Iconicity in Language and Speech

Dissertation
zur Erlangung des akademischen Grades

Doktorin der Philosophie (Dr. phil.)

eingereicht an der Sprach- und literaturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Aleksandra Ćwiek, M.A.

Prof. Dr. Peter Frensch
Präsident (komm.)
der Humboldt-Universität zu Berlin

Prof. Dr. Stefan Kipf
Dekan der Sprach- und
literaturwissenschaftlichen Fakultät

Gutachterinnen und Gutachter:

1. Prof. Dr. Manfred Krifka
2. Dr. Susanne Fuchs
3. Dr. Marcus Perlman

Disputation: 11. Februar 2022

Zusammenfassung

Iconicity in Language and Speech

von Aleksandra Ćwiek

Diese Arbeit befasst sich mit der Ikonizität in Sprache. Ikonizität bezeichnet eine Ähnlichkeit zwischen der sprachlichen Form und der Bedeutung (vgl. Perniss und Vigliocco, 2014). So wie eine Skulptur einem Objekt oder einem Modell ähnelt, kann auch der Klang oder die Form von Wörtern dem Ding ähneln, auf das sie verweisen. In dieser Arbeit untersuche ich das Potenzial des ikonischen Ausdrucks auf verschiedenen sprachlichen Ebenen – der Klangebene, der Ebene der Prosodie und der Wortebene.

Im Kapitel 1 erörtere ich die Idee der Ikonizität, der Arbitrarität gegenübergestellt, und skizziere ihre Rolle in der Kommunikation. Hier beginne ich damit, das ikonische Potenzial von Vokalisierungen zu demonstrieren.

Das Kapitel 2 ist eine Einleitung zu den experimentellen Arbeiten, die zu Zwecken dieser Doktorarbeit durchgeführt wurden. In diesem Kapitel konzentriere ich mich auf die Lautsymbolik. Ich berichte zunächst von der aktuellen Literatur und stelle anschließend die Untersuchung zur Lautsymbolik deutscher Pokémon-Namen vor. Die gewonnenen Ergebnisse stellen eine wichtige Ergänzung zu der wachsenden Disziplin der Pokémonastik dar (Kawahara u. a., 2018; Shih u. a., 2019). Auf der Grundlage der Ergebnisse schlage ich jedoch eine neue Hypothese zur lautsymbolischen Funktionalität von Lauten vor, die besagt, dass sich der lautsymbolische Ausdruck in verschiedenen Sprachen unterscheidet.

Darauf folgt im Kapitel 3 die Forschung zur ikonischen Prosodie. In diesem Kapitel untersuche ich eine neue Hypothese, dass die ikonische Prosodie der vertikalen Position und Größe in der Physiologie verwurzelt sein könnte. Auf der Basis der erhobenen Daten wurde keine Korrespondenz zwischen Objektgröße und Kieferöffnung gefunden. Die Ergebnisse für eine Korrespondenz zwischen vertikaler Objektposition und Kopfposition sind nicht schlüssig. Ich erörtere die möglichen anderen Faktoren, die die Ikonizität bestimmen könnten, sowie zukünftige Richtungen.

Im Kapitel 4 widme ich mich dem Thema der Ideophone. Ich stelle eine Datenbank mit deutschen Ideophonen, die aus Kinderbüchern gewonnen wurden, vor und führe mit Hilfe der Daten zwei Studien durch. Die erste Studie befasst sich mit der Klassifizierung von Ideophonen, wie sie von Dingemanse (2012) vorgeschlagen und von McLean (2020) überarbeitet wurde. Die Ergebnisse legen nahe, dass Ideophone im Deutschen multisensorische Eigenschaften aufweisen (Nuckolls, 2019). Die zweite Studie im Kapitel 4 untersucht, ob sich die Verwendung von Ideophonen in Büchern mit dem Alter verändert. Die Ergebnisse deuten darauf hin, dass Bücher für jüngere Kinder mehr Ideophone enthalten als Bücher für ältere Kinder, was die Hypothese des lautsymbolischen Bootstrapping im schriftlichen Bereich widerspiegelt (Imai und Kita, 2014).

In der Gesamtheit dieser Doktorarbeit befasse ich mich mit dem Oberthema der Ikonizität und ihrer Verbreitung. Das Hauptanliegen dieser Arbeit ist es, das Potenzial und die Bedeutung von Ikonizität in der heutigen Sprache zu untersuchen.

Abstract

Iconicity in Language and Speech

by Aleksandra Ćwiek

Iconicity refers to a resemblance between the linguistic form and the meaning of a referent (cf. Perniss and Vigliocco, 2014). Just like a sculpture resembles an object or a model, so can the sound or shape of words resemble the thing they refer to. In this thesis, I investigate the potential of iconic expression on different linguistic levels – the sound level, the prosodic level, and the word level.

While iconicity and arbitrariness have been debated and oftentimes confronted regarding language nowadays, many researchers discuss it with respect to language evolution. In Chapter 1, I give an insight into this ongoing debate. I depart from demonstrating the iconic potential of vocalizations for language evolution, next to gestures.

The following parts of my thesis are dedicated to iconicity in spoken and written language nowadays, assuming that iconicity is an important property of human language, present at different linguistic levels.

Starting from the smallest unit, the sound, in Chapter 2 I first examine the current literature and subsequently present work on the sound-symbolic nature of German Pokémon names. The obtained results are an essential addition to the growing body of research on Pokémonastics (Kawahara et al., 2018; Shih et al., 2019). However, based on the results, I propose a novel sound-symbolic functional load hypothesis suggesting that sound-symbolic expression differs across languages.

