pseudowords. Marginal locations are displayed numerically (white figures) and graphically for the four conditions *Spiky Frames* & *Angular Font*, *Spiky Frames* & *Curvy Font*, *Round Frames* & *Angular Font* and *Round Frames* & *Curvy Font*. Vertical bars report the confidence intervals for the four means. Significance levels of pairwise comparisons of these conditions are reported above. *P*-values have been corrected for multiple comparisons with Holm's method.

https://doi.org/10.1371/journal.pone.0208874.g005

multicollinearity between the predictors, which violated the assumptions of the models. We hence chose not to include these variables, rather than to complicate the statistical analysis with a selection of the best set of predictors (based on variance inflation factors).

In order to model error terms correctly, we compared different generalized mixed-effect regression models with response time as dependent variable. To do so, we initially relied on models with distributions belonging to the so-called exponential family, as made available by the ***glmer()*** function of the `lme4` package. We then switched to generalized additive models

Table 6. Likelihood ratio tests for the fixed predictors of response times for words in a Generalized Gamma gamlss model.

	Df	AIC	LRT	Pr(Chi)
Full model		58,349		
Type of Frame : Cat. Of Consonant	3	58,351	8.14	0.043
Type of Frame : Font	1	58,352	5.50	0.019
Cat. Of Consonant : Font	3	58,347	4.30	0.231
Trial position	1	58,348	107.84	0.373
Preceding response time	1	58,385	103.20	< 0.001
Subject (random)	40.2	59,630	1,890.92	< 0.001
Stimulus (random)	117.2	59,097	933.42	< 0.001
Frame (random)	0.0	58,349	0.00	< 0.001

Df stands for 'degrees of freedom', AIC for 'Aikake Information Criterion', and LRT for 'Likelihood ratio tests'.

https://doi.org/10.1371/journal.pone.0208874.t006

for location, scale and shape (GAMLSS) [32,34,35], available in the `gamlss` package. Details of why and how we compared these models are given in S1 Analysis.

We found that the *Generalized Gamma* (GG) distribution, which is a generalization of the *Gamma* (GA) and *inverse Gaussian* (IG) distributions, was an appropriate choice for error terms. Only the location parameter of the distribution was modelled with the previous predictors, other parameters of scale and shape were estimated by an intercept only. While location corresponds to the mean in *inverse Gaussian* and *Gamma* distributions, it does not in the *Generalized Gamma*. It was, however, proportional to it given that scale and shape were modelled by intercepts only in our approach.

While we report below the outputs of GG models, we also computed results for other distributions in order to assess effects over a range of models, and therefore increase our confidence in them.

A first model was run on the whole set of pseudowords (n = 5,100). Observations corresponding to normalized quantile residuals below -2.5 or above 2.5 were removed (see S1 Analysis), and the model was updated on the trimmed dataset (n = 5,035) before further computations were performed. This strategy, suggested by Baayen and Milin [36] and named model criticism, was preferred to a–priori trimming, since it better accounted for the specific, non-Gaussian, distribution of error terms of each generalized regression model. Assessments of the assumptions underlying the model were all satisfactory (see S1 Analysis).

The significance of the predictors was assessed with Likelihood ratio tests (LRT). The triple interaction was non-significant ($\Delta AIC = 4$, $Df = 2$, $LRT = 0.015$, $p = 0.99$), and double interactions were assessed once it was removed from the model.

Results are given in Table 5. The first column indicates which predictor term is dropped in the nested model. Except for the full model, the second column (Df) gives the difference of degrees of freedom between the full model and the nested model. The fourth column (LRT) reports the difference in deviance between these two models, and the fourth column (Pr(Chi)) the *p*-value of the χ^2 test on the difference of deviance. **Type of Frame**, **Category of Consonant** and **Font** are absent as main effects given the presence of their interactions. *P*-values suggested a significant interaction for **Type of Frame × Font**, but not for the other two interactions. This result was overall congruent with what was found with other distributions (IG, GA, Johnson's SU, see S1 Analysis).

To further understand the pattern of interaction between the type of frame and the font, we assessed the six possible contrasts between the four conditions ***Spiky Frames* & *Angular Font***, ***Spiky Frames* & *Curvy Font***, ***Round Frames* & *Angular Font*** and ***Round Frames* & *Curvy Font***. We first computed the estimated marginal locations of the response times in the four conditions, i.e. the locations adjusted for other variables in the regression models. For each contrast between two marginal locations, a z-test of the difference between these locations was then performed. We considered the Holm correction to decide which null hypotheses should be rejected when controlling for the inflated type I error rate due to multiple comparisons [37]. Fig 5 summarizes the values of the four marginal locations and the results of the six z-tests of difference.

The *a priori* congruent ***Round Frames* & *Curvy Font*** condition does not induce faster response times than the *a priori* incongruent conditions ***Round Frames* & *Angular Font*** and ***Spiky Frames* & *Curvy Font***. On the contrary, the *a priori* congruent ***Spiky Frames* & *Angular Font*** condition is faster than the two corresponding incongruent conditions, and also than the ***Round Frames* & *Curvy Font*** condition. Overall, the interaction is therefore due to the faster response times obtained in the ***Spiky Frames* & *Angular Font***, compared to the three other conditions.

Analysis of response times for words

We applied the same analytical procedure to words. Once again, a GG distribution appeared appropriate with respect to error terms.

In the initial model with all entries (n = 4,570), 43 entries had normalized quantile residuals higher than 2.5 or lower than -2.5, and were discarded in a second model (n = 4,527). Assessments of the assumptions underlying the model were all satisfactory.

Once again, the triple interaction **Type of Frame × Font × Category of Consonant** was non-significant (***Δ AIC*** = 1, ***Df*** = 3, ***LRT*** = 4.34, ***p*** = 0.23), and double interactions were assessed once it was removed from the model. Table 6 reports the various LRT performed.

P-values suggested a significant **Type of Frame × Font** interaction, no significant interaction for **Font × Category of Consonant**, and a **Type of Frame × Category of Consonant** interaction. Regarding **Type of Frame × Font**, computations of the marginal locations and of their contrasts are given in Fig 6. The pattern was reminiscent of what was observed for pseudowords. However the difference between ***Spiky & Agency FB*** and ***Spiky & Gabriola*** was at the 0.05 level before the Holm correction, and higher after. Rather than ***Spiky & Agency FB*** being significantly different from the three other conditions, the model therefore suggested a main effect of **Type of Frame**, with shorter response times for spiky frames than for round frames. Once again, models with different distributions (IG, GA, Johnson's SU) gave similar results qualitatively, although the significance of *p*-values varied from one model to the next. The JSU model in particular suggested both a main effect of **Type of Frame**, with shorter response times for spiky frames, and, as for pseudowords, shorter response times for ***Spiky & Agency FB*** compared to the three other conditions.

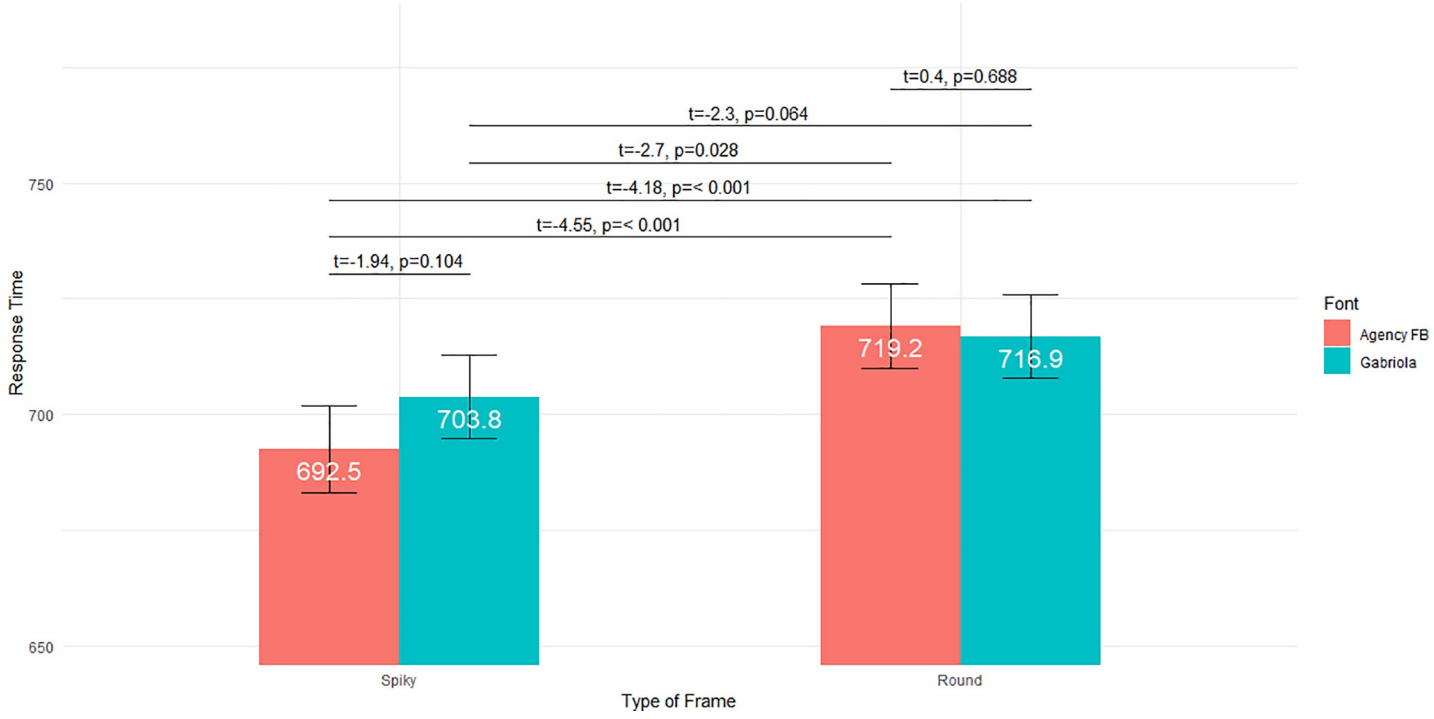


Fig 6. Interaction between Type of Frame and Font for words. Marginal locations are displayed numerically (white figures) and graphically for the four conditions *Spiky Frames* & *Angular Font*, *Spiky Frames* & *Curvy Font*, *Round Frames* & *Angular Font* and *Round Frames* & *Curvy Font*. Vertical bars report the confidence intervals for the four locations. Significance levels of pairwise comparisons of these conditions are reported above. *P*-values have been corrected for multiple comparisons with Holm's method.

https://doi.org/10.1371/journal.pone.0208874.g006

As for the **Type of Frame × Category of Consonant** interaction (Fig 7), the pattern of response times ran counter to our hypothesis (a), since for example round frames and sonorants led to longer response times than spiky frames and sonorants. This interaction was thus unsupportive of sound symbolism. What can be stressed is the case of voiced plosives, with in particular much longer response times for the **Round × Voiced Plosive** condition, compared to all other conditions. This effect likely explains why the interaction was significant with LR tests. We had no specific predictions for voiced plosives, and the result for the **Round × Voiced Plosive** condition is difficult to explain. Also, the **Type of Frame × Category of Consonant** interaction was not found when considering IG or GA distributions rather than GG, which casts doubts over its actual significance.

As a summary regarding our three oriented hypotheses, we did not get any interaction between **Font** and **Category of consonants**. In words, but not pseudowords, we found a statistically significant interaction between **Type of Frame** and **Category of consonants**, in conflict with sound symbolism. Finally, we observed a statistically significant interaction between **Type of Frame** and **Font** for both pseudowords and words. More precisely, with pseudowords, we observed faster responses for spiky frames and angular letters than for the three other conditions (spiky frames and curvy letters, round frames and angular letters, and round frames and curvy letters). For words, we saw rather a main effect of spiky frames.

While the amount of difference between the two fonts was not assessed a priori with a pretest, this result shows that differences were large enough to elicit a differentiated pattern of answers.

Subjects' reports about the experiment

Upon finishing the experiment, we asked subjects if they had noticed something special. If they answered yes, we then asked what was special and eventually if they had noticed that

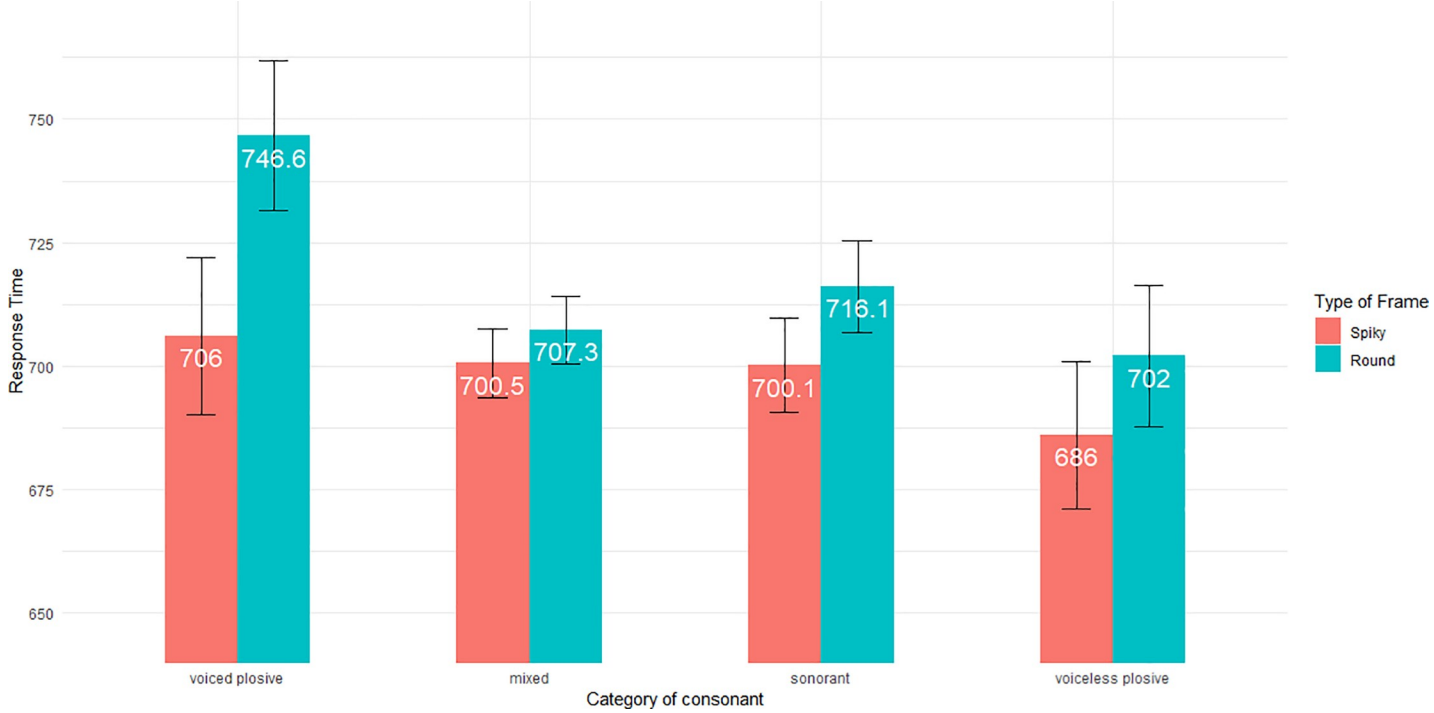


Fig 7. Interaction between Type of Frame and Category of Consonant for words. Marginal locations are displayed numerically (white figures) and graphically for height different conditions. Vertical bars report the confidence intervals for the four locations.

https://doi.org/10.1371/journal.pone.0208874.g007

there were different fonts. None of the 41 subjects spontaneously reported the existence of two fonts and only 10 subjects answered yes when explicitly asked. One subject reported faster answers when the frame was spiky, and another one when it was spiky and when the word had a negative valence.

Discussion

We did not observe the cross-modal interaction between phonology and visual form found in Westbury [9]. We obtained faster answers only when the spiky frames and the angular font were displayed together. This brought us closer to Cuskley et al.'s proposal of visual interaction effects, suggesting that such effects should be taken into account while investigating sound symbolism. Below, we focus on the visual processes possibly underlying the specific results we obtained, first with respect to geometric shapes in general, second with respect more specifically to written words and pseudowords. We then discuss the cognitive processes underlying sound symbolism, in relation to the transparency of the task, and in terms of low and high-level processes. We point in particular to the possibility of ideasthetic processes in addition to synaesthetic ones.

Visual processes

Low-level visual processes in tasks involving angular and curvy visual shapes. Faster answers for the combination of spiky frames and angular letters suggest an effect of visual saliency. Indeed, some studies highlight an attentional enhancement due to simple geometric shapes. For example, some minimal stimulus configurations enhance the capture and maintenance of attention [38]: downward-pointing stimuli (a downward pointing V or triangle) are more rapid to detect than other stimuli such as an upward-pointing V or triangle, or a circle. Moreover, this shape induces greater difficulty to disengage attention. This attentional modulation can be explained by a negative valence carried by angular configurations, especially downward-pointing stimuli. Negative stimuli are indeed known to be faster to detect and to retain attention for a longer time than positive stimuli. Armbruster, Suchert, Gärtner and Strobel [39] collected ratings about angular and curvy configurations and found that downward pointing triangles are judged as more negative and arousing than upward pointing triangles, which are in turn more negative and arousing than circles. These assessments are further in line with measures of peripheral physiological parameters. Difference in cognitive processing between upward pointing and downward pointing triangles are further evidenced by fMRI studies [40]

Different authors suggest a connection between geometric shapes and faces: facial features expressing anger are angular and diagonal forms (e.g. frowning eyebrows) including acute angles pointing downward, while happiness is characterized by curved patterns [41]. In particular, Bassili [42] showed that anger is characterized by a downward movement on the forehead due to frowning. Aronoff et al. [41] conducted a study in which masks of several cultures, either threatening or not, were evaluated. Cross-culturally, masks expressing threat contained more triangular eyes or visible teeth than nonthreatening masks. Some features are direct iconic representations of facial expressions (e.g. frowning) but others (e.g. pointed ears) seem to convey a subjective meaning that may derive from basic visual patterns involving two specific features: angularity and diagonality. Coelho, Cloete, & Wallis [43] evaluated in a more systematic way the impact of emotional content using different types of stimuli in a visual search paradigm. Comparing schematic faces with abstract faces built with straight or curved features, they reached the conclusion that subjects' answers are explained neither by resemblance to faces and the associated emotions, nor by judgments of valency, but rather by the

characteristics of lower-level visual cues. The orientation of features seems to be the key parameter regarding differences in detection speed: lines perpendicular to the edge are more rapidly detected, therefore more salient, than concentric features.