From the smallest units, I move on to iconic prosody in Chapter 3. Combining findings about the relation between head position and fundamental frequency, as well as the degree of jaw opening and formant frequencies, I test a novel hypothesis whether iconic prosody of vertical position and size may be rooted in physiology. No correspondence between size and jaw opening was found, and the results for correspondence between vertical position and head position are inconclusive. Therefore, I discuss other possible factors governing iconic prosody.

Finally, in Chapter 4, I investigate iconicity on the word level. Thus, I address the topic of ideophones – words which were said to be scarce in Western languages. I present a database of German ideophones collected from children's books and conduct two studies with the help of the data. The first study tackles the classification of ideophones, as proposed by Dingemanse (2012) and revised by McLean (2020). The results suggest that ideophones in German exhibit multisensory properties (Nuckolls, 2019). The second study within Chapter 4 investigates whether the use of ideophones in books changes with age. Here, the results suggest that books for younger children contain more ideophones than books for older children, thus echoing a sound-symbolic bootstrapping hypothesis in the written domain (Imai and Kita, 2014).

Throughout those chapters, I am concerned with the grand theme of iconicity and its widespread nature. The main point of this work is to explore the potential and the importance of iconicity in the language nowadays. All in all, the results of this dissertation show that iconicity is a crucial part of language.

Acknowledgements

I want to express my gratitude to my supervisors, Susanne Fuchs, Manfred Krifka, and Marcus Perlman, for their continuous support during my Ph.D. work. Susanne, you have never seized to inspire me professionally and personally, and I could not have wished for a better mentor during this period; thank you. Manfred, thank you for sharing your ideas and challenging my views, which let me grow as a scientist. Marcus, I am truly grateful that you kept inspiring me to become better and helped me realize what are the values of great research.

This work would not have been possible without the encouragement of Petra Wagner, who helped me discover my passion for science. I also thank Zofia Malisz for taking me by the hand in the earliest days of my research, and Mattias Heldner and Marcin Włodarczak, without whom I would not have made it on stage in Trondheim.

I also thank the members of the committee from the Humboldt University in Berlin for their time, energy, and help.

I am grateful to the team with whom I worked on vocalizations, especially to Susanne Fuchs, Marcus Perlman, and Bodo Winter. This project showed me that I am capable of big things! I feel honored and inspired to have worked with you. I thank Shigeto Kawahara for showing me that I can do research while watching my favorite childhood series. I also would like to thank Luisa Cimander for her invaluable help in the arduous collecting of the ideophone corpus.

I thank my colleagues from FB1 at Leibniz-Zentrum Allgemeine Sprachwissenschaft in Berlin for all your help and advice. I offer my special gratitude to Jörg Dreyer, who thought ahead to overcome all technical difficulties. I am also thankful to Heather Weston for her patience and support.

I am profoundly grateful to Monika Kotus, who supported me even in the darkest times, and thanks to whom I was able to find my path. Dziękuję Ci z całego serca. I also want to express my deep gratitude to Przemek Dobak, who continuously helped me stay on this path and regain my sense of self. Przemku, jestem Ci głęboko wdzięczna za Twoje wsparcie. I thank Zuzanna Dobak and Natalia Kuraszyńska for their friendship, love, and support.

I dedicate this thesis to my family. To my mother, Anna, and my father, Aleksander. Mamo, dziękuję Ci za nieustanną wiarę we mnie i danie mi możliwości do spełniania marzeń. Tato, dziękuję Ci za opiekę i pogodę ducha. To my everloving grandparents, Alicja and Longin. Kocham Was i dziękuję Wam.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Before I started working on this thesis, I was a proponent of the Saussurian view on linguistic signs. As a student, I was taught that language is arbitrary, and I took it as an axiom. However, later, I started reading on the topic of iconicity and became fascinated by the ubiquitous nature of this phenomenon. I believe that the best way for me to give this feeling to you is by sharing a moment from a solo journey I took to Japan. Follow me.

I quickly grasped that any preparation I did before coming there was not enough. The climate was nothing like in my home country, hot and humid; the sun was burning every piece of my skin. On the streets, I noticed plants that would never grow outside a controlled house environment during cold polish winters. At first, I was very surprised by the harsh sound of unknown animals, only later recognizing them as cicadas. Many people were wearing masks – and that before the pandemic. I rarely encountered familiar letters, and the surrounding orthography left me clueless. I felt that I could only rely on other people's knowledge of English or on the translation app on my phone. I felt alienated. One day, I woke up feeling sick, I believe as a result of sunstroke. I made my way to the nearest store to buy something that could help me. I used my phone for translating, however, the translation must not have been entirely correct, as the clerk could not understand what was wrong. Then, I intuitively started showing it. I frowned my eyebrows, closed my eyes, and turned down my lips in the sign of pain. I put the hand on my head and rotated it to express dizziness. Finally, I exhaled heavily and groaned in pain. I saw the clerk was beginning to understand. So, I started shivering my body and accompanied that with a fast periodic /b/-like sound, imitating the sound of shivering. In the end, I pointed to my stomach, puffed the cheeks, and expressed a muffled sound of pain.

I used iconicity to express something without having a common ground – I turned to manual gestures, facial expressions, and iconic vocalizations. I used signals that, by a sense of resemblance, transported the meaning despite the fact that the clerk and I did not share a common language. In the end, I received the help I needed, but most importantly, I felt connected. Instead of relying on language, I felt that I relied on something supralinguistic.

1.2 Definition and Outline

In this thesis, I investigate iconic expression on different linguistic levels – segmental, prosodic, and word-level. Iconicity is "any resemblance between certain properties of linguistic/communicative form (...) and certain sensori-motor and/or affective properties of corresponding referents" (Perniss and Vigliocco, 2014, 2). It is, hence, a correspondence between parts of the concept and parts of the way we refer to this concept. The expression of this correspondence is conveyed by iconicity, but so is the perception of it. There are iconic signals that are said to be understood across languages (e.g., Köhler 1929; Ohala 1994; Sapir 1929; also see Schmidtke et al. 2014 for a review), and some, which are considered language-specific (e.g., Dingemanse, 2019; Dingemanse et al., 2015; Iwasaki et al., 2007b). A linguistic background may thus be a filter to iconicity. However, current experimental evidence shows that iconicity seems to be an intrinsic mechanism extensively used to communicate additional meaning or enhance the existing one.

During the 54th Annual Meeting of the Societas Linguistica Europaea in September 2021, iconicity was voted the most favorite linguistic term by a variety of language scientists, from phonologists to syntacticians. Iconicity is a field of research that has greatly gained popularity in the past few decades (Dingemanse et al., 2015; Perniss et al., 2010). I argue that it is not a coincidence for two reasons. Firstly, because iconicity allows for grounding linguistic symbols at the dawn of language. And secondly, because iconicity is still abundant in languages nowadays.

The lion's share of the current work is concerned with the second argument – the goal of this thesis is to show that iconicity still is a highly important feature of languages today. Therefore, in Chapters 2, 3, and 4 I experimentally investigate different areas of iconicity separately: sound symbolism, iconic prosody, and ideophones, with a focus on their written form. However, iconicity has been widely discussed within language evolution research. Thus, I devote a special place for the debate on the role of iconicity at the origin of communicative systems in this chapter, in particular its Section 1.4. In the following section, I start by discussing iconicity in relation to arbitrariness.

1.3 Iconicity vs. Arbitrariness of a Linguistic Sign

Every linguistic novice is at some point confronted with semiotics – the theory of signs (e.g., Deely, 2006; Ogden and Richards, 1923; Peirce, 1955; Saussure, 2011). As Peirce (1955, 99) puts it:

> A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*.

(A) (B) (C)

FIGURE 1.1: Examples of (A) a symbol, (B) an index (natural), and (C) an icon. (A) Hammer and sickle, a symbol representing a Marxist–Lenninist ideology; (B) a 23,000 years-old human footprint found in New Mexico as an index of human presence in the region (original picture by Dan Odess from Zimmer 2021); (C) slightly smiling emoji used in the iOS environment.

A sign is thus in relation to the world. A sign signifies something and, with this act, evokes a kind of response in the person perceiving the sign. Peirce specified over 60 different types of signs, but I would like to focus on the three that most likely every linguistic student learns about (Peirce, 1955, 102). And so, there are signs that are purely conventional in nature, which he calls *symbols* (Peirce, 1955, 112ff.). Hammer and sickle are a symbol for the Marxist–Leninist ideology (cf. Figure 1.1a). Both the hammer and the sickle symbolize something, that is, however, not the object itself. The next type of a sign is an *index*, and this sign is contiguous to the object it represents (Peirce, 1955, 107ff.). Causal relationships, such as those between fire and smoke, are indexical. Thus, smoke is an index of fire, just as footprints are indexes of the creature that left them (cf. Figure 1.1b). However, I have to note at this point that I am only referring to natural indexes here but there also exist nonnatural indexes, where the causality of the relationship is replaced by intentionality, such as in the case of demonstrative pronouns "this" or "that" (cf. Peirce, 1955, 110; Grice, 1957). Lastly, Peirce (1955, 104ff.) recognizes an *icon* as the third type. Here, the denotation relies on the resemblance between the sign and the object; therefore, the sign possesses certain object characteristics. We know icons from our computers and perceive the characteristics of what they stand for. Even more so, nowadays, emoticons (i.e., emotion icons) are a perfect example of icons, as shown in Figure 1.1c.

According to Everett (2017, 97) indexes present an onset in the evolution of language, followed by icons. Both types are non-arbitrary; they differ, however, in the intention of use. Indexes are non-intentional – an animal simply leaves footprints, old meat just smells foul. A criminal would wish not to leave any indexes on the crime scene. For natural indexes, there is a physical linkage between the cause and the consequence, the signaller and the signal, but there is no intention. In icons, "intentionality meets representation" (Everett, 2017, 99). The need to represent something is enabled by icons, through a direct resemblance. Early *Homo* show this need in carvings and art forms they left behind (Chauvet et al., 1996; Joordens et al., 2015).

Icons bridged the way toward intentional and arbitrary symbols; as Everett (2017, 99) puts it, they "played a role in the transition to symbols that was so crucial to the invention of language."

The idea of icons as carriers of later linguistic symbols is by no means a new one. Already Plato in *Cratylus* (Sedley, 2003) describes the natural origin of names. Plato's work heavily resonates sound symbolism, one of the notions within iconicity: "[b]ut is air called *aer* because it raises (*airei*) things from the earth, or because it is always flowing (*aei rhei*), or because wind arises from its flow? (. . .) The word *aithhir* (ether) I understand in this way: because it always runs and flows about the air (*aei thei peri ton aera rheon*), it may properly be called *aeitheira*" (410b)[1]. Throughout the dialogue, Socrates exposes before Hermogenes the natural meaning of certain names. Not only the names of gods are tackled, e.g., "[t]hen Pan, who declares and always moves (*aei polon*) all, is rightly called goat-herd (*aipolos*), being the double-natured son of Hermes, smooth in his upper parts, rough and goat-like in his lower parts (...)" (408c-d), but also things or ideas, like "[w]isdom (*fronisi*); for it is perception (*noesis*) of motion (*phoras*) and flowing (*tu*); or it might be understood as benefit (*onesis*) of motion (*phoras*); in either case it has to do with motion. And *gnome* (thought), if you please, certainly denotes contemplation and consideration of generation (*gones nomesis*); for to consider is the same as to contemplate. Or, if you please, *noesis* (intelligence) is merely *hesis* (desire) *tou neou* (of the new); but that things are new shows that they are always being generated (...)" (411d).