With respect to our experiment, the previous studies would suggest a main effect of the font, with faster response times for angular letters, and a main effect of the type of frames, with faster response times for spiky frames. However, we observed a more complex pattern of effects with faster response times for pseudowords only when the sharp angles found in spiky frames co-occurred with those found in angular letters. This interaction suggests that although the visual saliency of such stimuli seems to play a role, it is a component of a more complex cognitive treatment.

Regarding pseudowords, a possible explanation is that while the dissimilarities between spiky/angular and round/curved shapes are not enough in the context of the lexical decision task to induce differences in response times, an asymmetrical priming effect takes place when angular letters are displayed within spiky frames: contrary to round frames, spiky frames first arouse attention to sharp angles and perpendicular lines, which then facilitates the processing of angular letters.

As for words, there was rather a main effect of the type of frame, with spiky frames corresponding to shorter response times, than a specific interaction between angular letters and spiky frames. A possible explanation for this different pattern of results echoes the ideas underlying dual-route models in reading: along a first route, words are processed more holistically, with access to the mental lexicon; unknown written forms–and therefore pseudowords–are deciphered along a second route on the basis of grapheme-phoneme association rules [22]. Along these lines, the processing of written words could be less impacted by graphemic features than the processing of written pseudowords. There would be therefore no priming effect of spiky frames on angular letters.

From processing geometric shapes to processing written pseudowords. Beyond these interpretations in terms of basic geometric features, an alternative or rather complementary explanation lies in the processing of written pseudowords in terms of linguistic stimuli, and not as arbitrary assemblages of basic shapes.

According to Dehaene and Cohen [44], an area localized in the left fusiform gyrus, in the visual occipital-temporal stream, appears to respond more to words than to other visual objects: the visual word form area, or VWFA. In Baker et al. [45] for example, English words and strings of consonants elicit stronger responses in English speakers than line drawings of things, numbers, or characters from another writing script (Hebrew letters and Chinese ideograms). As stated in Dehaene and Cohen [44], the VWFA would result from '*a putative mechanism by which a novel cultural object encroaches onto a pre-existing brain system*', in agreement with Dehaene [46]'s 'neuronal recycling' hypothesis. In other words, the VWFA would thus develop ontogenetically in preadapted brain areas to process the specific patterns of written linguistic stimuli. In underpinning their proposal, Dehaene and Cohen mention that the frequencies of intersections in writing systems follow a universal and natural frequency distribution, i.e. similar to what is found in natural images [47]. Hence, writing systems seem to follow rules enacted by more general visual capacities. Their treatment in the VWFA would therefore be an exaptation of an initial bias in favor of the recognition and treatment of geometric features that are close to those used in letter shapes: line junctions, by which an object occludes another. This is supported by the fact that the area analogous to VWFA in primates encodes intersections [44].

Szwed et al. [48] have underlined the primary role of line junctions. They investigated brain activations when perceiving objects and words while preserving either vertices or 'midsegments' in their drawing. For both objects and words, it appears that recognition relies

predominantly on line vertices, i.e. where line junctions occur. Activations following the display of stimuli with preserved vertices partially overlap the fusiform gyrus, which is involved in reading, even if it does not imply the VWFA directly.

As recalled by Newman and Twieg [49], a number of word reading studies have shown that '*pseudoword and real word reading tended to activate the same cortical network and that pseudoword reading is more effortful, producing more activation than real word reading*' (p. 39). The VWFA falls into such brain areas, with greater activations for pseudowords than for words. This suggests an implication of this area in a prelexical rather than lexical process [50]. A potential confounding factor is, however, that, as indicated by these authors, pseudowords are also accompanied by slower responses and longer activations. The greater BOLD (blood-oxygen-level dependent) signal observed in fMRI studies may therefore be due to a longer activation, and not a stronger one.

Although the VWFA does not respond to non-linguistic stimuli, Szwed et al. [48] showed that the vision of line junctions activates close neuronal structures in the fusiform gyrus. The spreading of activation to the VWFA that could follow is the possible neuronal basis for the asymmetrical priming effect we proposed earlier. Additionally, the frames used in our study did not result from a random placement of dots and either straight or curved lines as in Nielsen & Rendall [6,14] or in Monaghan et al. [16]. There could therefore be a bias due to the experimenters' involvement in the design of the frames, with features reminiscent of those coded by the fusiform gyrus or even the VWFA.

Transparency of the task and cognitive level of response

Implicit vs explicit protocols. While many studies have highlighted the existence of the bouba-kiki effect, our results did not. A possible explanation is that the implicit nature of our protocol explains the discrepancy with results from association tasks of other experiments. As already explained, tasks which do not explicitly ask the subjects to make associations dissimulate the phonetic and visual contrasts to a greater extent. One can reasonably admit that protocols can be evaluated along a continuum with respect to the transparency of their task. In other words, transparency is not a yes-or-no property. Along such a continuum, our protocol stands as rather opaque compared to others, which would explain the absence of sound symbolic effects.

Less transparent does not mean, however, that participants do not engage in metacognitive reasoning about the task. In our experiment, subjects were asked to perform a lexical decision task, without any reference to the frames or the fonts. Although metacognitive strategies may have taken place regarding the frames, we argue that the differences between the angular and curvy fonts were much less noticed, especially since none of our 41 subjects spontaneously reported that two different fonts were being used.

The discrepancy between our results and Westbury [9]'s remains to be accounted for, since our protocol derived from his and shared his implicitness. A first possibility lies in differences in terms of statistical approaches. In particular, the issue of non-independence was only partially addressed with the by-item and by-subject approaches used by Westbury. Another explanation relates to differences in controlling for the potential confounding factors (number of phonological/orthographic neighbors, preceding response time etc.). The difficulty of our task may be another reason: the contrasted graphemes of our two different fonts could have worked as a cognitive distractor and masked an intrinsically weak sound symbolic effect. Actually, our response times seem to be quite long for a lexical decision task (810 ms for words and 848 ms for pseudowords). For the sake of comparison, response times for a lexical decision task in French [51] are respectively 730 and 802 ms. We argue, however, that these differences are not

due to a greater difficulty of our task because of a lower readability of the fonts. On the one hand, the readability seemed to be equivalent for both fonts, since there was no main effect of the **Font** variable. On the other hand, response times were trimmed in Ferrand et al. [51]'s study, but not in ours. Given the likely right-skewed distribution of response times in the former, this likely explains the differences in mean response times.

Overall, our results support Nielsen and Rendall [14]'s argument that the strength of the bouba-kiki effect is related to the transparency of the testing protocols. In a very opaque procedure, sound symbolic associations, if they exist, may be too weak to be revealed statistically, even with a large number of observations.

Arguments in favor of low-level processes. What are the cognitive processes at play in sound symbolism, and were they underlying our subjects' answers despite the lack of significant sound symbolic interactions? More precisely, what is the 'level' of these processes?

A number of studies are in favor of low-level processes, which occur early and automatically in the processing of stimuli. Vainio, Tiainen, Tiippana, Rantala and Vainio [52] conducted experiments in which subjects were presented with objects differing on two dimensions–shape and size–and requested to produce isolated syllables or vowels according to one of the two preceding dimensions. The effect of the second dimension, which was not relevant to the task, was studied. The authors demonstrated that a spiky shape shortened the reaction time for the vocalization of /i/, /ti/ or /te/, *mostly* when participants correctly categorized the visual stimulus as little. Conversely, a round shape shortened the reaction time for the vocalization of /mɑ/, /me/ and /u/ *only* when participants correctly categorized the visual stimulus as big. These results supported correspondences between articulatory movements and visual features, and demonstrated an implicit impact of a non-relevant modality (shape) on a size-categorization task via an articulatory medium response.

A couple of studies suggest that sound symbolic associations can be detected in early neurophysiological processes. Kovic, Plunkett and Westermann [53] used a paradigm that consisted first in learning labels for two kinds of 'animals'–several exemplars of spiky and round creatures. Labels were either congruent or incongruent with the shape of the animals. In a second task, the four possible types of pairs were presented separately and subjects had to decide which pairs were correct according to the rules they had learned. Participants in the congruent condition responded quicker to congruent pairs than to incongruent ones, while participants in the incongruent condition were slower to reject congruent pairs than to reject incongruent ones. This revealed a bias in favor of sound symbolic pairs, regardless of the learning targets. This behavioral result was replicated in a setting with an ERP recording. A negative wave was found to appear between 140 and 180 ms in occipital regions for congruent pairs, which may indicate multimodal integration.

Asano et al. [54] also found cues of multimodal integration in 11-month-old infants which were presented with different audio-visual bouba-kiki associations. This was suggested by the increase, for congruent trials and between 1 and 300 ms, in the amplitude of oscillations recorded in centro-parietal regions. Additionally, a wave corresponding to N400 in adults–a well-known marker of semantic or conceptual incongruity–was found in central regions for incongruent pairs.

The previous results, and in particular the precocity of the brain activations, raise the question of the underlying physiological and psychological mechanisms for cross-modal correspondences. Spence [55] reviews various proposals, and cites Ramachandran and Hubbard [56]'s proposal that sound symbolic associations are explained by a low-level binding between visual and auditory representations, an instance of the more general phenomenon known as synaesthesia, which links sensory representations belonging to different modalities.

An argument in favor of synaesthesia is the possibility for 4-month infants to consistently map particular linguistic stimuli to particular shapes [7]. Chen, Huang, Woods & Spence [3] also explain differences in bouba-kiki associations between easterners and westerners by synaesthesia and underlying differences in perceptual experience.

Overall, there is thus a strong line of arguments in favor of low-level cognitive processes for sound symbolism, such as the building of low-level connections between sensory domains. One may then wonder why we did not observe significant sound symbolic associations in our study. Indeed, while higher-level processes could be affected by a non-transparent protocol and a time-controlled task, this should not be the case for lower-level ones, which take place during the early stages of the cognitive processing. This contradiction suggests that other processes may be at play in the case of written stimuli.

Synaesthesia, ideasthesia and the specific case of written representations of speech sounds. There are arguments against the previous explanations of sound symbolism in terms of synaesthesia. First, results in very young infants are debated, with experiments failing to reproduce effects found previously in similar populations [57]. Second, what is referred to as different perceptual experiences in Chen et al.'s study could well be different conceptual derivations from the same sensations, because of different cultural experiences and exposures as a whole. Third, some authors have questioned whether the inducer of a synaesthetic relationship belongs to the sensory or to the conceptual domain [58,59]. For example, in time-unit synaesthesia, in which inducers such as weekdays or months are associated with concurrents such as colors, time units are concepts without direct sensory inputs. In grapheme-color synaesthesia, it has also been shown that the assumed meaning of an ambiguous grapheme is what determines the associated synaesthetic colors [60]. Hence, in situations where concepts rather than sensory representations induce sensory activations, the term 'ideasthesia' could be more appropriate than 'synaesthesia' [59]. The latter would then be restricted to situations where only sensory representations are involved. In some cases, 'true' synaesthesia may therefore not be the right explanation for sound symbolic associations, as suggested below.

While we do not argue against synaesthesia in most cases of sound symbolism, we argue that the use of written words or pseudowords, instead of oral inputs, may actually rather constitute a case for ideasthesia, with its own specific features. Indeed, rather than directly accessing a phonetic form upon hearing an acoustic signal, reading linguistic units implies that a sound representation be reconstructed, in the case of pseudowords, or accessed, in the case of words stored in the subject's mental lexicon. This is reminiscent of the case of a conceptual rather than sensory inducer of a synaesthetic relationship with visual shapes, in the bouba-kiki case at least. This is true especially if ones consider internal representations of words or pseudowords to be made of phonemes rather than of phonetic units. Phonemes are indeed based on contrasts, and are therefore to some extent more abstract than acoustic representations–'abstract' is here a better characterization than 'conceptual'.

Such abstract contrastive representations may benefit, or perhaps require, explicit contrasts in order for their phonetic referents to engage in sound symbolic associations. In other words, presentation of two pseudowords or of two words differing along one or a few phonetic dimensions could help to emphasize the phonetic units to be matched by visual representations. In our own study, given the absence of linguistic contrasts–only one pseudoword or word was displayed at a given time–, sound symbolic associations may have been harder to trigger. The time limit to answer during a trial also perhaps prevented some associations that could have developed in the longer run with additional cognitive processing. This could hence explain why we did not see significant sound symbolic effects, while they can be observed in more explicit association tasks implying written linguistic material.

There is a priori no reason to mutually exclude low-level synaesthetic processes and higher-level ideasthetic ones, although how they occur simultaneously is an open question to us. Whether reinforcing or competing effects may take place is an interesting issue, which study would require carefully designed experimental settings to promote the various processes. A broader perspective would also consist in going beyond the opposition between low-level sensory processes and higher-level ones, and advocate for an embodied perspective on sound symbolism, where semiotic processes emerge from sensory representations without the unraveling of an abstracting process.

Conclusion

Our investigation, with a large corpus of data, well-balanced lists of stimuli and rigorous statistical analyses, fails to support sound symbolic associations that we were initially expecting on the basis of previous bouba-kiki studies. Rather, we observed at the visual level the possible consequence of interactions between angular shapes in frames and in letters, but not between round shapes and curvy letters. Beyond explanations of this phenomenon, different conclusions can be drawn regarding sound symbolism.

A first suggestion is that saliency effects and intra-modal correspondences should not be discarded as a possible source of interference when investigating sound symbolism with psycholinguistic experiments. What may appear on the surface as a cross-modal correspondence may indeed turn out to be partly based on phenomena that are not related to sound symbolism. Also, sound symbolic effects may be masked by such phenomena.

A second proposal rests on the existence of different processes leading to sound symbolic associations, with some taking place at a lower level of cognitive processing, for example with crossmodal synaesthetic correspondences, while others rely on more abstract representations and necessitate the right environment to become manifest. This could be the case of ideasthetic processes, especially when written material rather than oral one is involved in the experimental design. Different cases of sound symbolism may thus actually point to differing underlying cognitive processes, and may display different properties upon their respective investigations.

Supporting information

S1 Table. List of pseudowords.
(DOCX)

S2 Table. Properties of the lists of pseudowords.
(DOCX)

S3 Table. List of words.
(DOCX)

S4 Table. Properties of the lists of words.
(DOCX)

S1 Dataset. Dataset for pseudowords.
(XLSX)

S2 Dataset. Dataset for words.
(XLSX)

S1 Structure. Structure of the datasets.
(DOCX)

S1 Protocol. Additional details of the experimental design.
(DOCX)

S1 Analysis. Details of the statistical analysis.
(DOCX)

Acknowledgments

The authors thank Enora Luce-Bassetti for her help with designing and running the experiment, and Sébastien Flavier for his help with the conception of the experimental scripts. They also thank the three reviewers for their very useful and relevant comments and suggestions.

Author Contributions

Conceptualization: Léa De Carolis, Egidio Marsico, Christophe Coupé.

Data curation: Léa De Carolis.

Formal analysis: Vincent Arnaud, Christophe Coupé.

Investigation: Léa De Carolis, Egidio Marsico, Christophe Coupé.

Methodology: Léa De Carolis, Egidio Marsico, Vincent Arnaud, Christophe Coupé.

Software: Christophe Coupé.

Supervision: Egidio Marsico, Christophe Coupé.

Visualization: Christophe Coupé.

Writing – original draft: Léa De Carolis, Egidio Marsico, Vincent Arnaud, Christophe Coupé.

Writing – review & editing: Léa De Carolis, Vincent Arnaud, Christophe Coupé.

References

1. Köhler W. Gestalt Psychology (2nd Ed.). New York: Liveright; 1947.
2. Bremner AJ, Caparos S, Davidoff J, de Fockert J, Linnell KJ, Spence C. "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. Cognition. 2013; 126: 165–72. https://doi.org/10.1016/j.cognition.2012.09.007 PMID: 23121711
3. Chen Y-C, Huang P-C, Woods A, Spence C. When "Bouba" equals "Kiki": Cultural commonalities and cultural differences in sound-shape correspondences. Sci Rep. 2016; 6: 26681. https://doi.org/10.1038/srep26681 PMID: 27230754
4. Davis R. The fitness of names to drawings. A cross-cultural study in Tanganyika. Br J Psychol. 1961; 52: 259–268. https://doi.org/10.1111/j.2044-8295.1961.tb00788.x PMID: 13720232
5. Tarte RD. Phonetic symbolism in adult native speakers of Czech. Lang Speech. 1974; 17: 87–94. https://doi.org/10.1177/002383097401700109
6. Nielsen AKS, Rendall D. The sound of round: Evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. Can J Exp Psychol. 2011; 65: 115–124. https://doi.org/10.1037/a0022268 PMID: 21668094
7. Ozturk O, Krehm M, Vouloumanos A. Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. J Exp Child Psychol. 2013; 114: 173–86. https://doi.org/10.1016/j.jecp.2012.05.004 PMID: 22960203
8. Rogers SK, Ross AS. A cross-cultural test of the Maluma—Takete phenomenon. Perception. 1975; 4: 105–106. https://doi.org/10.1068/p040105 PMID: 1161435
9. Westbury C. Implicit sound symbolism in lexical access: evidence from an interference task. Brain Lang. 2005; 93: 10–19. https://doi.org/10.1016/j.bandl.2004.07.006 PMID: 15766764