In the probably most well-known model of the sign, Saussure (2011) depicts the two sides of the sign – the signifier, which is the sound image, and the signified, which is the concept itself. Saussure's sign underlies two principles. The first principle is the arbitrary nature of the sign, by which he means that the motivation of the signifier is arbitrary "in that it has no natural connection with the signified" (Saussure, 2011, 69). This idea has been famously echoed by Hockett (1960, 90) in his 13 design features of human language, who says "[t]he word 'salt' is not salty nor granular; 'dog' is not 'canine'; 'whale' is a small word for a large object; 'microorganism' is the reverse." In later works, Hockett argues that visual communication is not necessarily arbitrary. It affords iconicity thanks to its multidimensionality as compared to a unidimensionality of speech (Hockett, 1978, 274f.). In this assumption, Hockett is unanimous with Saussure's second principle of the sign: the linear nature of the signifier (Saussure, 2011, 70). The auditory signifier is said to be but a line. Hockett's words, "[i]n one-dimensional projection, an elephant is indistinguishable from a woodshed" (Hockett, 1978, 275) corroborate this. Furthermore, both Hockett (1978, 274) and Saussure (2011, 70) mention prosodic prominence as a possible second dimension of speech, however, they contradict it could be the case, as "this is an illusion; the syllable and its accent constitute only one phonational act" (Saussure, 2011, 70). However, accents – on word- and sentence-level – have various prosodic

[1]Greek text indicated in the cursive font was transliterated from the original here and in further quotations. It does not depict the diacritics.

expressions in different languages (e.g., Ćwiek and Wagner, 2018; Szalontai et al., 2016), and, therefore, do not constitute a basis for a reliable argument. Rather, we must consider acoustic parameters as dimensions of speech: time, frequency, and intensity, as displayed on a spectrogram. Modern technology enables a visualization of acoustic signals in a three-dimensional space (e.g., *Real-time 3D Spectrogram now available in SignalScope for iOS* 2015), whereby the x is the frequency, y is the time, and z the amplitude (represented by colors and height). It clearly shows that speech is not to be treated as unidimensional. Thus, speech, just as visual communication, has the potential for iconic representation.

The main misconception about iconicity that led to its criticism in the past (e.g., Hockett, 1978; Pinker and Bloom, 1990) is that iconicity contradicts the arbitrariness of a sign, a fundamental theory in semiotics (Saussure, 2011). With the striking evidence on the widespread nature of iconicity (e.g., Blasi et al., 2016; Ćwiek et al., 2021; Perniss et al., 2010; Schmidtke et al., 2014), it is more plausible to think of both forces as complementary (Arbib et al., 2008; Dingemanse et al., 2015; Perniss and Vigliocco, 2014). Where iconicity cannot be afforded, arbitrariness comes into play (Gasser, 2004; Lupyan and Winter, 2018; Monaghan et al., 2014). In Peircian terms, icons serve the formation of symbols; they serve the conventionalization which leads to the apparent arbitrariness (cf. Abralin, 2021; Everett, 2017). Neither arbitrariness, nor iconicity is exclusive – and they may be opposing forces, but still, both function together.

Iconicity is more of a need in a language community. This need restricts arbitrariness, which is caused only by convention. Arbitrariness is not a principle; it is a side effect of establishing symbols by convention (Abralin, 2021). Repetitions of the signal lead to its conventionalization, and it has been shown that this process has a deleterious effect on iconicity in the given modality (Caldwell and Smith, 2012; Frishberg, 1975; Tomaszewski, 2006) or in another modality (Little et al. 2017; but see Perlman et al. 2015b; Verhoef et al. 2014). In turn, as Bodo Winter says in his recent Abralin Ao Vivo talk, "[l]anguages don't want to be arbitrary, but they find themselves to be arbitrary" (Abralin, 2021, 1:04:53).[2]

Leiss (1997, 136) points out that arbitrariness was nothing more than a convention itself, taken for granted by early philosophers such as Thomas Hobbes, but also John Locke. Subsequent researchers simply stuck to this convention. However, with the amounting evidence, iconicity becomes harder to deny. Current research shows that various characteristics may be expressed iconically, e.g., size (Gallace and Spence, 2006; Ohala, 1994; Pisanski, 2014), spatial relations (Mudd, 1963; Pisanski et al., 2017b; Pratt, 1930), or velocity (Perlman et al., 2015a; Speed et al., 2018). Sound symbolism (cf. Chapter 2), as a special case of iconicity, has been shown to be a powerful way to code meaning such as size (Kawahara et al., 2018; Pitcher et al., 2013; Ultan, 1978), distance (Traunmüller, 1996; Woodworth, 2009), and even various

[2]The talk is available on YouTube at: `https://youtu.be/R1ETw21oCGE` (last accessed: September 28, 2021).

product characteristics (Klink, 2000, 2001; Pathak and Calvert, 2020). Not to mention cross-modal correspondences (Spence, 2011), such as the well-known bouba/kiki effect (Köhler, 1929; Ramachandran and Hubbard, 2001), which is based on the correspondence between sound and shape. "Bouba" is famously connected with a round shape, and "kiki" with a spiky shape – across languages, across ages (cf. Bremner et al. 2013; D'Onofrio 2014; De Carolis et al. 2018; Fort et al. 2018; Maurer et al. 2006, but see Rogers and Ross 1975; Styles and Gawne 2017). Recently, Sidhu et al. (2021b) found that phonemes [b] and [u] predict object roundness, and the phoneme [k] its spikiness. Their results corroborate previous research (D'Onofrio, 2014; McCormick et al., 2015; Westbury et al., 2018) and speak for the crossmodal character of the effect in question — the effect whose foundations are the articulatory-acoustic properties of individual segments (Knoeferle et al., 2017). Those properties do not go unnoticed in the auditory perception. Abrupt (i.e., spiky) spectral changes caused by voiceless stops relate to the spikiness in the visual domain, thus, the relationship is based on iconicity. Uninterrupted voicing and, thus, fundamental frequency (Kohler, 1982; Westbury and Keating, 1986) evoke a sense of smoothness in perception, relating this to smoother, i.e., rounder objects (Ćwiek et al., 2022).

So why is iconicity so wide-spread and so pluripotential? In the next chapter, I will outline the role of iconicity for the foundations of communication. First, I will briefly discuss the accounts on language origin, focusing on the *gesture-first hypothesis* (e.g., Hewes, 1973). I will then move to the use of iconicity in animal communication. Finally, I will present recent evidence showing that the vocal domain has the iconic potential similar to the gestural domain.

1.4 Iconicity from the Evolutionary Perspective

There exist various perspectives on how language began. One of the widely-discussed ideas is the gestural or gesture-first hypothesis (Corballis, 2002; Stokoe, 2001; Żywiczyński and Wacewicz, 2019). It assumes that the gestural communication preceded vocal communication, and it has been pursued by scholars for centuries (e.g., Condillac 2001; Wundt 1973; also see Kendon 2017, 163). Experiments with primates in the second half of the 20th century served researchers to advocate the gesture-first hypothesis (Hewes, 1973). Gardner and Gardner (1969) were able to teach a chimpanzee pieces of the American Sign Language. Apes were, however, never able to learn to speak. Moreover, ape anatomy may not allow for producing all speech sounds as we know them. However, even with the limited anatomical circumstances, apes should be able to invent new vocal signals if the origin of language includes the oral–aural modality. Early research reports that apes do have various alarm calls (e.g., Cheney and Seyfarth, 1980; Seyfarth et al., 1980a; Seyfarth et al., 1980b), but their repertoire is limited. Seyfarth et al. (1980b) show that vervet monkeys use different alarm calls for various predators – leopards, eagles, and snakes – and they do, in fact, respond differently to each of these calls. But, as noted by

Hewes (1973), primate vocal behavior differs drastically from their gestural behavior. Apes control their forelimb movement and easily learn new gestures; however, they lack the same degree of control over their vocal apparatus (Hewes, 1973, 6f.).

A similar case can be observed when looking at the process of language acquisition in children – children learn gestures before they learn speech (Meguerditchian et al. 2011; see also Kendon 2017). After the publication of the book *Baby Signs: How to Talk to Your Baby Before Your Baby Can Talk* (Acredolo and Goodwyn, 1996), teaching hearing babies a rudimentary set of signs has boomed. Babies, who are intended to speak in the future but are unable to do it yet, can so establish a communication system with their caretakers (Barnes, 2010). Unlike this simplified signing system, sign languages are fully developed languages, with grammar, phonology, or prosody (e.g., Meir et al., 2007; Sandler, 1996, 1999; Sandler et al., 2020). An example of Nicaraguan Sign Language (Senghas et al., 2005), or even more recently Al-Sayyid Bedouin Sign Language (e.g., Sandler et al., 2020) – both of which emerged in deaf communities – show that there is a capacity to develop a full-fledged communication system within the gestural modality only (cf. Tomasello 2008; see Kendon 2017, 165). Without a common communication system, we can rely on indexes and icons to get a system off the groud. If it is possible to do it based on gestural signs only, the visual modality would bear prime before the vocal modality.

The importance of gestures in everyday speech is another argument echoed in favor of the gesture-first hypothesis. Even though we primarily use speech to communicate, gestures play a crucial role in the communicative act. Co-speech gestures are tightly coordinated with speech, they are mapped onto the prosodic structure and align with prominent units (e.g., Bosker and Peeters 2021; Mol et al. 2009; Perniss et al. 2015; Willems and Hagoort 2007; see Wagner et al. 2014 for an overview). Experimental evidence shows that speakers convey more information in the gestures when communicating face-to-face (Bavelas et al., 2008; Gerwing and Allison, 2011). And, according to the *lexical retrieval hypothesis*, gestures aid speech production, i.e., lexical retrieval, during hestitation (Butterworth and Beattie 1978; Dittmann and Llewellyn 1969; after Wagner et al. 2014).

An upright posture in the *Homo erectus* ca. 2 million years ago freed up the hands to collect food, carry things, move things out of the way, scavenge, and so on (see Niemitz, 2010, for an overview). It might also have made it even easier to communicate with gestures. However, then, the hands are occupied with communicating, and one cannot perform other actions, such as crafting, carrying and so on. When using vocal communication instead, the hands are free to craft, gather food, eat, or perform other manual tasks. The superiority of the oral–aural modality is evident in conditions when parties do not see each other, like in the canopy of trees or at night (Ćwiek et al., 2021, 7). It also enables communication without being in the same. We can hear an acoustically prominent signal, like a scream, across large distances; however, we cannot always see the gesture from far away.

On the other hand, gestural communication is silent. Imagine a special unit approaching a terrorist hide-out. Their chunky boots step lightly on the concrete staircase. The commander gives signals with his hands. He may even rhythmically count down until the moment of the strike. Before the strike, the unit has to remain as silent as it can. Now, imagine our ancestors hunting. They crept softly through the bushes, barely letting any sound slip, watching every step, every branch that could crack under their feet. In such a situation, they had a purpose in remaining silent and they would most likely sign to communicate with each other. There is the right place and time to choose one modality over the other.