10. Knoeferle K, Li J, Maggioni E, Spence C. What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. Sci Rep. Springer US; 2017; 7: 1–11. https://doi.org/10.1038/s41598-017-05965-y PMID: 28717151
11. Aveyard ME. Some consonants sound curvy: Effects of sound symbolism on object recognition. Mem Cognit. 2012; 40: 83–92. https://doi.org/10.3758/s13421-011-0139-3 PMID: 21948332
12. Fort M, Martin A, Peperkamp S. Consonants are more important than vowels in the Bouba-kiki effect. Lang Speech. 2015; 58: 247–266. https://doi.org/10.1177/0023830914534951 PMID: 26677645
13. Ahlner F, Zlatev J. Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. Sign Syst Stud. 2010; 38: 298–348.
14. Nielsen AKS, Rendall D. The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. Lang Cogn. 2012; 4: 115–125. https://doi.org/10.1515/langcog-2012-0007
15. Nobile L. Phonemes as images. An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In: Hiraga MK, Herlofsky WJ, Shinoara K, Akita K, editors. Iconicity: East meets West. Amsterdam: John Benjamins; 2015. pp. 71–91. https://doi.org/10.1075/ill.14.04nob
16. Monaghan P, Mattock K, Walker P. The role of sound symbolism in language learning. J Exp Psychol Learn Mem Cogn. 2012; 38: 1152–1164. https://doi.org/10.1037/a0027747 PMID: 22468804
17. Sidhu DM, Pexman PM. A prime example of the Maluma/Takete effect? Testing for sound symbolic priming. Cogn Sci. 2017; 41: 1958–1987. https://doi.org/10.1111/cogs.12438 PMID: 27766662
18. Crystal D. A Dictionary of Linguistics and Phonetics (6th ed). Oxford: Blackwell; 2008.
19. Cuskley C, Simner J, Kirby S. Phonological and orthographic influences in the bouba–kiki effect. Psychol Res. 2015; 9: 389–397. https://doi.org/10.1007/s00426-015-0709-2
20. Turoman N, Styles SJ. Glyph guessing for 'oo' and 'ee': Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. R Soc Open Sci. 2017; 4. https://doi.org/10.1098/rsos.170882 PMID: 28989784
21. Nielsen AKS, Rendall D. Parsing the role of consonants versus vowels in the classic Takete-Maluma phenomenon. Can J Exp Psychol. 2013; 67: 153–63. https://doi.org/10.1037/a0030553 PMID: 23205509
22. Coltheart M. Modeling Reading: The dual-route approach. In: Snowling MJ, Hulme C, editors. The science of reading. Oxford: Blackwell; 2005. pp. 6–23. https://doi.org/10.1002/9780470757642.ch1
23. Gigerenzer G, Krauss S, Vitouch O. The null ritual. What you always wanted to know about significance testing but were afraid to ask. The Sage Handbook of quantitative methodology for the social sciences. Thousans Oaks, CA: Sage; 2004. pp. 391–408. https://doi.org/10.4135/9781412986311.n21
24. New B, Pallier C, Ferrand L, Matos R. Une base de données lexicales du français contemporain sur internet: LEXIQUE™. Annee Psychol. 2001; 101: 447–462. https://doi.org/10.3406/psy.2001.1341
25. Luce PA, Pisoni DB. Recognizing spoken words: The neighborhood activation model. Ear Hear. 1998; 19: 1–36. https://doi.org/10.1097/MPG.0b013e3181a15ae8.Screening PMID: 9504270
26. Coupé C. BALI: A software tool to build experimental material in psycholinguistics. Proceedings of the Architectures and Mechanisms for Language Processing (AMLaP) Conference 2011 Sept 1–3, Paris, France. 2011.
27. Mathôt S, Schreij D, Theeuwes J. OpenSesame: an open-source, graphical experiment builder for the social sciences. Behav Res Methods. 2012; 44: 314–24. https://doi.org/10.3758/s13428-011-0168-7 PMID: 22083660
28. R Development Core Team. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2017. Available: http://www.r-project.org
29. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009.
30. Wickham H. Reshaping data with the reshape package. J Stat Softw. 2007; 21: 1–20. https://doi.org/10.1016/S0142-1123(99)00007-9
31. Bates D, Maechler M, Bolker BM, Walker S. Fitting Linear Mixed-Effects Models using lme4. J Stat Softw. 2015; 67: 1–48. https://doi.org/10.18637/jss.v067.i01
32. Stasinopoulos M, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing using GAMLSS in R. CRC Press/Taylor & Francis Group; 2017.
33. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. Ecol Monogr. 1984; 54: 187–212. https://doi.org/10.2307/1942661
34. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape (with discussion). Appl Stat. 2005; 54: 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x

35. Rigby RA, Stasinopoulos DM, Lane PW. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat. 2007; 23.

36. Baayen RH, Milin P. Analyzing reaction times. Int J Psychol Res. 2010; 3: 12–28. https://doi.org/10.1287/mksc.12.4.395

37. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979; 6: 65–70. https://doi.org/10.2307/4615733

38. Larson CL, Aronoff J, Stearns JJ. The shape of threat: Simple geometric forms evoke rapid and sustained capture of attention. Emotion. 2007; 7: 526–534. https://doi.org/10.1037/1528-3542.7.3.526 PMID: 17683209

39. Armbruster D, Suchert V, Gärtner A, Strobel A. Threatening shapes: The impact of simple geometric configurations on peripheral physiological markers. Physiol Behav. 2014; 135: 215–221. https://doi.org/10.1016/j.physbeh.2014.06.020 PMID: 24976454

40. Larson CL, Aronoff J, Sarinopoulos IC, Zhu DC. Recognizing threat: A simple geometric shape activates neural circuitry for threat detection. J Cogn Neurosci. 2009; 21: 1523–1535. https://doi.org/10.1162/jocn.2009.21111 PMID: 18823242

41. Aronoff J, Barclay AM, Stevenson LA. The recognition of threatening facial stimuli. J Pers Soc Psychol. 1988; 54: 647–655. https://doi.org/10.1037/0022-3514.54.4.647 PMID: 3367283

42. Bassili JN. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. J Pers Soc Psychol. 1979; 37: 2049–2058. https://doi.org/10.1037/0022-3514.37.11.2049 PMID: 521902

43. Coelho CM, Cloete S, Wallis G. The face-in-the-crowd effect: When angry faces are just cross(es). J Vis. 2010; 10: 1–14. https://doi.org/10.1167/10.1.7 PMID: 20143900

44. Dehaene S, Cohen L. Cultural recycling of cortical maps. Neuron. 2007; 56: 384–398. https://doi.org/10.1016/j.neuron.2007.10.004 PMID: 17964253

45. Baker CI, Liu J, Wald LL, Kwong KK, Benner T, Kanwisher N. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc Natl Acad Sci. 2007; 104: 9087–9092. https://doi.org/10.1073/pnas.0703300104 PMID: 17502592

46. Dehaene S. Evolution of human cortical circuits for reading and arithmetic: The "'neuronal recycling"' hypothesis. In: Dehaene S, Duhamel J-R, Hauser M, Rizzolatti G, editors. From Monkey Brain to Human Brain. Cambridge: MIT Press; 2005. pp. 133–157.

47. Changizi M, Zhang Q, Ye H, Shimojo S. The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. Am Nat. 2006; 167: E117–E139. https://doi.org/10.1086/502806 PMID: 16671005

48. Szwed M, Dehaene S, Kleinschmidt A, Eger E, Valabrègue R, Amadon A, et al. Specialization for written words over objects in the visual cortex. Neuroimage. 2011; 56: 330–344. https://doi.org/10.1016/j.neuroimage.2011.01.073 PMID: 21296170

49. Newman SD, Twieg D. Differences in auditory processing of words and pseudowords: An fMRI study. Hum Brain Mapp. 2001; 14: 39–47. https://doi.org/10.1002/hbm.1040 PMID: 11500989

50. Dehaene S, Le Clec'H G, Poline J-B, Le Bihan D, Cohen L. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport. 2002; 13: 321–325. PMID: 11930131

51. Ferrand L, New B, Brysbaert M, Keuleers E, Bonin P, Méot A, et al. The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudo words. Behav Res Methods. 2010; 42: 488–496. https://doi.org/10.3758/BRM.42.2.488 PMID: 20479180

52. Vainio L, Tiainen M, Tiippana K, Rantala A, Vainio M. Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. Psychol Res. 2017; 81: 827–839. https://doi.org/10.1007/s00426-016-0778-x PMID: 27306548

53. Kovic V, Plunkett K, Westermann G. The shape of words in the brain. Cognition. Elsevier B.V.; 2010; 114: 19–28. https://doi.org/10.1016/j.cognition.2009.08.016 PMID: 19828141

54. Asano M, Imai M, Kita S, Kitajo K, Okada H, Thierry G. Sound symbolism scaffolds language development in preverbal infants. Cortex. Elsevier Ltd; 2015; 63: 196–205. https://doi.org/10.1016/j.cortex.2014.08.025 PMID: 25282057

55. Spence C. Crossmodal correspondences: a tutorial review. Atten Percept Psychophys. 2011; 73: 971–95. https://doi.org/10.3758/s13414-010-0073-7 PMID: 21264748

56. Ramachandran VS, Hubbard EM. Synaesthesia—A window into perception, thought and language. J Conscious Stud. 2001; 8: 3–34.

57. Fort M, Weiss A, Martin A, Peperkamp S. Looking for the bouba-kiki effect in prelexical infants. In: Ouni S, Berthommier F, Jesse A, editors. Proceedings of the 12th International Conference on Auditory-Visual Speech Processing. INRIA; 2013. pp. 71–76.

58. Mroczko-Wąsowicz A, Nikolić D. Semantic mechanisms may be responsible for developing synesthesia. Front Hum Neurosci. 2014; 8: 509. https://doi.org/10.3389/fnhum.2014.00509 PMID: 25191239

59. Nikolić D. Is synaesthesia actually ideasthesia? An inquiry into the nature of the phenomenon. Proceedings of the Third International Congress on Synaesthesia, Science & Art, Granada, Spain, April 26–29. 2009.

60. Dixon MJ, Smilek D, Duffy PL, Zanna MP, Merikle PM. The role of meaning in grapheme-colour synaesthesia. Cortex. 2006; 42: 243–252. https://doi.org/10.1016/S0010-9452(08)70349-6 PMID: 16683498

S1 Table. List of pseudo-words.

Voiced plosives	Voiceless plosives	Sonorants 1	Sonorants 2
abude	acape	amane	aloul
adibe	apipe	anoul	aloum
adude	apute	immal	amil
agade	atupe	immim	aname
badie	catte	imoul	iloum
baga	couk	innim	imale
bigu	cuke	lalla	imane
boube	icute	lami	linni
bougu	ikak	lanou	loume
bube	ikite	linou	louni
buda	ikuk	loula	lula
bugue	ipipe	loune	lulle
dagou	itape	lul	lumue
dide	kipou	lummu	malla
digou	pouke	lumou	mimue
doudi	pouki	malue	mounu
douga	puc	minnu	mune
dubu	pukou	muma	nal
gabe	pupue	munou	namie
gagou	puti	nalle	namue
gouga	puttu	nalli	nanu
gubi	quipe	namme	nilou
gugou	tapou	nannu	ninne
guibe	ticou	noune	noul
ibibe	touca	nune	numue
ibude	toucu	oulim	nunie
idabe	touki	oumul	oulil
idide	toutu	ounul	oumal
igabe	tutou	umam	oumum
ubibe	upipe	umine	uline
ugade	utape	unnim	umane
ugude	utate	unum	umoum

S2 Table. Properties of the lists of pseudo-words.

	Voiced plosives	Voiceless plosives	Sonorants 1	Sonorants 2
Structure (count)				
CVC	6	6	6	6
CVCV	14	13	13	13
VCVC	12	13	13	13

Nb of letters				
Mean	4,75	4,78	4,78	4,78
Min	4	3	3	3
Max	5	5	5	5
Nb of phonemes				
Mean	3,81	3,81	3,81	3,81
Min	3	3	3	3
Max	4	4	4	4
Nb of orthographic neighbors				
Mean	3,38	3,22	3,28	3,28
Min	0	0	0	0
Max	15	18	12	18
Nb of phonological neighbors				
Mean	7,84	8,59	8,19	8,16
Min	0	0	0	0
Max	25	56	24	31
Average frequency of phonological neighbors				
Mean	5,35	4,95	5,21	5,23
Min	0,00	0,00	0,01	0,00
Max	57,16	42,64	26,83	40,26
Average frequency of orthographic neighbors				
Mean	2,66	3,33	3,54	3,26
Min	0,00	0,00	0,00	0,00
Max	40,68	43,43	50,60	41,28
Maximum frequency of phonological neighbors				
Mean	27,80	70,86	74,15	56,17
Min	0,00	0,00	0,01	0,00
Max	289,63	573,72	453,95	537,44
Maximum frequency of orthographic neighbors				
Mean	16,71	16,53	18,83	11,42
Min	0,00	0,00	0,00	0,00
Max	341,35	349,32	349,46	143,45

S3 Table. List of words.

Plosives	Sonorants	Mixed 1	Mixed 2

agape	anime	amibe	aneth
agate	laine	anode	atoll
audit	lama	atome	atone
bader	lame	autel	banni
bague	lime	balai	bilan
battu	limer	bonne	bile
bec	lino	canot	caler
bidet	lune	coma	connu
biper	mamie	demi	culot
cabot	manie	dune	donne
cadet	manne	gaine	galop
coque	mener	gamma	gomma
coter	menu	goule	gone
coupe	mille	idole	goulu
daube	mime	item	hamac
digue	mimer	kilo	idem
dodo	minet	laque	imite
dodu	mini	ligue	lac
duc	minot	lubie	laide
duper	minou	lutte	lobe
gober	molle	mater	loti
godet	moule	matou	meute
gouda	moulu	midi	noter
otite	mule	nappe	nuque
papi	mulot	nette	patte
petit	muni	obole	peine
picot	naine	opine	peler
pub	nomme	paume	poney
tabou	nonne	piler	poule
tague	nul	polo	puma
tipi	nulle	puni	tamis
tique	ulule	tenu	utile

S4 Table. Properties of the lists of words.

	Plosives	Sonorants	Mixed 1	Mixed 2
Structure (count)				
CVC	10	15	10	11
CVCV	18	15	14	14
VCVC	4	2	8	7
Nb of letters				
Mean	4,69	4,63	4,72	4,75
Min	3	3	4	3
Max	5	5	5	5

Nb of phonemes				
Mean	3,69	3,53	3,69	3,66
Min	3	3	3	3
Max	4	4	4	4

Nb of orthographic neighbors				
Mean	8,34	12,25	8,50	9,13
Min	1	3	0	0
Max	19	25	20	23

Nb of phonological neighbors				
Mean	16,28	19,34	17,19	17,91
Min	1	5	2	3
Max	37	34	35	55

Average frequency of phonological neighbors				
Mean	13,02	26,24	20,35	22,45
Min	0,08	0,05	0,01	0,08
Max	99,33	180,04	196,66	143,46

Average frequency of orthographic neighbors				
Mean	21,16	33,69	25,50	35,94
Min	0,22	0,12	0,00	0,00
Max	347,75	453,51	475,49	968,89

Maximum frequency of phonological neighbors				
Mean	413,48	1461,11	902,11	2008,79
Min	0,47	0,41	0,02	0,24
Max	4394,70	14946,48	14946,48	18188,15

Maximum frequency of orthographic neighbors				
Mean	290,64	531,83	398,44	782,32
Min	0,34	0,14	0,00	0,00
Max	6882,16	9587,97	9587,97	23633,92

Word frequency in books				
Mean	25,11	12,72	19,00	21,76
Min	0,00	0,07	0,00	0,14
Max	653,78	142,09	294,53	388,24

S1 Structure. Structure of the datasets.

Dataset for pseudowords:

- **Subject**: Subject ID (integer)
- **Stimulus**: Written form displayed on screen (string)
- **PhonologicalForm**: Phonological form corresponding to the written form, according to the convention used in *Lexique 3.81* (string)
- **Frame**: Frame ID, e.g. ‘Curve4’ or ‘Spike17’ (string)
- **FrameType**: Type of Frame, either ‘Curvy’ or ‘Spiky’ (string)
- **OccOrSon**: Whether the consonants are both ‘plosive’ or both ‘sonorant’ (string)
- **Voicing**: Whether the consonants are both ‘voiced’ or both ‘voiceless’ (string)
- **ConsonantCat**: Categoy of the consonants, ‘sonorant’, ‘voiced_plosive’ or ‘voiceless_plosive’ (string)
- **TrialPosition**: position of the trial during the subject’s test (integer, from 1 to 256)
- **Font**: Font used for the written form, either ‘Gabriola’ or ‘Agency FB’ (string)
- **ResponseTime**: Response time in ms (double)
- **SOA**: Stimulus-onset asynchrony in ms (integer)
- **PrecedingResponseTime**: preceding response time in ms (double)
- **Structure**: Structure of the pseudoword, ‘CVC’, ‘CVCV’ or ‘VCVC’ (string)
- **Consonants**: Consonants used to build the pseudoword (string)
- **Vowels**: Vowels used to build the pseudoword (string)
- **NbLetters**: Number of letters in the pseudoword (integer)
- **NbPhonemes**: Number of phonemes in the pseudoword (integer)
- **StimuliList**: List of stimuli to which the pseudoword belongs (integer)
- **NbPhon**: Number of phonological neighbors (integer)
- **NbOrtho**: Number of orthographic neighbors (integer)
- **AvFrPhon**: Average frequency of occurrence of the phonological neighbors (double)
- **AvFrOrtho**: Average frequency of occurrence of the orthographic neighbors (double)
- **MaxFrPhon**: Maximum frequency of occurrence of the phonological neighbors (double)

- **MaxFrOrtho**: Maximum frequency of occurrence of the orthographic neighbors (double)
- **MedFrPhon**: Median frequency of occurrence of the phonological neighbors (double)
- **MedFrOrtho**: Median frequency of occurrence of the orthographic neighbors (double)
- **Gender**: Subject's gender, either 'M' or 'F' (string)
- **Laterality**: Subject's handedness, either 'L' or 'R' (string)
- **LateralityScore**: Subject's laterality score, from -100 (left-handed) to 100 (right-handed) (double)

Dataset for words:

- **Subject**: Subject ID (integer)
- **Stimulus**: Written form displayed on screen (string)
- **PhonologicalForm**: Phonological form corresponding to the written form, according to the convention used in *Lexique 3.81* (string)
- **Frame**: Frame ID, e.g. 'Curve4' or 'Spike17' (string)
- **FrameType**: Type of Frame, either 'Curvy' or 'Spiky' (string)
- **OccOrSon**: Whether the consonants are both 'plosive', both 'sonorant' or 'mixed' (string)
- **Voicing**: Whether the consonants are both 'voiced', both 'voiceless' or 'mixed' (string)
- **ConsonantCat**: Categoy of the consonants, 'sonorant', 'voiced_plosive', 'voiceless_plosive' or 'mixed' (string)
- **TrialPosition**: position of the trial during the subject's test (integer, from 1 to 256)
- **Font**: Font used for the written form, either 'Gabriola' or 'Agency FB' (string)
- **ResponseTime**: Response time in ms (double)
- **SOA**: Stimulus-onset asynchrony in ms (integer)
- **PrecedingResponseTime**: preceding response time in ms (double)
- **Structure**: Structure of the word, 'CVC', 'CVCV' or 'VCVC' (string)
- **Consonants**: Consonants used to build the word (string)

- **Vowels**: Vowels used to build the word (string)
- **voicing**_1st_consonant: Voicing of the first consonant, either 'voiced' or 'voiceless' (string)
- **voicing**_2nd_consonant: Voicing of the second consonant, either 'voiced' or 'voiceless' (string)
- **NbLetters**: Number of letters in the word (integer)
- **NbPhonemes**: Number of phonemes in the word (integer)
- **FrMovies**: Frequency of occurrences in movies (double)
- **FrBooks**: Frequency of occurrences in books (double)
- **StimuliList**: List of stimuli to which the word belongs (integer)
- **NbPhon**: Number of phonological neighbors (integer)
- **NbOrtho**: Number of orthographic neighbors (integer)
- **AvFrPhon**: Average frequency of occurrence of the phonological neighbors (double)
- **AvFrOrtho**: Average frequency of occurrence of the orthographic neighbors (double)
- **MaxFrPhon**: Maximum frequency of occurrence of the phonological neighbors (double)
- **MaxFrOrtho**: Maximum frequency of occurrence of the orthographic neighbors (double)
- **MedFrPhon**: Median frequency of occurrence of the phonological neighbors (double)
- **MedFrOrtho**: Median frequency of occurrence of the orthographic neighbors (double)
- **Gender**: Subject's gender, either 'M' or 'F'(string)
- **Laterality**: Subject's handedness, either 'L' or 'R' (string)
- **LateralityScore**: Subjec's laterality score, from -100 (left-handed) to 100 (right-handed) (double)

In both datasets, **FrMovies, FrBooks**, **NbPhon**, **NbOrtho**, **AvFrPhon**, **AvFrOrtho**, **MaxFrPhon**, **MaxFrOrtho**, **MedFrPhon**, **MedFrOrtho**d are given in, or computed from, *Lexique 3.81*.