Nevertheless, modality preference aside, there is a growing body of research showing the potential of voluntary behavior in the vocal apparatus of primates, against some of the previously mentioned accounts (e.g., Hewes, 1973). The recent evidence shows that primates exhibit voluntary control over their lip and tongue movements, including rhythmic lip-smacking (Bergman 2013; Fitch 2010; Ghazanfar et al. 2012; see Kendon 2017, 167). Koko, a well-known human-fostered female gorilla, exhibited advanced control over her breathing behavior (see Perlman, 2017, for an overview). She would blow air gently onto the face of someone she liked and snort onto the face of someone she did not like. According to Kendon (2017, 164), ape communication with gestures is overrated, as "gestures of a depictive or referential nature have not been reliably observed". Also, successful experiments with primates learning symbolic communication (Wallman, 1992), like the one with the male bonobo Kanzi (Hopkins and Savage-Rumbaugh, 1991), only prove that primates are capable of learning symbols in extraordinary conditions. They do not, however, show that apes develop and use these symbols naturally (Kendon, 2017, 164).

Recent research also reports that primates expand their vocal repertoire with novel vocalizations. Lameira et al. (2015) showed that Tilda, an ex-entertainment orangutan, developed two novel vocalizations, which, in their structure, resembled human speech rhythm. Hopkins et al. (2007) showed that chimpanzees in captivity develop novel signals to capture the attention of caretakers. The results of Lameira et al. (2015) and Hopkins et al. (2007) show that primates developed spontaneous species-alike signals in captivity based on iconic imitation, e.g., of the speech rhythm. Therefore, current evidence suggests that both the visual and the oral–aural modality play a role in primate communication. Novel signals can be developed in both modalities.

The big question in the gesture-first theory is how did it happen that we primarily use the oral–aural modality nowadays if we had started with gesturing in the first place? In nature, nothing goes to waste, and economy is the basic principle. "Unconstrained, a motor system tend to default to a low-cost form of behavior" (Lindblom, 1990, 413). If there already exists a system, nature adapts it to what it is needed for, as adaptation is more economical than novel creation. The eyes in all organisms of the world evolved from simple photoreceptors that only distinguished light from dark in the beginning (Darwin, 2009a; Lee et al., 2011). So, why would we have speech that requires fine motoric coordination of the articulators if we had already had an excellent communication system with gestures? All higher mammals use vocal organs in order to communicate; evolutionarily, this was only a matter of time until *Homo* would do the same (Kendon, 2017, 166). The true question is about the complexity of the communication system we use. As mentioned, articulatory movements require enormous precision that did not evolve from one day to the other. According to the *social complexity hypothesis*, the more complex the social system, the more complex the communication system (Kendon, 2017, 167). Therefore, we may assume that our vocal apparatus evolved along with the expanding complexity of the social structure. The onset of communication included two systems alongside – gestural and oral–aural – which worked in cooperation to establish the creation of communicative symbols (e.g., Kendon, 1991, 2011, 2017; McNeill, 2012; Sandler, 2013).

However, redundancy is expensive, so why would nature allow two systems to do the same thing? In a recent study, Macuch Silva et al. (2020) showed that multimodal communication is more efficient than gesture-only communication (but see Fay et al., 2013; Fay and Lim, 2010; Fay et al., 2014). Using a referential communication task, the researchers asked participants to describe vocal or visual stimuli in three different conditions – gesture-only, non-linguistic vocalization-only, and multimodal. For both stimuli types, the most efficient mode of communication was by using both gesturing and vocalizing. The multimodal condition was optional, i.e., the participants could choose to use a single modality of the two or both modalities. Some participants initially used only one modality, and when they felt they could not reach the communicative goal, they turned to multimodality. This adaptation, despite multimodality being apparently less economical, allowed them for a more efficient communication. These results demonstrate that the gestural and oral–aural communication are not redundant; rather, the two modalities support one another (Macuch Silva et al. 2020; also see Sandler 2013). The gestural and the oral–aural communication have different semiotic affordances that both serve the goal of communication (Pouw et al., 2021, 270).

Language is, however, not only multimodal to the core, but it is also iconic (cf. Perlman, 2017). In Chapter 1.3, I outlined that icons are intentional (cf. Everett, 2017, 99). They are intentional in that they represent something by the direct link, through the immediate resemblance (Peirce, 1955, 104ff.). Iconicity is fundamental to the visual modality – we craft figures of people to represent them, we draw pictures to depict the scenery – and sign languages exhibit a high degree of iconicity (e.g., Taub, 2001). But, can the oral–aural modality be fundamentally iconic?

Evidence shows that languages generally start off being iconic but shift towards arbitrariness with time. This process takes place in the gestural modality as well (e.g., Frishberg, 1975; Pietrandrea, 2002, for examples in sign languages). We observe the language – spoken or written – from the synchronic perspective, and resources on language history date back only a few thousand years. Let us say that we take the onset of the *Homo* genus as the beginning of the communication system.[3] The earliest of the *Homo*, *Homo habilis*, is said to have walked the earth in the Early Pleistocene about 2.8 million years ago (Everett, 2017, 35). The oldest proto-writing system discovered in Jiahu, China, dates back to 6,600 BC and is based on a set of iconic carvings (Li et al., 2003). Iconic graphical representations become symbolic through the repeated use (Garrod et al., 2007). Current writing systems do not bear that high a degree of iconic resemblance, and neither does speech. During development, languages constantly change. The system becomes more and more complex, sounds and their combinations shift, and as a result, "iconicity can be obscured or overlaid" (Kendon, 2017, 167). Flaksman (2020) points out various causes of the process she calls *de-iconization*. Semantically, iconic words lose their character when they become embedded into the structure of the language, e.g., receive verbal endings (Flaksman, 2020, 83). Similarly to Kendon (2017, 167), Flaksman (2020, 86ff.) mentions that "sound changes tend to obscure the original sound-sense link existing in iconic word" (Flaksman 2020, 87; see also Klimova 1986; Voronin 2006).

Through repeated intentional action, icons undergo ritualization in the sense of Haiman (1994, 4). The form is emancipated from the meaning; it becomes a mere symbol of what it initially was, and it only denotes the meaning (Haiman, 1994, 16). We perceive language as arbitrary now because large parts of the vocabulary do not exhibit iconic motivation (Lupyan and Winter, 2018). The iconic resemblance between the form and the meaning was diluted over time; therefore, the form is perceived as symbolic rather than as iconic. It does not, however, mean that spoken languages are not iconic. It only implies that to search for iconicity, we need to look further than on the surface.

In the following chapter, I will give insight into iconicity in animal communication. The research reported here grounds the message of the current chapter – that language began as multimodal and iconic.

1.4.1 Iconicity in Non-Human Communication

Animals communicate using a variety of modalities, from vibration, like the spiders dancing (Barth, 1997), to vision, like the birds showing off their feathers (Osorio and Vorobyev, 2008). However, visual signals are ineffective in dense rainforests; therefore, many species turn to olfaction to communicate over greater distances. Plants serve as a prime example. The Sumatran rainforest is home to the biggest flowering plants on Earth, including the *Rafflesia* species and *Amorphophallus titanum*. The flowers of *Rafflesia* species both look and smell like rotting flesh (see Figure 1.2). The flowering cycle of these plants is very short and lasts no longer than a couple of days.

[3] Even though the earliest proto-primate is dated to have existed around 58 million years ago (Everett, 2017, 36).

FIGURE 1.2: *Rafflesia Arnoldii* in bloom. *Rafflesia* flowers have five petals, the structure of which resembles rotting flesh. A strong smell of rotting flesh is emitted from the inside of the flower to attract pollinators. The photo was taken by Frank (`https://www.flickr.com/photos/genlab/8524285694/in/album-72157632904563046/`), and is available under a CC-BY-NC-ND 2.0 license (`https://creativecommons.org/licenses/by-nc-nd/2.0/`).

The olfactory and visual mimicry attracts insects that search for feces or dead animals to lay their eggs. Insect visitors attracted in such a way help to pollinate plants that grow far from each other. Researchers found that the mechanism evolved independently in different species on earth (Jürgens et al., 2013). To boost the olfactory signal, the central part of the of *Amorphophallus titanum* flower, which can reach up to 3 meters in height, heats itself during prime ripeness. With outside temperature at 27°C, Barthlott et al. (2009) noted the surface temperature of the plant at 36°C. Heating the flower allows the plant to transmit the olfactory signal to further distances and, thus, to attract more insects. It seems like the visual communication is not enough. Smell serves as an additional channel for plants and many animals, as well. However, mammals are equipped with another means – vocal organs, which, according to Darwin (2009b) were made by nature to allow communication. Similar to olfaction, the oral–aural modality is beneficial in conditions with limited vision.

Even though the flowers of *Rafflesia* directly resemble rotting flesh, this behavior would have to be intentional for it to be an icon, according to Everett (2017). Without the intention, such signals remain indexes – just like the smell of the prey is an index left for the predator. Vocal properties are a similar index for body size in many animal species (Bowling et al., 2017; Gingras et al., 2013) and in humans (Pisanski et al., 2014a,b). The relationship is based on an inverse correlation between size and fundamental frequency (F0), whereby smaller vocalizers have a higher F0, and larger vocalizers have a lower F0. This link constitutes a basis for Ohala's "frequency code"

(Ohala, 1982, 1984, 1994), in which he postulates that this physiology-based relationship is responsible for iconic prosody (cf. Chapter 3). Frequency code implies that the originally indexical sign has transformed into an icon through the intentionality of use. The question is now whether animals can use their vocalizations intentionally or whether their vocalizations are rather reflex-like.

Dogs whine to show submissiveness or pain and bark in order to signal threat. Pigs squeal or grunt, depending on the situation. Animals have various vocalizations at their disposal that differ in their fundamental frequency and other qualities (e.g., Seyfarth et al., 1980b, for vervet monkeys). Forstmeier et al. (2009) show that zebra finch vocalizations are inherited and signal both body size and sex. However, there exist studies reporting that some species use vocal modulation for deception. Bee et al. (2000) found that male green frogs significantly lower their F0 to defend their territories from other males. The effect was the strongest in the smaller specimen and in response to large-male stimuli. Deception was also observed in two bird species – *Lanio versicolor* and *Thamnomanes schistogynus* – by Munn (1986). These species dishonestly employ alarm calls, which are used for predators, in order to deter other birds and to increase their chance of hunting flies. Also, Matrosova et al. (2007) found that pups of two different species of squirrel imitate the qualities of adult vocalizations (i.e., by lowering the fundamental frequency), most likely in order to cheat potential juvenile-hungry predators. In the visual modality, snapping shrimp (Hughes, 2000) and hermit crabs (Arnott and Elwood, 2010; Laidre, 2009) use dishonest signals to bluff their factual size.