S1 Protocol. Additional details of the experimental design.

Exclusion criteria for words

For words, we avoided all the sequences of graphemes with diacritics and implying the pronunciation of one of four nasal French vowels – [ɛ̃, œ̃, ɔ̃, ɑ̃]. For pseudo-words, the same constraints applied but we additionally limited the selection of vowels to [a, i, y, u]. We avoided the use of the written vowels "e" and "o" because of cross-country variations in their pronunciation [61]. The letter "e" was however used at the end of some strings when it was mute and helped produce "natural-looking" pseudo-words (e.g. "dide" – [did]). This also made pseudo-words slightly longer on average, although this was not a wanted outcome of the process.

Keys to provide answers

Two keys were used by subjects to enter their choice whether a word or a pseudo-word was displayed: one on the left of the AZERTY keyboard ("q"), one on the right ("l"). In a 'right-oriented' version, "l" was used to answer "word" and "q" to answer "pseudo-word". The 'left-oriented' version was the opposite. Participants were asked to choose which version they preferred.

Constraints on the order of presentation of the stimuli

The order of presentation of the stimuli was constrained to avoid repetition effects as follows:

- no more than seven occurrences of one type of frame (i.e., spiky or rounded) in a row;
- no more than two occurrences of the same frame one after the other;
- no more than six words in a row, and no more than six pseudo-words in a row;
- no more than six exemplars of a given category of consonants in a row.

S1 Analysis. Details of the statistical analysis.

Trimming of the response times

Response times for both pseudo-words and words did not follow a Gaussian distribution, were bounded to the left and skewed in favor of longer response times (see Fig 1). It therefore did not make

sense to drop response times distant by more than 2.5 or 3 times the standard deviation from the mean response time, as this would have mostly trimmed longer response times, and hardly any shorter ones. This would have erased potentially important information contained in the thick right tail of the distribution, and would have possibly hidden some relevant effects.

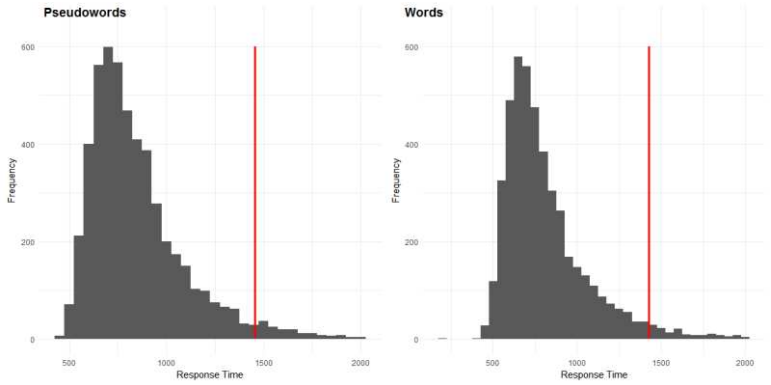


Fig 1. Non-Gaussian distribution of response times for pseudo-words and words. Entries on the right of the red vertical line are distant by more than 2.5 times the standard deviation from the mean response time.

Inclusion of additional predictors in the models

Baayen and Milin [36] have provided suggestions considering the appropriate modelling of response times. In particular, they have considered the trial position and the response time of the preceding trial as possible predictors, and '*found that including variables such as Trial and Preceeding RT in the model not only avoids violating the assumptions of linear modeling, but also helps improving the fit and clarifying the role of the predictors of interest*'. More precisely, including these effects help prevent temporal patterns of correlation of response times.

We therefore chose to consider these two predictors in our regression models, in addition to **Font**, **Type of frame**, **Category of consonants** and their interactions. For a subject's first answer or after a failure to answer within 2000ms, the preceding response time was replaced by their average response time during the experiment.

Choosing an appropriate approach to model response times

Inadequacy of a linear mixed effect model

The most straightforward regression modelling approach to response times is to consider a linear (mixed) model to relate them to predictors. However, an analysis of the residuals of such a model shows that the required assumptions of normality and homoscedasticity of the residuals are violated, as seen in Figure 2. The outputs of the model are therefore not to be trusted, even if it is robust to some extent to such problems.

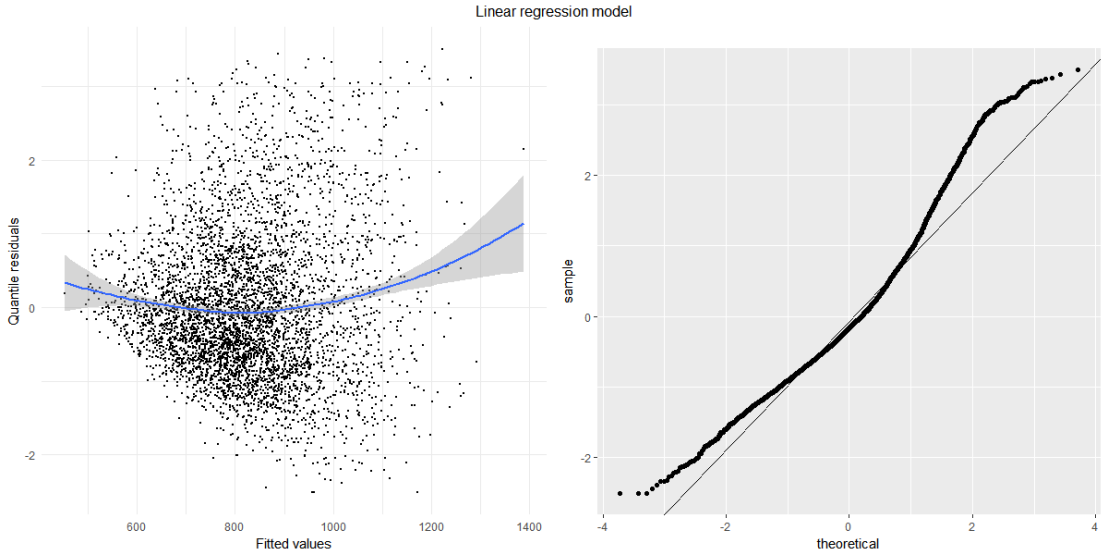


Fig 2. Distribution of the residuals of a linear regression model for pseudo-words against the linear predictor (left) and quantile-quantile plots of these residuals (right). The left panel displays the heteroscedasticity, and the right panel the non-normality of these residuals.

A common solution to this issue is to apply a logarithmic or inverse transformation to the response times [62–65]. The resulting variable then often presents a Gaussian profile, which makes it fit for linear regression. However, as explained by Lo & Andrews [66], this approach is problematic, because '*statistically significant differences on the transformed metric are uninformative as to whether significant differences exist on the original untransformed metric and vice versa*' (p. 3). In other terms, the significance of a predictor with respect to the logarithm or inverse of response times do not tell us about the significance of the relationship between this predictor and the untransformed response times.

All in all, linear regression models are therefore not well suited to response times.

Shifting from linear to generalized linear mixed effect models

A possible solution to avoid transformation of the dependent variable consists in relying on generalized linear mixed regression models (GLMM), which offer more appropriate modelling of non-Gaussian distributions of the error terms, as well as a link function to relate the linear combination of predictors to the observed response. As suggested by Lo & Andrews, the *inverse Gaussian* and *Gamma* distributions make sense at a conceptual level for response times, as they adequately describe the time it takes for an event of interest to occur – pressing a key to answer in our case. Additionally, they advise to choose an *identity* link function – i.e., no transformation – to reflect the fact that models in mental chronometry directly link response times to mental processes.

We therefore first considered the ***glmer()*** function of the `lme4` package, since it provided the *inverse Gaussian* (IG) and *Gamma* (GA) distributions to test with our data. We however experienced convergence issues, which given attempts with other datasets seemed to stem from the combination of these distributions with an *identity* link function. This led us to shift to generalized additive models for location, scale and shape (GAMLSS), as offered in the `gamlss` package, which did not suffer from such problems, and also allowed to consider a much wider range of distributions for error terms.

Generalized additive models for location, scale and shape

Generalized additive models for location, scale and shape (GAMLSS)[32,34,35] are an extension of generalized additive mixed models (GAMM) which allow to consider a wide range of options for the conditional distribution of the dependent variable (which corresponds to the distribution of error terms), while GLMM and GAMM are restricted to the exponential family of distributions [34]. Distributions offered in the `gamlss.dist` package differ on the number of parameters which can be modelled – up to four. These parameters are classically noted μ, σ, ν and τ, and correspond respectively to the location, the scale and the shape (the last two parameters) of the distribution. They are related, though not always equal, to the four moments of a distribution: mean, variance, skewness and

kurtosis. They can be modelled, either with linear parametric, non-linear parametric or non-parametric (smooth) functions of the predictors.

As for the *Poisson* distribution for example, the only parameter that can be modelled is the location parameter, which is equal to the mean of the distribution. The scale and shape of the distribution cannot be modelled independently, since in a *Poisson* distribution the variance is equal to the mean, the skewness to the square root of the mean, and the excess kurtosis (the kurtosis minus 3) to the inverse of the mean. In the well-known Gaussian distribution, the mean and the variance of the distribution are independent from each other, and can be modelled separately, while the skewness and kurtosis are fixed.

We relied on GAMLSS to analyze the response times of our experiment and find an appropriate distribution for the location parameter, and left aside modelling options such as smooth terms. We modelled random effects with a specific smoothing function, in which a local maximum likelihood estimation is performed to shrink the fitted values of the factor predictor to the overall mean [35].

As previously, we first considered *IG* and *GA* distributions and followed the trimming procedure described in the methodological section. Although better than what was observed with a *Gaussian* distribution (NO), residuals were still not adequate enough to consider the adoption of either distribution, this for both pseudo-words and words. It appeared that the problem had likely to do with the strong skewness of the distribution of response times. This led us to envisage other distributions, and especially the Generalized Gamma (GG) distribution, a 3-parameter distribution of which the *IG* and *GA* distributions are two specific instances, and the 4-parameter Johnson's SU (JSU) distribution.

Table 1 and Table 2 summarize the adequacy of various distributions for pseudo-words and words, respectively. Figure 3 and Figure 4 display the corresponding quantile-quantile plots of the residuals. For both pseudo-words and words, the lowest AIC was obtained with the GA distribution. Normalized quantile residuals of this distribution, however, did not closely follow a normal distribution, as it was also the case for the IG distributions. The models with the GG and JSU distributions had higher AIC but a near-normal distribution of residuals. Among the two, the GG distribution led to a lower AIC, again

both for pseudo-words and words, and we therefore chose it as our target distribution, to be reported in the article. We however investigated the output of all models, and always found similar results for the **Type of Frame** x **Font** interaction depicted in the results of this study, although sometimes significance was not reached. This was a solid argument in favor of the existence of this interaction, beyond the singularity of a given model and a given dataset. Other interactions were significant in the JSU model, but did not match our hypotheses with respect to sound symbolism. The **Type of Frame** x **Category of Consonant** interaction found in the GG model for words was absent in the IG and GA models, and was unsupportive of sound symbolic hypotheses too.

Table 1. Number of parameters, number of trimmed observations, global deviance, used degrees of freedom and AIC for GAMLSS models for pseudo-words with various distributions (same predictors and predicted values).

Distribution	*# parameters*	*# deleted observations*	*Global deviance*	*df*	*AIC*
inverse Gaussian (IG)	2	101	63,755	168.7	64,092
Gamma (GA)	2	124	63,551	168.7	63,889
Generalized Gamma (GG)	3	65	64,374	172.0	64,374
Johnson's SU (JSU)	4	45	65,080	177.3	65,435

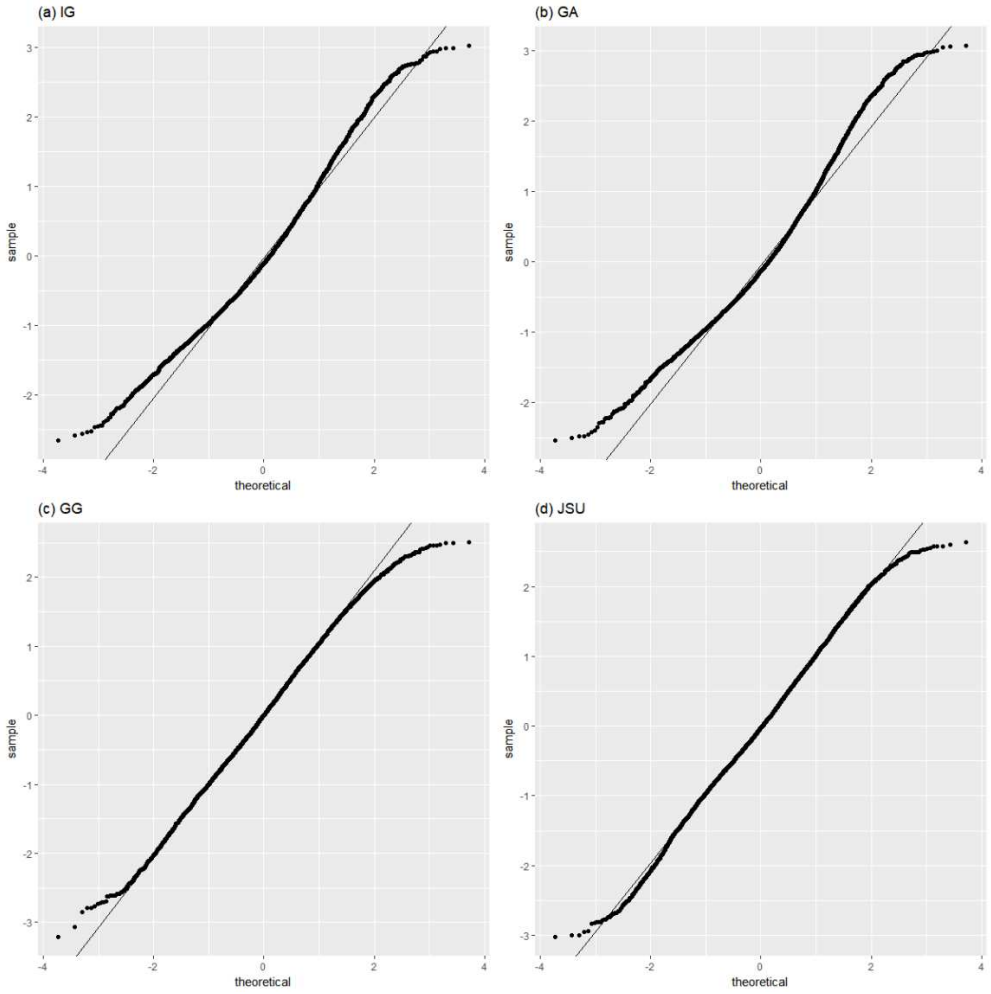


Fig 3. Quantile-quantile plots of residuals for models for pseudo-words with various distributions: IG (a), GA (b), GG (c) and JSU (d).

Table 2. Number of parameters, number of trimmed observations, global deviance, used degrees of freedom and AIC for GAMLSS models for words with various distributions (same predictors and predicted values).

Distribution	*# parameters*	*# deleted observations*	*Global deviance*	*df*	*AIC*
inverse Gaussian (IG)	2	102	56,957	175.7	57,308
Gamma (GA)	2	124	56,772	176.2	57,125
Generalized Gamma (GG)	3	43	57,996	177.4	58,350
Johnson's SU (JSU)	4	36	58,420	176.2	58,772

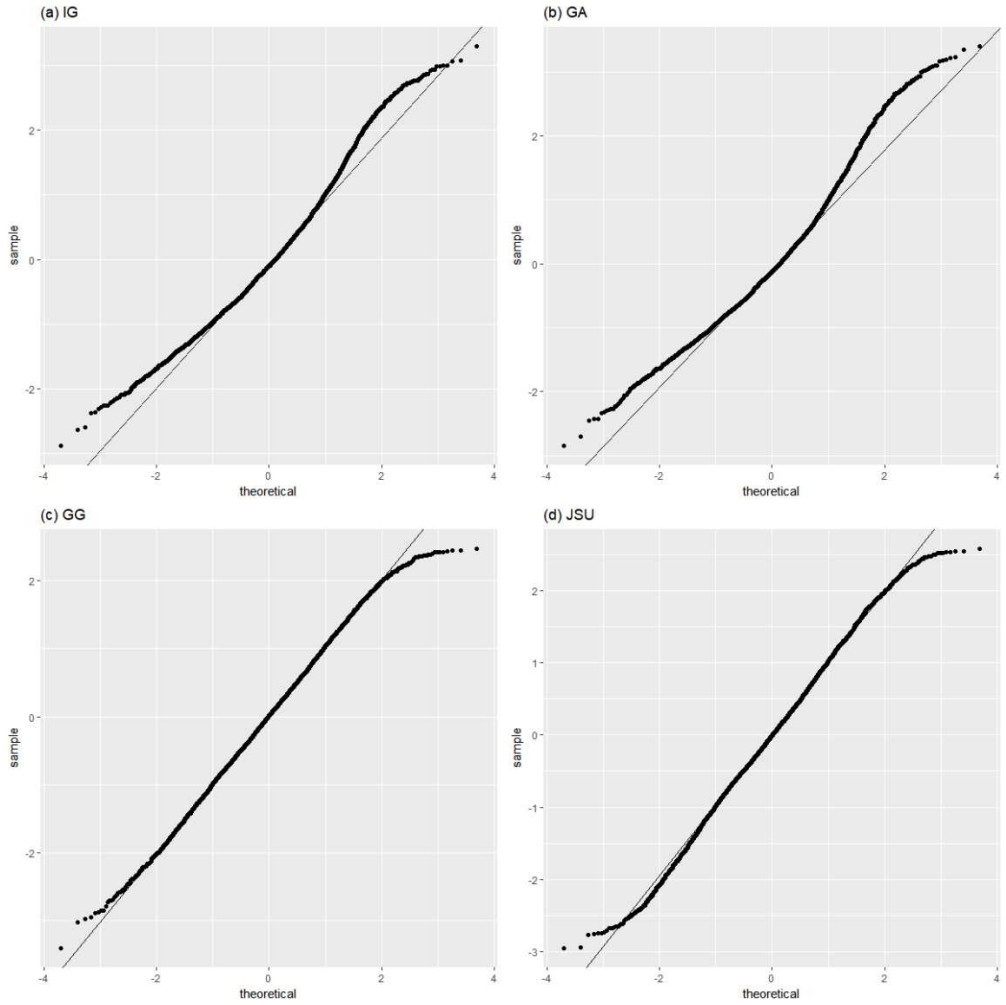


Fig 3. Quantile-quantile plots of residuals for models for words with various distributions: IG (a), GA (b), GG (c) and JSU (d).

Checking the assumptions of the regression models

In addition to the normality of the residuals, other assumptions must be satisfied for a model to be valid: homoscedasticity of the residuals, linearity of the continuous fixed effects, absence of strong multicollinearity, and normal distribution of the modes of each random effect. We checked them for the GG GAMLSS models for pseudo-words and words.

As an illustration, Figure 4 displays residuals of the model for pseudo-words against the linear predictor to assess homoscedasticity. Figure 5 provides the quantile-quantile plots for the modes of the three random effects of the model for pseudo-words. Finally, Figure 6 allows to assess the linear relationship between response times and both **Trial Position** and **Preceding Response Time**.

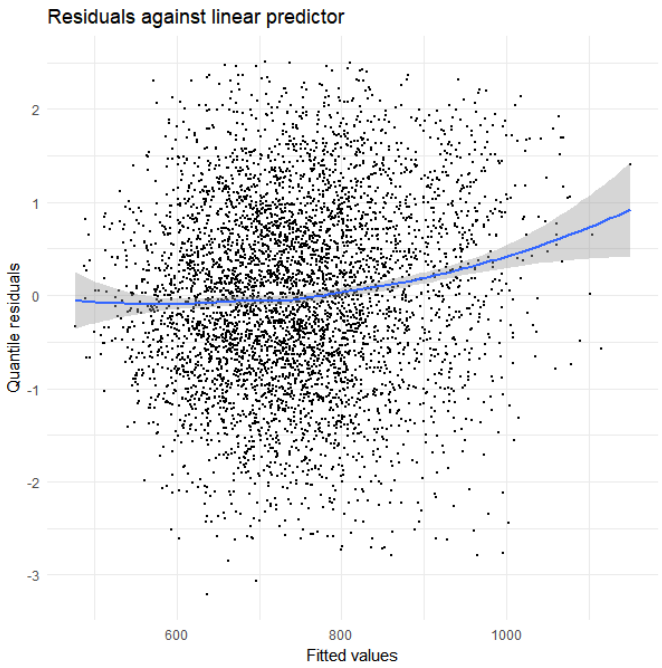


Fig 4. Residuals of the GG GAMLSS model for pseudo-words against the linear predictor.

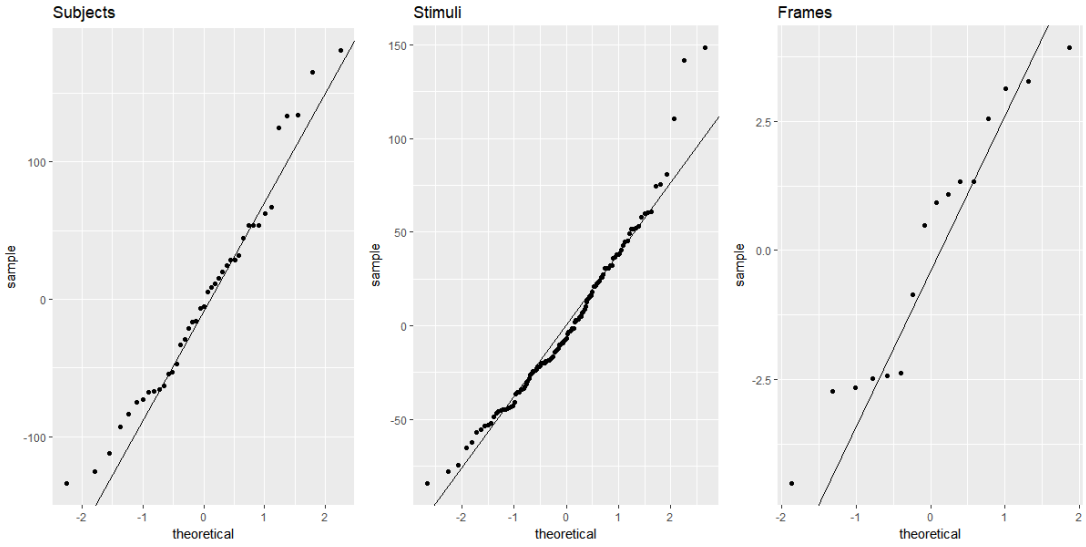


Fig 5. Quantile-quantile plots for the modes of the three random effects of the GG GAMLSS model for pseudo-words. From left to right, the modes of the **Subject**, **Stimulus** and **Frame** random effects are displayed, respectively.

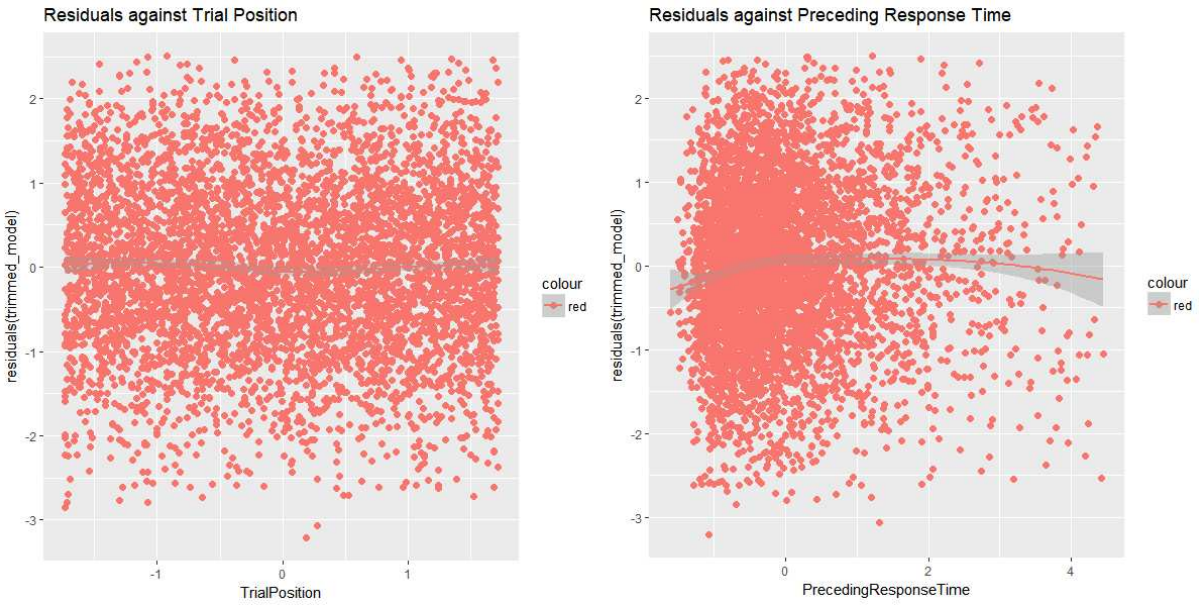


Fig 6. Residuals of the GG GAMLSS model for pseudo-words against Trial Position (left) and Preceding Response Time (right).

Assessing significance in GAMLSS models

As suggested by Stasinopoulos et al. [32], we relied on a series of Likelihood Ratio Test (LRT) to assess the significance of the predictors of the model, and in particular the significance of the three interactions of interest, namely **Type of Frame × Font**, **Type of Frame × Category of consonant**, **Font × Category of Consonant**. For each predictor, the deviance of the full model was compared to the deviance of the nested model without the predictor, testing the hypothesis that the two models have the same likelihood with the assumption that the difference of deviances is approximately χ^2 distributed. Dropping the target predictor without controlling for amount of shrinkage of the random effects would have led the nested model having a lower deviance than the full model, but random effects with higher degrees of freedom. In order to avoid this bias and produce correct differences between the two models in terms of degrees of freedom, we fixed the equivalent degrees of freedom of the random effects in the nested model to the values found in the full model. Doing so led to correct differences in degrees of freedom.

A predictor that appears to be significant must be interpreted cautiously if it is part of a significant higher-order interaction. Indeed, it is uneasy to interpret the effect of a variable when the size of this effect depends on the values of another variable – simple effects must replace main effects in this

case. Because of this, we first assessed the significance of the triple interaction **Type of Frame × Font × Category of Consonant**. We had to verify that it was not significant to drop it and consider double interactions.

Discussion

1. Summary and additional comments on the three studies

1.1. *First study*

The main aim of the first study was to assess whether animals could be good candidates for eliciting *motivated* associations, on the basis of the assumption that animals may have had represented an important component of human communication at its onset (e.g., threats, sources of food). Besides, this type of stimuli represented a more ecological approach in comparison to spiky and round shapes (bouba-kiki tasks) and was a way for indirectly evaluating a potential remnant of early communications.

Many controls were made in order to avoid confounding effects: 1) pictures of animal were presented in levelled shades of grey against a white background; 2) each pair, contrasting one modality (e.g. size), was controlled with respect to other parameters (i.e. repulsiveness, dangerousness and biological class); 3) animals were presented in various sizes, respecting likely mental representations of their respective sizes. These controls were made possible by preliminary assessments collected from participants via online surveys, in which pictures had to be evaluated along these various parameters. The contrasted modalities were randomly presented, and filler pairs were also added in order to further mask the aim of the study, and thus to undertake a more implicit approach in comparison to most studies. Sixty-four target pseudo-words were generated in VCVC form, which permitted almost every combination of the vowels and consonants selected in this study [i, a, u, y, p, b, t, d, k, g, f, v, s, z, ʃ, ʒ, ʁ, m, n, l] with a syllabic reduplication (e.g. [ipip]). The matching between one pseudo-word and a pair of animals was random, but the analyzes were on specific segments for each conceptual contrast.

After correction for multiple tests no significant association appeared. Two major explanations were proposed: either the type of presentation – 2x1 – was not appropriate to bring to light *motivated* associations (i.e. because of the absence of a segmental contrast), or the associations were masked by the multi-dimensionality of the pictures of animals, despite the controls that were applied.

A second experiment was conducted in order to determine which one of these two possibilities could have led to an absence of associations. In this respect, a paper-and-pencil

task was conducted using labels instead of pictures of animals (e.g. 'a small animal'), while preserving the type of presentation (2x1). In such circumstances, seven hypotheses were significant after correction, among the nine that were tested. On this basis, it seems that: 1) this type of presentation is appropriate to bring to light *motivated* associations (as it has been shown in other studies, e.g. Aveyard, 2012); 2) the VCVC structure is also appropriate (as shown before in Fort et al., 2015); 3) pictures of animals are too complex because of their multi-dimensionality. More importantly, the study brings evidence that using features of animals is a relevant approach to the study of *motivation*, namely biological class (fish and birds), emotional aspects (dangerousness and repulsiveness) and size. Moreover, an analysis of the French lexicon revealed that the associations found were not explained by specific frequencies of segments in French animal names according to the concepts that were evaluated. Indeed, no contrast related to these concepts led to statistically significant differences in frequency of occurrence (e.g. there is no more [i] in names of birds than in names of fish).

1.1.1. Cross-linguistic approach to motivation

Animal names in Huambisa (at least those denoting birds and fish) present specific frequencies of segments, on which foreign students rely on for making a choice (Berlin, 1994). In parallel, presenting pseudo-words constructed on the basis of these specific frequencies to French speakers (in our study) led to similar patterns of associations between segments and the biological classes of bird and fish. Concomitantly, however, these frequencies do not appear in the French lexicon of animal names. This led us to an assumption about language evolution.

First, early communications during prehistory would have taken advantage of *motivated* signs to easily and efficiently express important meanings (e.g., food, danger). Then, a later language could have preserved these forms of communication, or not, depending on its speakers' lifestyle. Indeed, Huambisa speakers live in the Amazonian jungle, they are surrounded by multiple wild species and have persistent practices of hunting and fishing, whereas French speakers are part of the Western World, in which industrialization and changes in modes of consumption may have lessened the constrains on communication about 'primordial' needs and threats.

Assuming that languages may follow different diachronic scenarios depending on speakers' lifestyle, we hypothesized that there would be more *motivated* and systematic relations in languages spoken by people keeping a traditional way of living, closer to nature, in comparison to languages spoken by people living in more industrialized countries, with a more distant relation to nature and wildlife.

A way to assess this assumption is to analyze the frequencies of the segments composing words, for different categories (e.g. names of bird or fish, to follow Berlin's study), in two groups of languages: one group of languages spoken by speakers with a 'traditional way of life', and another group of languages spoken in industrialized countries. This work is in progress but was not ready yet to be included in this thesis. However, the difficulties that it raises are exposed here.

There are two possible paths to evaluate frequencies of occurrence within one language. The first one is to look for specific segmental frequencies based on expectations, i.e. on the basis of previous studies. However, some segments may not be relevant for some languages, while others may. This solution may thus be misleading, in accordance with Styles and Gawne (2017)'s conclusions, i.e. the necessity to fit the segmental composition of a language. The second path is to try to detect 'abnormal' frequencies without expectancies regarding particular segments. However, this approach leads to a major problem, namely the inflation of Type I errors (false positives) when conducting multiple statistical tests. Accounting for this issue leads to more stringent criteria to reach significance for each test, and therefore to an increased number of false negatives, i.e. an increased likelihood of failing to detect interesting results. This difficulty has been highlighted by Monaghan et al. (2012)'s study: testing 18 linguistic features with 18 different tests led to two significant results, which were, however, no longer significant after correction. Additionally, in this study, features were perhaps not the optimal level at which associations should have been searched for. It can be recalled that Knoeferle et al. (2017) reported, for example, different effects of voicing depending on manner (voiced plosives are more associated to round than voiced fricatives). As a result, to test all the segments of a given language with multiple tests is problematic (even more than testing all features since there are more segments than features), as the classical way to protect oneself from false positives fails to deliver any positive output as soon as they are more than a few segments considered (unless some motivated associations are very strongly apparent in the target lexicon). We are, however, trying to address this issue with another approach, namely penalized regression models, where the penalization process bypasses the need to conduct multiple statistical tests. A penalized logistic regression model can thus be used to predict a category (e.g. bird or fish) between two features given the absence or presence of many features – which are the predictors of the model.

Our first study underscored the influence of methodological choices. While pictures of animals were too complex to induce *motivated* associations, labels pointing at conceptual

features, such as big, dangerous, bird, etc., led to several significant results. However, some aspects differed between the two experiments of this first study. In the second experiment, the pseudo-words were written (respecting the orthographical conventions of French) rather than auditory stimuli, and the purpose of the task was less implicit, since the conceptual contrasts were explicitly presented. It is quite likely that these two parameters were what enhanced the significance and effect size of the associations. It is also possible that they led to overestimated effects because of the transparency of the task, as previously proposed by Nielsen and Rendall (2012) for bouba-kiki effects. Finally, another aspect was not assessed in this study, namely the possible interactions between vowels and consonants.

The second study therefore moved towards further assessing the impact of methodological differences across studies and aimed at providing some answers regarding these issues.

1.2. *Second study*

Studies about *motivated* associations differ between each other on multiple aspects: the population of participants, the segments composing the linguistic stimuli, the concepts to be associated with these stimuli (even though most of them are about round and spiky shapes), and, more interestingly to us, the type of presentation. Building on the wide methodological variety across studies, the aim of the second study was to investigate the impact of the type of presentation, while preserving the same population, segments and concepts. More precisely, it aimed at assessing the role played by the phonetic and conceptual contrasts in the participants' answers in terms of *motivated* associations. Four protocols were thus assessed with zero, one or two contrasts: 1x1, 2x1, 1x2, 2x2. This study originated in the previous one, since like it, it involved associations about animal features (i.e. dangerousness, repulsiveness, size and biological class). More precisely, the same 10 different associations were assessed with each protocol, each association corresponding to an oriented hypothesis to be confirmed or not with participants' answer. It addressed some methodological issues about the second experiment: 1) in lieu of a paper-and-pencil task, it was computerized, allowing to control presentation, randomize trials and measure response times; 2) pseudo-words were orally presented; 3) the different concepts and segments – either alone or in contrast – were randomly presented in order to modestly hide what was being studied; 4) the analyzes accounted for both the main effects of segments (vowels and consonants) and interactions between them. Only a small selection of pseudo-words was proposed per conceptual contrast.

Results revealed that: 1) there were no contradicting results across protocols; 2) two protocols led to higher numbers of significant effects and higher effect sizes: 1x2 and 2x2, which suggests a beneficial effect from the presence of a phonetic contrast when it comes to highlighting motivated associations; 3) there were only two interaction effects between vowels and consonants; 4) vowels seemed to have overall more impact than consonants. The last two points have, however, to be considered cautiously, since only a few segments were used in this study, and it is possible neither to generalize the stronger effect of vowels, nor to minimize the potential impact of interactions between consonants and vowels.

The fact that no result went against the hypotheses means that the type of presentation did not influence the orientation of the associations made by participants. However, the significance and size of the effects varied depending on the type of presentation. There are several explanations about these differences. The most 'efficient' protocol was 2x2, followed by 1x2. It seems to suggest the preeminent role of phonetic contrasts, since 1x1 and 2x1 do not contain one. However, this may also be explained by the presence of *both* contrasts. Indeed, in 2x2, the conceptual contrast is undoubtedly present. As for 1x2, the conceptual contrast is somehow also present, because labels are inherently linked to their opposite (at least for dangerousness, repulsiveness and size contrasts). For example, the label 'a small animal' is intrinsically opposed to its opposite 'a large animal'.

All in all, some associations are well confirmed throughout the different protocols (e.g. between vowels and size), while some others are more questionable (e.g. those about biological class), since less associations appeared to be significant.

The 1x1 protocol led to the least number of significant associations (four), and this may be explained either by the fact that it is the most implicit task – since there was no contrast within a trial – or by the difference in terms of cognitive processing at play, i.e. a judgement instead of a choice. Regarding biological class (fish *vs*. bird), no association appeared to be significant in 1x1 and 2x1. As a reminder, the hypotheses tested were based on Berlin (1994)'s experiment, in which pairs of Huambisa words were presented to non Huambisa-speaking students who had to guess which one was referring to a bird, while knowing that the second one referred to a fish. Both phonetic and conceptual contrasts may be required for subjects to produce motivated associations about fish and birds. However, the results of the paper-and-pencil experiment of our first study do not fully correspond to those obtained in this study. Although more segments were used in the first study [s, f, t, p, i, a], the main difference lies in the modality of presentation of the pseudo-words: they were provided as written stimuli, as they

were in Berlin's study. It is therefore possible that, for some reason, this modality enhances the associations regarding biological class.

The previous results led us to ponder over the cognitive mechanisms underlying the associations, such as the processing of written forms, in addition to the processing of contrasts. We therefore designed another study in which participants were presented with a lexical decision task, which corresponds to a 1x1 presentation (one shape and one pseudo-word), even more implicit than a judgment task (because the supraliminal presentation of a shape was irrelevant). The relevance of using an implicit task is in agreement with Nielsen and Rendall (2012)'s criticism that classical bouba-kiki tasks are too transparent which leads to overestimating the effects.

1.3. *Third study*

This experiment is an extension of the one conducted by Westbury (2005), which consisted in a lexical decision task in which the linguistic stimuli were displayed in frames of different shapes. These frames appeared to enhance the processing of these stimuli according to their consonantal composition (e.g. a spiky shape speeded the processing of a pseudo-word composed of voiceless plosives). In addition to the type of frame and phonetic composition, another variable was considered to investigate the influence of the shapes of letters, in order to evaluate Cuskley et al. (2015)'s explanation of the bouba-kiki effect (e.g. 'k' is spikier than 'm'). Hence, this study aimed at assessing the implicit impact of both the type of frame and the shapes of letters on the processing of pseudo-words.

Pseudo-words and words were composed of sonorants [l, m, n], voiced plosives [b, d, g] or voiceless plosives [p, t, k], and were presented in two different fonts, an angular one (*Agency FB*) or a curvy one (*Gabriola*) in two possible frames, a round one or a spiky one. Multiple potentially confounding factors were controlled regarding the strings of letters (e.g., number of phonemes, syllabic structure, orthographic neighborhood), within and between groups of words and pseudo-words. A trial began with the display of a frame in which – after 1 to 3 seconds (SOA) – appeared a string of letters, which remained visible until the participant provided a response or reached the time limit (2 seconds).

Analyzes revealed an interaction between the type of frame and the font in the processing of pseudo-words, and simple effects indicated more precisely that this was due to faster response times in one case: when pseudo-words were presented with the angular font in

a spiky frame. This result runs counter to *motivated* associations and seems in line with Cuskley et al.'s proposition about the influence of the shapes of letters. However, since it only concerned angular letters in spiky frames, and not curvy letters in round frames, we suggested another explanation: a visual saliency priming effect from spiky frames on the processing of angular letters. Indeed, perpendicular lines are more rapidly detected than concentric features (Coelho, Cloete, & Wallis, 2010). Moreover, according to the neuronal recycling hypothesis (Dehaene, 2005), the VWFA – the brain area specialized in the visual processing of words – would have been recycled for this purpose because it had a related function: the processing of geometrical features as line junctions, which is supported by the fact that primates' analogous area encodes intersections (Dehaene & Cohen, 2007). In addition, pseudo-words would require stronger activation of brain areas in comparison to words (Newman & Twieg, 2001), which could explain the difference we reported between words and pseudo-words: as for words, there was indeed rather a main effect from frames with spiky frames speeding processing. This difference between words and pseudo-words may be explained in regards to the dual-route hypothesis: words are processed holistically, while pseudo-words are processed through grapheme-phoneme mappings.

The first conclusion of this study is that experimenters working on *motivated* associations with spiky and round shapes should consider low-level visual processing, because it could influence their results. For example, we can wonder whether the difference between 'congruent' and 'incongruent' matching pairs could be explained by an enhancement of the processing of spikiness, instead of both spikiness and roundness. Hence, round and spiky shapes should be analyzed distinctly, and not together.

There are several possible explanations for the absence of *motivated* associations. First, the type of presentation – 1x1 – offers no phonetic contrast while it could be needed according to our second study. Second, the protocol may be too implicit, since frames and fonts were irrelevant to the task from the participants' point of view. However, different studies have highlighted the possibility to produce *motivated* associations in implicit protocols. For example, the study by Kovic et al. (2010) involved an implicit learning and these authors found *motivated* associations. However, some learning meta-strategies could have been devised by participants. In our case, such strategies were unlikely because of the variability of the pairings between the modalities of the three variables (e.g., spiky-angular-sonorants, round-angular-sonorants, etc.). Concomitantly, the variability of pairings may also explain the longer response times compared

to those usually obtained in lexical decision tasks: Garner interference effect could explain a decrease in overall performance.

Several studies have brought to light an enhancing effect of one modality on another in sound-shapes associations. First, Marks (1987) reported an influence of pitch on the classification of round and spiky shapes. Second, in the study conducted by Hung et al. (2017), participants were faster to determine the location of a masked visual shape when an oral pseudo-word previously presented was congruent with the shape. Third, Sidhu and Pexman (2017) obtained a supraliminal priming effect from a pseudo-word on the categorization of a shape. Overall, these three studies suggest an effect of audition on vision in shape categorization tasks. This may be the reason why we did not obtain an influence from the frame on the processing of pseudo-words. If the task had been to categorize one shape, following the presentation of a word or a pseudo-word, we could perhaps have obtained results in line with these three previous studies. Also, it is possible to juxtapose this idea with the results we obtained in the second study. We found in particular larger effect sizes and more significant results with the 1x2 protocol than with the 2x1 protocol. One may reasonably assume that 1) presenting one concept and two pseudo-words lead to cognitive processes where the concept influences the choice between phonetic forms, and similarly that 2) presenting one pseudo-word and two concepts lead to cognitive processes where the phonetic form impacts on the choice between two concepts. In that case, semantic information would influence phonetic judgments less than phonetic information influences semantic judgment.

The three studies that compose this thesis aimed at evaluating *motivated* associations implying ecological concepts (studies 1 and 2) and the differences induced by different methodological settings in the experimental study of *motivation* (studies 1, 2 and 3). While each study contains its own discussion regarding the results, as well as possible perspective, the following discussion aims at completing some elements that have already been considered in the introduction – or in regard to the experimental studies – and at examining others that have not been yet surveyed. This allows to open this work to other research fields that can contribute to the evaluation and investigation of the nature of *motivated* associations, and of the related cognitive processes.

2. Broadening the scope

The following section further discusses elements regarding 1) cross-linguistic, cross-cultural similarities and their methodological implications; 2) the origin and evolution of language through cross-linguistic and interspecies investigations; 3) the cognitive nature of *motivated* associations in connection with pathologies such as aphasia, autism spectrum disorder, or dyslexia; 4) evidence coming from studies using neuro-imagery and their broader theoretical implications: 5) embodied cognition and 6) linguistic relativity (in relation with synesthesia).

2.1. *Cross-linguistic and cross-cultural studies and their methodological implications*

2.1.1. Cross-linguistic studies

Tzeng, Nygaard and Namy (2017) showed English-speaking participants spoken words from 10 different languages (Albanian, Dutch, Gujarati, Indonesian, Korean, Mandarin, Romanian, Tamil, Turkish, and Yoruba) denoting different meanings along several dimensions: large-small; round-spiky; fast-slow; moving-still. In each trial, one spoken word of one of the 10 languages was presented, with the related pair of antonyms translated in English (e.g. 'big' and 'small'). Participants had to determine which of the two antonyms was the proper translation. For each possible meaning, participants chose the correct translation significantly more than expected at chance level (with a mean agreement across words of 0.65). However, this study was based on a previous one, conducted by DeFife, Nygaard and Namy (n.d., in Tzeng et al., 2017), in which the authors had obtained an even higher mean agreement (0.85). This difference may be explained by the fact that whereas in the initial study, the semantic dimension was always the same within one participant (e.g. always 'big' and 'small'), in the second study the dimension varied across trials. '*Listeners in this study were required to switch attention to different sets of linguistic and semantic features, as well as to different speaker characteristics from trial to trial, which rendered it more difficult for listeners to selectively attend to particular sound characteristics or semantic dimensions to inform their decisions*' (p. 2199). This confirms our insights about the studies we conducted using animal features: changing the dimensions we evaluated across trials allowed us to mask them in order to avoid some metacognitive strategies (e.g. the awareness of what is evaluated, the need for consistency, etc.) It is thus possible that stronger effect sizes, or higher number of significant associations, would appear with another protocol in which the assessed dimension would be constant across trials.

In another experiment, Tzeng et al. (2017) showed participants one foreign word with two pairs of antonyms, resulting in a forced-choice between four possibilities of signification. For example, a word which means 'pointy' was presented with the translations 'pointy', 'round', 'fast' and 'slow'. The overall performance was higher than 25% for every meaning, and significant for seven out of eight; only words for 'slow' were not significantly mapped with the proper meaning. Moreover, when the proper meaning was not chosen, participants chose significantly more 'moving' for words that meant 'pointy' and the meanings 'pointy' and 'moving' for words that meant 'fast'. These cross-modal mappings may be explained by a common (possibly amodal) dimension such as intensity. To quote the authors, '*semantic relatedness may also be a product of correlated features across referents. For example, if small things also tend to be fast, then the observed crossdimensional mappings may be a product of priming or generalization based on these associations*' (p. 2211).

This means that some associations can be explained by indirect related associations, and it is possible to try to infer some new *motivated* associations based on already established ones.

2.1.2. Cross-cultural studies

A recent study conducted by D'Anselmo, Prete, Zdybek, Tommasi and Brancucci (2019) focused on the guessability of foreign languages in two distinct populations (Italian and Polish speakers) in order to assess the possible discrepancy due to cross-cultural and cross-linguistic differences. Words (verbs, nouns and adjectives) of four unrelated languages (Finnish, Japanese, Swahili and Tamil) were orally presented to participants with three possible translations: the real meaning, its antonym and a distractor. Both Italian and Polish speakers guessed significantly higher than chance (35.31% compared to 33.33%, the chance level) the meaning of the words, with no significant difference between the two groups. Analyses per language, however, revealed significant guessability only for Finnish and Japanese, even if recognition rates for all four languages exceeded 33.33% of correct answers. This difference may be explained by the fact that Finnish and Japanese both '*seem to possess a rich ideophonic vocabulary*' (p. 6). Overall, nouns and verbs significantly exceeded chance level but not adjectives, even though there were some differences between the four languages. An interaction exists between languages and categories of words, but this will not be further developed here. The most interesting result of this study is indeed for us that there are no differences between the two populations studied, Italian and Polish speakers, which suggests a common sensitivity across speakers of different languages.

In comparison, the study by Nygaard et al. (2009, presented in section 2.6 of the General introduction) – in which English-speaking participants learned Japanese words – revealed similar performances for both actual and opposite meanings, in comparison to unrelated meaning. The results were in favor of a relation between words and their semantic fields, rather than between words and their specific meanings (even though performance was better for actual meanings than for opposite meanings, the difference was not statistically significant). Unfortunately, D'Anselmo et al. did not present results about errors and whether they were different for antonyms and distractors. If there were significantly more errors in favor of antonyms, it would represent an additional argument in favor of Nygaard et al.'s conclusions, leading to further understanding of whether the segmental composition of words better foreshadows the semantic field or the meaning itself.

While Italian and Polish speakers similarly guessed the meaning of Finnish and Japanese words, which is in favor of a potential universal sound symbolism, one may wonder whether Finnish and Japanese speakers would be more accurate at guessing the meaning of Japanese and Finnish words, respectively. Indeed, since their respective languages are more iconic or symbolic – which enables significantly more correct guesses in speakers of other languages – Finnish and Japanese speakers may be even more sensitive to iconicity or symbolism in another language. The study by Imai et al. (2008) revealed higher matching rates and more consistency in choices for congruent pairings in Japanese speakers in comparison to English speakers. But Japanese speakers' higher sensitivity may be explained by the linguistic exposure to, and the learning of regularities of, their own language, since ideophones in this study were created on the grounds of the description of Japanese mimetics. In order to disentangle the two possible explanations, it would thus be interesting to evaluate whether Japanese and Finnish speakers are better than Italian and Polish speakers at guessing the meanings of words of Finnish and Japanese, respectively.

2.2. *Language emergence and evolution*

2.2.1. Evidence of motivation through language change

As a reminder, Monaghan et al. (2011, presented in section 2.6 of the General introduction)'s study highlighted the complementarity of arbitrariness and systematicity, and assessed the necessity of systematicity in order for the advantage of arbitrariness to show up.

Another study provides insight about *motivation* and language evolution. Johansson and Carling (2015) analyzed 30 languages from the Indo-European family (13 contemporary

languages and 17 reconstructed languages) with respect to their deictic lexemes (i.e. words denoting persons or locations). Their hypotheses were based on the frequency code, which states that high frequencies are associated to smallness (and proximity) and low frequencies to largeness (and distance) (cf. Figure 3 presenting the ordering of the segments). The authors looked for *motivated*, non-*motivated* and reversed *motivated* forms. They found that the majority of forms were *motivated* (70.2%) (e.g. in Proto-germanic, a proximal form is '(h)iz' and its distal counterpart is 'sa'). They also added that '*genetic explanations, inherited phonetic forms of the deictic terms, for the high motivated support can be disregarded due to the diversity in rebuilding of forms*' (p. 26). The authors eventually concluded that '*based on the results of this study it seems very likely that iconicity is involved in the rebuilding of deictic systems and forms in Indo-European languages, both contemporary and historically, and it is highly likely that this is the case for other language families as well*' (p.27).

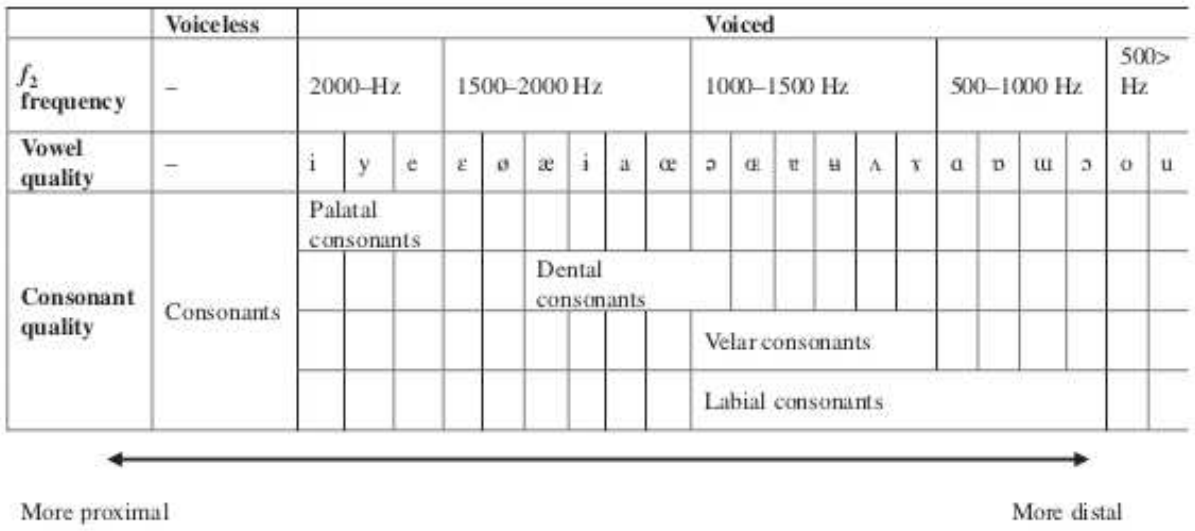


Figure 3. Ordering of segments in function of the proximal-distal continuum. Table extracted from Johansson & Carling (2015).

In addition to the assumption that *motivation* could have facilitated the emergence of language, there is thus evidence in favor of a diachronic influence of *motivation* in the evolution of languages. Following this, language evolution is not a process where initial *motivation* would only decrease or be maintained in some languages or areas of the lexicon. In addition to such processes, there is indeed the possibility for *motivated* associations to arise in languages as a regular output of language change. Such a dual perspective brings additional complexities and nuances to the whole picture of *motivation*.

2.2.2. Non-human primates

Pitch-luminance mappings, known to exist in humans, were also looked for in chimpanzees in a study carried out by Ludwig, Adachi and Matsuzawa (2011). Participants (humans and chimpanzees) had to categorize black and white stimuli while high- and low-pitched sounds were simultaneously displayed, as in the study conducted by Melara (1989). Humans and chimpanzees both had a better performance for congruent trials compared to incongruent ones. Incongruent trials resulted in longer response times in humans (with no difference as for errors), whereas chimpanzees made more errors in this condition (with no difference as for response times). This difference is explained by the authors by behavioral differences between the two species: humans try to be as accurate as possible while chimpanzees are more impulsive. Nevertheless, this study demonstrated that crossmodal correspondences are not specific to humans and are thus not explained by cultural or linguistic mediation. It represents an argument in favor of structural mediation instead of statistical learning, given the lack of natural correspondences between pitch and luminance in the natural environment. More interestingly, it is an argument in favor of an implication of *motivation* in language emergence. '*Our findings in the present study suggest that natural tendencies to systematically map certain dimensions (here, pitch–luminance) were already present in our nonlinguistic ancestors. Thus, such cross-modal mappings might indeed have influenced the emergence of language*' (p. 20663).

2.3. *Studies within impaired individuals*

Studies in diverse population may also provide insights about the relation between *motivated* signs and crossmodal correspondences.

2.3.1. Aphasic patients

In studies reported in the introductory chapter, it has been reported that *motivated* words enhance learning of foreign words in adults (e.g. Nygaard et al., 2009) and facilitate word generalization in children (e.g. Kantartzis et al., 2011). The following study brings to light their particular status in adults suffering from aphasia.

In a study conducted by Meteyard, Stoppard, Snudden, Cappa and Vigliocco (2015), English speakers with aphasia (of three types: anomic, Broca or conduction) were presented with different tasks (repetition, reading aloud, auditory lexical decision and visual lexical decision). While there were no differences between iconic and control words within control participants in the four tasks, results revealed overall better performances for iconic words

compared to controls words within aphasic participants, more precisely in two tasks: reading aloud and auditory lexical decision. Authors supposed that iconicity has a stronger influence when phonology-semantics mappings are involved, instead of only phonology (repetition) or only semantics (visual lexical decision task). Two explanations were thus proposed. First, iconic words could be characterized by additional connections between the semantic system and systems dealing with *modality-specific representations*. Such a redundancy would minimize damage in case of brain injury. Second, iconic words would be characterized by direct connections between their phonological forms and their *modality-specific representations*. There would exist therefore an extra route for them, possibly in the right hemisphere more involved in crossmodal correspondences, which would also shield them from aphasia (which often results from brain injuries in the left hemisphere). Authors concluded that '*iconicity provides an opportunity for greater embodiment in language processing*' (Meteyard et al., 2015, p. 266).

In any case, this bring to light the potential advantage that *motivated* words may represent for rehabilitation of patients with aphasia.

While aphasic individuals present an advantage for *motivated* associations, other cognitive impairments are instead associated with difficulties for this type of associations.

2.3.2. Dyslexic individuals

Drijvers, Zaadnoordijk and Dingemanse (2015) conducted a study in order to determine if dyslexic individuals perform as controls in a bouba-kiki task for the reason that they present impaired crossmodal processing and since reading depends on effective mappings between graphemes and phonemes (McNorgan, Randazzo-Wagner, & Booth, 2013). The experiment consisted in displaying two visual stimuli, a round one and a spiky one, and producing an oral pseudo-word. The control group produced significantly more *motivated* associations than dyslexic patients (73 *vs*. 60%). The authors proposed, as a possible explanation, an impairment of the processes underlying crossmodal correspondences, namely abstraction and coupling of different modalities (i.e. segmental and conceptual). At a neural level, they proposed an implication of the angular gyrus – which seems impaired in dyslexic patients (Pugh et al., 2000) – that is the brain area possibly involved in *motivated* associations according to Ramachandran & Hubbard (2001b).

2.3.3. Autism spectrum disorder

A difference in performance between individuals with autism spectrum disorders (ASDs) and controls would also point to a deficit of multisensory integration in the former.

Occelli, Esposito, Venuti, Arduino and Zampini (2013) conducted a study in which low- and high-functioning ASD patients were compared to controls in a bouba-kiki task. All of them were Italian children aged from 5 to 20 years old. Participants were presented with a 2x2 protocol and thus had to select one shape among two with one oral pseudo-word among two. Controls produced significantly more *motivated* associations (about 85%)[37] than high-functioning ASDs (about 69%), who in turn produced significantly more *motivated* associations than low-functioning ASDs (about 52%). The level of the latter group did not differ from chance level. These results demonstrate that *motivated* pairings are '*affected by the presence of ASD, and even more by the comorbidity between retardation and ASD*' (p. 237).

In addition to confirming our intuition that any linguistic impairment, like dyslexia, could have influenced our results, this is a neuropsychological evidence that *motivated* associations are part of a larger family, namely crossmodal correspondences. However, all these studies do not rule out the possibility that language impairment itself influences the results, instead of a crossmodal deficiency. Indeed, low verbal IQ in low-functioning ASDs and grapheme-phoneme correspondences may explain the lower amount of *motivated* associations, instead of a general crossmodal impairment. To better assess this issue, it would be needed to evaluate different crossmodal correspondences in these populations, as between pitch and elevation, in order to evaluate whether the impaired crossmodal correspondences are exclusively the ones implying language, or whether they are more general. However, in the case of ASDs '*the present findings seem to point to poorer capabilities of patients with ASD to integrate information across different sensory modalities, consistently with previous behavioral and neuroimaging studies*' (Occelli et al., 2013, p. 238).

2.4. *Brain imagery's evidence for multimodal integration*

One way to better assess to which extent *motivated* associations are crossmodal correspondences is brain imagery.

In the study conducted by Kovic et al. (2010) previously outlined (in section 3.4.2 of the General introduction), the authors replicated their protocol – an implicit learning task of congruent (match) or incongruent (mismatch) pairs of bouba-kiki shapes and pseudo-words – using EEG, which measures event-related potentials (ERP)[38]. The authors found a strong

[37] Percentages reported here have been read on a graphic of the publication.
[38] ERP are electrical responses that are measured and averaged per electrode. The wave form is composed of a series of positive and negative peaks which correspond to the polarity of the responses. Hence P or N appoint to the polarity of the wave and the following number refers to the time after exposure.

negative response in occipital areas, around 160 ms after exposure, in the case of a congruent trial, whatever the learning (matching or mismatching). For them, this conveys an intermodal integration. They also found a N400 response in the case of a mismatch, depending thus on the learning and not on congruency. N400 are known to appear for unexpected stimuli (e.g. the last word of the following sentence would elicit one: *the bird flies in the* ***table***).

Asano et al. (2015) also conducted an EEG study with 11-month-old Japanese children to whom they showed a visual round or spiky shape followed by a pseudo-word that was either congruent or incongruent (as always, according to previous studies). They found evidence in favor of multisensorial integration, more precisely higher-amplitude gamma frequencies in centro-parietal regions (1-300 ms), greater synchronization between brain areas in incongruent condition and a similar response as N400 in the incongruent condition (350-550 ms).

Lockwood and Tuomainen (2015) conducted another study with Japanese speakers using EEG and real words instead of pseudo-words, more precisely mimetics compared to arbitrary adverbs. P2 responses were greater for mimetics compared to adverbs, and the authors argued that it reflected the multisensory integration of the two modalities.

Overall, these studies suggest a particular response for congruent trials, possibly indicating multimodal integration, and another response for incongruent trials, echoing incongruity detection (according to learning in Kovic et al., 2010, and to congruence in Asano et al., 2015).

Kanero, Imai, Okuda, Okada and Matsuda (2014) conducted an fMRI study with Japanese speakers, using orally presented mimetics, (non-mimetic) verbs and (non-mimetic) adverbs. They found specific activations for mimetics – compared to verbs and adverbs – in one location in particular, the right superior temporal sulcus. However, since stimuli involved motions and one of the functions of this brain area is motion processing, the specificity of the activation for *motivated* relations needs to be further assessed. In a second experiment, the authors compared mimetics referring to movements to others referring to static shapes. They found stronger activations in the same area for mimetics referring to both movements and static shapes for congruent trials, compared to incongruent ones. This means that this area may be '*a critical hub for processing Japanese mimetic words, and possibly sound symbolism in general*' (p. 7).

Besides the previous perspectives, there is another position according to which *motivated* associations reflect *embodied cognition* (though both may coexist), as it may be highlighted by the following fMRI studies.

2.5. *Embodied cognition*

Ramachandran and Hubbard (2001b) proposed, as an explanation of some *motivated* associations, that some words or segments are synkinetic mimicries of what they denote. For example, [i] in words denoting smallness are produced by a small aperture of the mouth, mimicking a narrow distance between two fingers. Another example is the concept of 'you' in several languages (e.g., 'vous' and 'tu' in French, 'thoo' in Tamil), which pronunciation is accompanied by an outward movement of the lips induced by the vowels [u] and [y].

This is more generally in line with the embodied theory, according to which cognition is influenced by the entire body which involves sensory-motor representations. As exemplified by Lupyan and Bergen (2015), the embodied cognition theory states that '*comprehending a word like "eagle" activates visual circuits that capture the implied shape, canonical location, and other visual properties of the object, as well as auditory information about its canonical sound*' (p. 7). They later added that '*not only are perceptual, motor, and affective systems activated during meaning construction, but that this activity plays a functional role in comprehension*' (p. 7).

Experimental evidence can support this theory with respect to *motivated* associations. First, Osaka, Osaka, Morishita, Kondo and Fukuyama (2004) conducted an fMRI study using mimetics expressing pain that were compared to pseudo-words. They found that brain areas involved in the sensation of pain were more activated with mimetics, more precisely the anterior cingular cortex, the prefrontal cortex, the insula and somatosensory areas. The coactivation of the prefrontal cortex and the anterior cingular cortex suggests a functional connectivity. The authors suggested, more precisely, that the activation of the prefrontal cortex expresses the semantic retrieval of information about pain from long term memory systems via attention, producing an imaginary pain. However, it is surprising that this study compares mimetics expressing pain to pseudo-words expressing neither semantic nor sensorial information. Lockwood and Dingemanse (2015)'s review presented several studies conducted by the same authors using fMRI (see Table 8). These studies showed that ideophones activate specific brain areas depending on their semantic field. However, one should note that all these studies, as Lockwood and Dingemanse (2015) added, also compared ideophones with pseudo-words.

Hence, these results must be considered cautiously. A comparison with verbs and adverbs expressing pain would be for instance more informative and reliable.

Table 8. Results of various studies using fMRI, reported in Lockwood and Dingemanse (2015)'s review.

Ideophonic expressions	Specific brain areas	Common brain area
Laughter	Visual cortex, extrastriate cortex, premotor cortex, striatal reward area	Visual cortex and premotor cortex
Pain	Cingulate cortex (the pain related area)	
Crying	Laughter areas + inferior frontal gyrus and anterior cingulate cortex	
Gaze direction	Frontal eye field	
Manner of walking	Extrastriate visual cortex	

At the behavioral level, the study of Šetić and Domijan (2007) is of interest. These authors proposed a semantic judgment task where participants had to categorize words according to what they denote, either a flying animal or a non-flying animal. Words were presented either at the top or at the bottom of the screen. Results revealed an interaction between the meaning of the word and the spatial position: words denoting flying animals were processed faster at the top position compared to the bottom position, and similarly, non-flying animal words were processed faster at the bottom position, compared to the top position. Since this interaction could indicate an influence of the type of answer – 'flying' being related to a top position – the authors conducted a second study in which the categorization was not about spatial position. Participants had to categorize words in two categories: living and non-living entities. Hence, in addition to the previous words for animals, the authors added words for inanimate objects also related to top or bottom spatial position (e.g. 'moon' and 'floor', respectively). Since some participants were engaged in the first experiment, analyses were restricted to words denoting non-living entities. Likewise, there was an interaction between the spatial position on the screen and the spatial position commonly encountered for the objects. Response times were faster when the spatial position on the screen was congruent with the spatial position of the objects, compared to the opposite displayed position. The results of experiment 2 thus replicated those of experiment 1 with answers unrelated to spatial position (i.e. living and non-living). As the authors underlined, '*it should be noted, however, that neither*

study permits us to distinguish whether results should be attributed to the interference due to the conflicting information or to facilitation due to the consistent information' (p. 308).

These results can be explained by the theory of perceptual simulation proposed by Barsalou (1999), whereby lexical processing activates perceptual representations. This is in opposition with amodal theories according to which higher-level cognitive representations are non-perceptual. Rather, Barsalou supports the hypothesis that during experience, associative brain areas record pattern of sensory-motor activations (as well as proprioception and introspection) that are later reactivated via a simulator, i.e. a 'frame' – '*an integrated system of perceptual symbols that is used to construct specific simulations of a category*' (p. 590) and the simulations it produces, even in the absence of the perceptual input. For example, the first perception of a car produces a frame composed of the overall form and its components; the perception of another car will update this frame, adding details and new elements; and so on indefinitely.

Gibbs (2003) proposed the *embodiment premise* that is the '*embodied understanding of language*' (p. 12). More precisely, embodied information is involved in language processing and evidence in that respect is presented in Gibbs' review. For example, the processing of metaphors would be built on embodied knowledge. As claimed by Gibbs, '*processing linguistic meaning is not a matter of understanding what words mean, but includes the perception of physical objects, physical events, the body, and other people in interaction*' (p. 13). For example, judging the semantic correctness of the expression 'aim a dart' is speeded by producing the handshape for 'pinch' (Klatzky, Pellegrino, McCloskey, & Doherty, 1989).

2.6. *Synesthesia and linguistic relativity*

Moos et al. (2014)'s study was rapidly mentioned in section 3.3 of the General introduction. They brought to light common associations in synesthetes and non-synesthetes (although stronger ones for the former), between colors and acoustic features of vowels. The following study further assesses this matter.

In addition to acoustic variations of vowels (with variable F1 and F2), Cuskley, Dingemanse, Kirby and Leeuwen (2019) aimed at evaluating the role of specific vowels instead of formant variation within vowels. They conducted an online study with a large sample of Dutch speakers (over a thousand participants) using a more fine-grained color-space than the selection of 16 colors used in Moos et al.'s experiment.

Participants were presented with 16 vocalic sounds, three times each, and had to pick a color in a color space for each sound (number of trials: 48). The repetition permitted to evaluate the consistency within participants.

Moos et al.'s results about vocalic variations of F1 and F2 were replicated but vowel categories predicted even better the choice of color (the associations are summarized in Table 9). Lightness better predicted the choice than the two other axes (red-green and blue-yellow). There were differences between synesthetes and non-synesthetes, with stronger and more extremes associations for the former (e.g. even lighter colors for front vowels and even darker colors for lower vowels).

Table 9. Associated hues in accordance with specific vowels in Cuskley et al. (2019)'s study.

Vowels	Associated hues
[e, ɪ, ɛ, ø]	Light, yellow, green
[i]	Even lighter and yellower
[u, ɔ]	Bluer and darker
[u, ɔ, a]	Redder
[ɑ]	The reddest

Some participants may seem to be synesthetes because of their consistency, but this consistency can exist across vowels (e.g. systematically choosing blue hues whatever the vowel). Hence, the authors calculated the correlation between the two spaces (the vocalic space and the color space) in order to evaluate the mapping structure, in addition to consistency. Five different profiles of participants appeared depending on their consistency and mapping structure. For example, a low level of structure and a low consistency characterized participants who chose different colors for a given vowel (i.e. lack of consistency) and similar colors for distant vowels (i.e. lack of structure). At the other extremity, a participant who presented a high level of structure with a high-consistency meant that they chose distant colors for distant vowels consistently across trials. Overall, participants who showed consistency tended to also exhibit structured mappings (whether they were synesthetes or non-synesthetes). This approach permits to reveal that consistency is not necessarily an evidence for being synesthete, and that the overall structure of the mappings should be considered. Hence, this study also provided a better way of identifying synesthetes within participants.

The fact that vocalic categories are more influential than acoustic differences within categories is in favor of an implication of learning and activation of concepts. However, in the red-green dimension, acoustic differences explained associations in synesthetes while vocalic categories did not. It is hence possible that some synesthetes are sensible to acoustic variations (in addition to categories), which would explain this difference between synesthetes and non-synesthetes. These two elements may seem contradictory (or complementary) regarding the debate between synesthesia and ideasthesia.

These results are consistent with previous ones collected with English (Moos et al. 2014) and Korean (Kim, Nam, & Kim, 2018) speakers. As proposed by Cuskley et al., '*this opens up the possibility of a degree of linguistic relativity in cross-modal associations and synesthetic experience*' (p. 12). Indeed, while linguistic relativity refers to the different representations induced by the spoken language, it is also a way to investigate the possible existence of invariants in representations of the world across individuals.

In this line of thinking, the work conducted by Berlin and Kay (1999) is insightful. They analyzed basic color terms of 98 languages and found 22 combinations out of 2048 possibilities of 11 basic color terms[39]. Languages had at least two basic terms (when there are only two, they are white and black, or more precisely light and dark). If one language had three terms, the third one was red. If it had four terms, the fourth was either green or yellow. If it had five terms, it was the five previously mentioned colors. If it had six terms, the sixth was blue. The seventh term was brown. From there, every combination of the remaining four terms were possible (pink, purple, orange and grey). While the authors concluded that '*the eleven basic color categories are pan-human perceptual universals*', they also added '*but we can offer no physical or psychological explanation for the apparently greater perceptual salience of these particular eleven color stimuli nor can we explain in any satisfying way the relative ordering among them*' (p. 109).

To return to Cuskley et al. (2019)'s findings, it is possible to draw a parallel with Berlin and Kay's observations. The axis that predicted best the associations was the light-dark one which corresponds to the two first basic color terms. Then, the hues that are reported to be specifically associated to some vowels are red, green, yellow and blue (without ordering), which

[39] The four major criteria for a word to be considered as a basic color term are the following: 1) to be monolexemic (the meanings of its parts do not predict its meaning e.g. not as 'bluish'); 2) its meaning is not included in another basic term (e.g. not as 'scarlet'); 3) its use is not restricted to some objects (e.g. not as 'blond'); 4) its meaning needs to be salient, i.e. stable and established across speakers, (e.g. not as 'the color of my car').

correspond to the following four basic color terms. No other associated color was reported in Cuskley et al.'s study. However, it is possible that the description of the associated colors was 'biased' by the authors who labeled, according to their own categorization, the average coordinates in the color space of the participants' choices. For this reason, it would be interesting to further this issue, i.e. to assess whether 'primary' color terms have a special status which leads individuals to associate more segments compared to other colors cross-culturally, cross-linguistically and whatever their synesthetic profiles. The study conducted by Moos et al. used the 11 basic color terms and added five others but, unfortunately, the authors did not report the results per color. Moreover, it was restricted to vowels, but consonants would also be of interest.

3. Conclusion

This thesis consists in several contributions to the methodological approach to *motivated* associations. First, animal features or classes (bird and fish) can elicit associations, thus permitting to evaluate emotional aspects, among others. This type of investigation allows to indirectly assess theories about language evolution and emergence. A second experiment highlighted the differences induced by different types of presentation of the linguistic and non-linguistic stimuli, and thus the differences provoked by the presence or the absence of phonetic and conceptual contrasts, which should have implication for future research. Third, the final study brought to light potential perceptual biases such as visual saliency and priming effects, which may also be taken into account in future studies. Furthermore, it would be relevant to better assess the relation between *motivated* associations and 1) more general crossmodal correspondences, 2) synesthesia and ideasthesia, as well as 3) the theories of linguistic relativity and embodied cognition. These theoretical frames are not mutually exclusive to each other. Rather, they represent complementary approaches to the study of *motivated* associations. *Motivation* potentially represents a key-driver underlying language emergence and evolution, through the exploitation of a cognitive phenomenon consisting in unifying experiences in their multimodality. Also, the study of impaired individuals and brain imagery can be complementary to behavioral data (psycholinguistics) and lexicon studies.

Even though arbitrariness is an undoubtedly fundamental feature of language, there is, as a conclusion, strong evidence for *motivation*, which highlights a specific facet of the cognitive functioning of our species and of its evolution.

References

Adelman, J. S., Estes, Z., & Cossu, M. (2018). Emotional sound symbolism: Languages rapidly signal valence via phonemes. *Cognition*, *175*, 122–130.

Ahlner, F., & Zlatev, J. (2010). Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies*, *38*(1), 298–348. https://doi.org/10.12697/sss.2010.38.1-4.11

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, *63*, 196–205. https://doi.org/10.1016/j.cortex.2014.08.025

Asano, M., & Yokosawa, K. (2011). Synesthetic colors are elicited by sound quality in Japanese synesthetes. *Consciousness and Cognition*, *20*(4), 1816–1823. https://doi.org/10.1016/j.concog.2011.05.012

Aveyard, M. E. (2012). Some consonants sound curvy: effects of sound symbolism on object recognition. *Memory & Cognition*, *40*(1), 83–92. https://doi.org/10.3758/s13421-011-0139-3

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(04), 577–660. https://doi.org/10.1017/S0140525X99002149

Bergen, B. K. (2004). The Psychological Reality of Phonaesthemes. *Language*, *80*(2), 290–311. https://doi.org/10.1353/lan.2004.0056

Berlin, B, & Kay, P. (1999). *Basic color terms: their universality and evolution.* CSLI.

Berlin, Brent. (1994). Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In Leanne Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 76–93). New York: Cambridge University Press.

Blasi, D. E., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound – meaning association biases evidenced across thousands of languages. https://doi.org/10.1073/pnas.1605782113

Bolinger, D. L. (1950). Rime, Assonance, and Morpheme Analysis. *WORD*, *6*(2), 117–136. https://doi.org/10.1080/00437956.1950.11659374

Bottini, R., Barilari, M., & Collignon, O. (2019). Sound symbolism in sighted and blind. The role of vision and orthography in sound-shape correspondences. *Cognition*, *185*(February), 62–70. https://doi.org/10.1016/j.cognition.2019.01.006

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape-sound matches, but different shape-taste matches to Westerners. *Cognition*, *126*(2), 165–172. https://doi.org/10.1016/j.cognition.2012.09.007

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

Chen, Y. C., Huang, P. C., Woods, A., & Spence, C. (2016). When "bouba" equals "kiki": Cultural commonalities and cultural differences in sound-shape correspondences. *Scientific Reports*, *6*(December 2015), 1–9. https://doi.org/10.1038/srep26681

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition*, *102*, 341–360. https://doi.org/10.1016/j.cognition.2006.01.001

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding Emotions in Speech with the Size Code. A Perceptual Investigation. *Phonetica*, *65*(4), 210–230.

Coelho, C. M., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect: When angry faces are just cross(es). *Journal of Vision*, *10*(1), 1–14.

Corballis, M. C. (2002). *From hand to mouth: the origins of language*. Princeton: Princeton University Press. https://doi.org/10.1017/s0022226702221982

Cuskley, C., Dingemanse, M., Kirby, S., & Leeuwen, T. (2019). Cross-modal associations and synesthesia : Categorical perception and structure in vowel – color mappings in a large online sample. *Behavior Research Methods*, 1–25. https://doi.org/https://doi.org/10.3758/s13428-019-01203-7

Cuskley, C., Simmer, J., & Kirby, S. (2015). Phonological and orthographic influences in the bouba-kiki effect. *Psychological Research*. https://doi.org/DOI 10.1007/s00426-015-0709-2

D'Anselmo, A., Prete, G., Zdybek, P., Tommasi, L., & Brancucci, A. (2019). Guessing Meaning From Word Sounds of Unfamiliar Languages: A Cross-Cultural Sound Symbolism Study. *Frontiers in Psychology*, *10*(March), 1–11. https://doi.org/10.3389/fpsyg.2019.00593

Davis, R. (1961). The fitness of names to drawings. A cross-cultural study in Tanganyika. *British Journal of Psychology*, *52*(3), 259–268. https://doi.org/10.1111/j.2044-8295.1961.tb00788.x

DeFife, L. C., Nygaard, L. C., & Namy, L. L. (n.d.). Cross-linguistic consistency and within-language variability of sound symbolism in nature languages. *Manuscript in Preparation*.

Dehaene, S. (2005). Evolution of human cortical circuits for reading and arithmetic: The 'neuronal recycling' hypothesis. In S. Dehaene, J.-R. Duhamel, M. Hauser, & G. Rizzolatti (Eds.), *From Monkey Brain to Human Brain*. Cambridge: MIT Press.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., & Merikle, P. M. (2003). The role of meaning in grapheme-colour synaesthesia. *Cortex*, *42*, 243–252.

Drijvers, L., Zaadnoordijk, L., & Dingemanse, M. (2015). Sound-Symbolism is Disrupted in Dyslexia: Implications for the Role of Cross-Modal Abstraction Processes. In *CogSci* (pp. 602–607).

Edquist, J., Rich, A. N., Brinkman, C., & Mattingley, J. B. (2006). Do synaesthetic colours act as unique features in visual search? *Cortex*, *42*(2), 222–231. https://doi.org/10.1016/S0010-9452(08)70347-2

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 1–14. https://doi.org/10.1167/7.5.7.Introduction

Everaert-Desmedt, N. (2011). La sémiotique de Peirce. In Louis Hébert (Ed.), *Signo*. Retrieved from http://www.signosemio.com/peirce/semiotique.asp

Fitch, T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fónagy, I. (1961). Communication in poetry. *Word*, *17*, 194–218.

Fónagy, I. (1983). *La vive voix. Essais de psycho-phonétique*. Paris: Payot.

Fort, M., Martin, A., & Peperkamp, S. (2015). Consonants are More Important than Vowels in the Bouba-kiki Effect. *Language and Speech*, *58*(2), 247–266. https://doi.org/10.1177/0023830914534951

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191–1203.

Garner, W. R. (1974). Attention: the processing of multiple sources of information. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception Vol. 2* (pp. 23–59). New York: Academic Press. https://doi.org/10.1016/b978-0-12-161902-2.50009-2

Gasser, M., Sethuraman, N., & Hockema, S. (2011). Iconicity in Expressives : An Empirical Investigation. *Empirical and Experimental Methods in Cognitive/Functional Research*, (July), 163–180.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language*, *84*(1), 1–15. Retrieved from http://www.sciencedirect.com/science/article/pii/S0093934X02005175

Gomi, T. (1989). *An illustrated dictionary of Japanese onomatopoeic expressions*. The Japan Times.

Gould, S. J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(20), 10750–10755. https://doi.org/10.1073/pnas.94.20.10750

Hamano, S. (1998). *The sound-symbolic system of Japanese*. Stanford, CA & Tokyo: CSLI & Kuroshio Publisher.

Haynie, H., Bowern, C., & LaPalombara, H. (2014). Sound Symbolism in the Languages of Australia. *PLoS ONE*, *9*(4), e92852. https://doi.org/10.1371/journal.pone.0092852

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–135. https://doi.org/10.1017/S0140525X0999152X

Hinton, L., Nichols, J., & Ohala, J. (1994). *Sound Symbolism*. Cambridge: Cambridge University Press.

Hirata, S., Ukita, J., & Kita, S. (2011). Implicit Phonetic Symbolism in Voicing of Consonants and Visual Lightness Using Garner's Speeded Classification Task. *Perceptual and Motor Skills*, *113*(3), 929–940. https://doi.org/10.2466/15.21.28.pms.113.6.929-940

Hung, S. M., Styles, S. J., & Hsieh, P. J. (2017). Can a Word Sound Like a Shape Before You Have Seen It? Sound-Shape Mapping Prior to Conscious Awareness. *Psychological Science*, *28*(3), 263–275. https://doi.org/10.1177/0956797616677313

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*(1), 54–65. https://doi.org/10.1016/j.cognition.2008.07.015

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS ONE*, *10*(2). https://doi.org/10.1371/journal.pone.0116494

Iwasaki, N., Vinson, D. P., & Vigliocco, G. (2007). What do English Speakers Know about gera-gera and yota-yota?: A Cross-linguistic Investigation of Mimetic Words for Laughing and Walking. *Japanese-Language Education around the Globe*, *17*(6), 53–78.

Johansson, N., & Carling, G. (2015). The de-iconization and rebuilding of iconicity in spatial deixis: An Indo-European case study. *Acta Linguistica Hafniensia*, *47*(1), 4–32. https://doi.org/10.1080/03740463.2015.1006830

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS ONE*, *9*(5), 1–8. https://doi.org/10.1371/journal.pone.0097905

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cognitive Science*, *35*(3), 575–586. https://doi.org/10.1111/j.1551-6709.2010.01169.x

Keller, R. (1998). *A theory of linguistic signs*. Oxford University Press UK.

Kim, H. W., Nam, H., & Kim, C. Y. (2018). [i] is Lighter and More Greenish Than [o]: Intrinsic Association Between Vowel Sounds and Colors. *Multisensory Research*, *31*, 419–437.

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., & Doherty, S. (1989). Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. *Journal of Memory and Language*, *28*(1), 56–77. https://doi.org/10.1016/0749-596X(89)90028-4

Knoeferle, K., Li, J., Maggioni, E., & Spence, C. (2017). What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-05965-y

Köhler, W. (1947). *Gestalt Psychology (2nd Edition)*. New York City: Liveright.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*(1), 19–28. https://doi.org/10.1016/j.cognition.2009.08.016

Kunihira, S. (1971). Effects of the Expressive Voice on Phonetic Symbolism. *Journal of Verbal Learning and Verbal Behavior*, *10*, 427–429. https://doi.org/10.1016/S0022-5371(71)80042-7

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*, *6*(August), 1–14. https://doi.org/10.3389/fpsyg.2015.01624

Lockwood, G., & Tuomainen, J. (2015). Ideophones in Japanese modulate the P2 and late positive complex responses. *Frontiers in Psychology*, *6*(July), 1–10. https://doi.org/10.3389/fpsyg.2015.00933

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661–20665. https://doi.org/10.1073/pnas.1112605108

Lupyan, G., & Bergen, B. (2015). How Language Programs the Mind. *Topics in Cognitive Science*, 1–17. https://doi.org/10.1111/tops.12155

Marks, L. E. (1987). On Cross-Modal Similarity: Auditory-Visual Interactions in Speeded Discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384–394. https://doi.org/10.1037/0096-1523.13.3.384

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The Handbook of Multisensory Processes.* (pp. 85–105). MIT Press. https://doi.org/10.1063/1.3246835

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology*, *50*, 125–145. https://doi.org/10.1016/S0167-8760

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception*. https://doi.org/10.1068/p2866

Martino, G., & Marks, L. E. (2001). Synesthesia : Strong and Weak. *Current Directions in Psychological Science*, *10*(2), 61–65.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Developmental Science*, *9*(3), 316–322. https://doi.org/10.1111/j.1467-7687.2006.00495.x

McNorgan, C., Randazzo-Wagner, M., & Booth, J. R. (2013). Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Frontiers in Human Neuroscience*, *7*, 1–12.

Mehler, J., Jusczyk, P., Lamsertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143–178.

Melara, R. D. (1989). Dimensional Interaction Between Color and Pitch. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(I), 69–79. https://doi.org/10.1037//0096-1523.15.1.69

Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: correspondences between linguistic and nonlinguistic dimentsions. *Memory & Cognition*, *18*(5), 477–495.

Melara, R. D., & O'Brien, T. P. (1987). Interaction Between Synesthetically Corresponding Dimensions. *Journal of Experimental Psychology: General*, *116*(4), 323–336. https://doi.org/10.1037/0096-3445.116.4.323

Meteyard, L., Stoppard, E., Snudden, D., Cappa, S. F., & Vigliocco, G. (2015). When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia*, *76*, 264–275. https://doi.org/10.1016/j.neuropsychologia.2015.01.042

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*(3), 325–347. https://doi.org/10.1037/a0022924

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(5), 1152–1164. https://doi.org/10.1037/a0027747

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500. https://doi.org/10.1016/0163-6383(93)80007-U

Moos, A., Smith, R., Miller, S. R., & Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder. *I-Perception*, *5*(2), 132–142. https://doi.org/10.1068/i0626

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

Morton, E. S. (1977). On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds. *The American Naturalist*, *111*, 855–869. https://doi.org/10.1086/283219

Mroczko-Wasowicz, A., & Nikolić, D. (2014). Semantic mechanisms may be responsible for developing synesthesia. *Frontiers in Human Neuroscience*, *8*, 1–13. https://doi.org/10.3389/fnhum.2014.00509

Mroczko, A., Metzinger, T., Singer, W., & Nikolic, D. (2009). Immediate transfer of synesthesia to a novel inducer. *Journal of Vision*, *9*(12), 1–8. https://doi.org/10.1167/9.12.25

Newman, S. D., & Twieg, D. (2001). Differences in Auditory Processing of Words and Pseudowords: An fMRI Study. *Human Brain Mapping*, *14*(1), 39–47. https://doi.org/10.1002/hbm.

Nielsen, A., & Rendall, D. (2011). The sound of round: evaluating the sound-symbolic role of consonants in the classic Takete-Maluma phenomenon. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Expérimentale*, *65*(2), 115–124. https://doi.org/10.1037/a0022268

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*(2012), 115–125. https://doi.org/10.1515/langcog-2012-0007

Nikolić, D. (2009). Is Synaesthesia Actually Ideaestesia ? An Inquiry Into the Nature of the Phenomenon. In *Proceedings of the Third International Congress on Synaesthesia, Science & Art*. Granada, Spain.

Nobile, L. (2015). Phonemes as images: An experimental inquiry into shape-sound symbolism applied to the distinctive features of French. In *Iconicity: East Meets West* (pp. 71–91). https://doi.org/10.1075/ill.14.04nob

Nuckolls, J. B. (1999). The Case for Sound Symbolism. *Annual Review of Anthropology*, *28*(1999), 225–252.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, *112*(1), 181–186. https://doi.org/10.1016/j.cognition.2009.04.001

Occelli, V., Esposito, G., Venuti, P., Arduino, G. M., & Zampini, M. (2013). The takete–maluma phenomenon in autism spectrum disorders. *Perception*, *42*, 233–241.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. Macneilage (Ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*, 1–16.

Ohala, J. J. (1997). Sound Symbolism. *4th Seoul International Conference on Linguistics [SICOL]*, (Ching 1982), 98–103. https://doi.org/10.1525/jlin.1996.6.1.109

Osaka, N., Osaka, M., Morishita, M., Kondo, H., & Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. *Behavioral Brain Research*, *153*, 123–127.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound-shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, *114*(2), 173–186. https://doi.org/10.1016/j.jecp.2012.05.004

Parise, C. V., & Spence, C. (2009). "When birds of a feather flock together": Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS ONE*, *4*(5), 1–7. https://doi.org/10.1371/journal.pone.0005664

Peirce, C. S. (1931). *Elements of Logic. The Collected Papers of Charles Sanders Peirce, vol. 2*. (C. Hartstone & P. Weiss, Eds.). Cambridge, Mass.: Belknap Press of Harvard University Press. https://doi.org/10.1038/135131a0

Peña, M., Mehler, J., & Nespor, M. (2011). The Role of Audiovisual Processing in Early Conceptual Development. *Psychological Science*, *22*(11), 1419–1421. https://doi.org/10.1177/0956797611421791

Perfors, A. (2004). What's in a Name? The effect of sound symbolism on perception of facial attractiveness. *Proceedings of CogSci*, *26*(26), 1617.

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Hillsdale, N.J: Erlbaum. https://doi.org/10.1038/ncb3241

Pugh, K. R., Mencl, W. E., Jenner, A. R., Katz, L., Frost, S. J., Ren Lee, J., … Shaywitz, B. A. (2000). Functional neuroimaging studies of reading and reading disability (developmental dyslexia). *Mental Retardation and Developmental Disabilities Research Reviews*, *6*, 207–213. https://doi.org/10.1146/annurev.psych.54.101601.145128

Ramachandran, V. S., & Hubbard, E. M. (2001a). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings. Biological Sciences / The Royal Society*, *268*(1470), 979–983. https://doi.org/10.1098/rspb.2001.1576

Ramachandran, V. S., & Hubbard, E. M. (2001b). Synaesthesia — A Window Into Perception, Thought and Language. *Journal of Consciousness Studies*, *8*(12), 3–34.

Reilly, J., Hung, J., & Westbury, C. (2017). Non-Arbitrariness in Mapping Word Form to Meaning: Cross-Linguistic Formal Markers of Word Concreteness. *Cognitive Science*, *41*, 1071–1089. https://doi.org/10.1111/cogs.12361

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

Sakamoto, M., & Watanabe, J. (2018). Bouba/Kiki in touch: Associations between tactile perceptual qualities and Japanese phonemes. *Frontiers in Psychology*, *9*, 1–12. https://doi.org/10.3389/fpsyg.2018.00295

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

Saussure, F. de. (1916). *Cours de linguistique générale*. (C. Bally & A. Sechehaye, Eds.). Paris: Payot.

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in Psychology*, *5*, 80. https://doi.org/10.3389/fpsyg.2014.00080

Šetić, M., & Domijan, D. (2007). The influence of vertical spatial orientation on property verification. *Language and Cognitive Processes*, *22*(2), 297–312. https://doi.org/10.1080/01690960600732430

Sidhu, D. M., & Pexman, P. M. (2017). A Prime Example of the Maluma/Takete Effect? Testing for Sound Symbolic Priming. *Cognitive Science*, *41*(7), 1958–1987. https://doi.org/10.1111/cogs.12438

Smith, L. B., & Sera, M. D. (1992). Analysis of the Polar Structure of Dimensions. *Cognitive Psychology*, *24*, 99–142.

Spector, F., & Maurer, D. (2013). Early sound symbolism for vowel sounds. *I-Perception*, *4*(4), 239–241. https://doi.org/10.1068/i0535

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Styles, S. J., & Gawne, L. (2017). When Does Maluma/Takete Fail? Two Key Failures and a Meta-Analysis Suggest That Phonology and Phonotactics Matter. *I-Perception*, *8*(4), 1–17. https://doi.org/10.1177/2041669517724807

Tallerman, M. (2011). Protolanguage. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford Handbook of Language Evolution*. Oxford: Oxford University Press.

Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Language and Speech*, *14*(3), 266–276.

Tarte, R. D. (1974). Phonetic symbolism in adult native speakers of czech. *Language and Speech*, *17*(1), 87–94. https://doi.org/10.1177/002383097401700109

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology*, *64*(12), 2392–2404. https://doi.org/10.1080/17470218.2011.605898

Turoman, N., & Styles, S. J. (2017). Glyph guessing for ‘oo’ and ‘ee’: Spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society Open Science*, *4*(9). https://doi.org/10.1098/rsos.170882

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). The Specificity of Sound Symbolic Correspondences in Spoken Language. *Cognitive Science*, *41*(8), 2191–2220. https://doi.org/10.1111/cogs.12474

Vainio, L., Tiainen, M., Tiippana, K., Rantala, A., & Vainio, M. (2017). Sharp and round shapes of seen objects have distinct influences on vowel and consonant articulation. *Psychological Research*, *81*(4), 827–839. https://doi.org/10.1007/s00426-016-0778-x

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language*, *93*(1), 10–19. https://doi.org/10.1016/j.bandl.2004.07.006