However, this kind of deception may be interpreted as a mere adaptation bearing an evolutionary advantage rather than an intention. Some frog species were shown to adapt their vocalizations due to external circumstances, such as traffic noise (Cunnington and Fahrig, 2010; Lengagne, 2008; Parris et al., 2009). Vocal plasticity in noise-polluted areas has also been signaled for some bird species (Brumm et al., 2009; Moseley et al., 2018). Nevertheless, as Seyfarth and Cheney (2003) notice, the evolutionary advantage differs on the side of the vocalizer and the listener. On the callers' side, in favor are those specimen who can influence the behavior of others for their advantage (Seyfarth and Cheney, 2003, 168). On the listeners' side, in favor of the natural selection are those, who are better at recognizing the relevant information. Callers may not have an insight into the mental state of their kins, and they may not intend to inform others of their mental state with their calls (Seyfarth and Cheney, 2003, 159). This behavior, however, does not limit listeners in extracting this information from the vocalization anyway. So, "[l]isteners acquire information from signalers who do not, in the human sense, intend to provide it" (Seyfarth and Cheney, 2003, 168). Interestingly, Anderson et al. (2012) studied two types of song in the song sparrow – the soft song and the broadcast song. The former is a known signal of aggression within the species. Such a signal is only economical when it effectively intimidates a part of the competitors (Enquist 1985, after Anderson et al. 2012, 1446). However, Anderson et al. (2012) found that all rivals responded aggressively to the soft song. Thus, the soft song is not economical, as it does not intimidate the rivals; however, it is still sustained. These results are in line with Seyfarth and Cheney (2003, 168) and suggest that even if the signal is not intentional from the caller, it may be perceived as intentional by the listener.

There is no definite answer whether chimpanzees have an insight into what others are thinking, i.e., theory of mind. However, Seyfarth and Cheney (2003, 166f.) note that they may, in fact, be able to recognize the mental state of others. To the best of my knowledge, there exist two studies that explicitly tested the intentionality of chimpanzee calls, i.e., whether chimpanzees adapt their calls to the knowledge their kins have (Crockford et al., 2012; Schel et al., 2013). The results of both studies suggest that chimpanzees are aware of the knowledgeability of the other members of the group (but see Fischer and Price, 2017; Fitch, 2017). Based on their results, Schel et al. (2013) conclude that chimpanzees' alarm calls are intentional. I mentioned before that chimpanzees in captivity can develop novel signals to capture the attention of their caretakers (e.g., Hopkins et al., 2007). Combining Hopkins et al.'s 2007 observations with the results by Schel et al. (2013), we can assume that such novel signals are true intentional icons. A sign that is a pure indexical reflex stemming from some innate mechanism in frogs (e.g., Bee et al., 2000), might be taking a turn toward intentional use in other species. Additional support for this hypothesis are the results by Byrne and Corp (2004), who showed that, in primates, the size of the neocortex predicts the use of deception, i.e., intentionality. Their results suggest that the costly neocortical expansion was motivated by increasing social sophistication in primate groups (Byrne and Corp, 2004, 1697).

As put forth by Perlman (2017), there is mounting evidence showing that primates have control over their vocal behavior and are able to intentionally produce vocalizations to receive the attention of the caretakers (Leavens et al., 2010; Taglialatela et al., 2011; after Perlman, 2017, 381). Perlman (2017, 381) further refers to Lameira et al. (2016), who found that the orangutan Rocky developed a novel species-unlike vocal signal called a "wookie". When playing a "do-as-I-do" game, Rocky used this signal as a response towards the demonstrator, and even more so, Rocky modulated his fundamental frequency (Lameira et al., 2016; after Perlman, 2017, 381). In a 2012 study, Perlman et al. report that a mother gorilla Bawang uses iconic push-gestures to guide her infant Barney around their enclosure. The analysis reveals that Bawang uses force and trajectory of gesture as an iconic signal to enforce a movement in a given direction on her infant son. According to Perlman et al. (2012, 65), the understanding of the communicative intention need not take place, as the context itself creates the framing of the intention. Finally, Perlman and Clark (2015) analyzed the vocal and breathing-related behavior of Koko, a female human-fostered gorilla. After examining 439 instances of such behavior, the authors identified a set of characteristic vocal and breathing-related behaviors used by Koko in different situations. Manual gestures frequently accompanied their use. Altogether, Perlman and Clark's (2015) results might demonstrate the early stages of a full-fledged volitional control over vocalization.

Thus, primates have all of the prerequisites to construct iconic signals, but in the end, the interpretation of iconicity lies in the listener's hand. As Peirce (1955) formulates it, it all starts with indexes, which, by the intentionality, are formed to icons (cf. Everett, 2017, 97ff.). Whether the intention comes from the signaller or is only present at the interpreter's side is still a matter of discussion. However, the evidence provided in this section shows that the mental capacity of great apes, as shown by their behavior in numerous studies, allows them to use icons – intentionally.

Ultimately, some studies presented above also imply that intentionality, as a prerequisite to iconicity (Everett, 2017; Peirce, 1955), may not necessarily be the responsibility of the caller. The caller may not have the intention to reveal some information, but this information may still be recognized by their kins, i.e., the intention is perceived (Seyfarth and Cheney, 2003, 168). In this sense, we may perceive animal calls as intentional and recognize these calls as iconic, even though they are not meant this way. Are they, then, iconic or not? Is the perception of iconicity enough to call a signal "iconic"? The question seems philosophical, and to answer it, we have to remember that iconicity is "*any* resemblance between certain properties of linguistic/communicative form (...) certain sensori-motor and/or affective properties of corresponding referents"[4] (Perniss and Vigliocco, 2014, 2). Hence, a *perceived* resemblance between the form and the meaning is just as iconic. Now, it remains to ask a question of how vocal iconicity may serve symbol grounding in human communication.

1.4.2 Human Iconic Vocalizations

In the previous sections, I showed that iconicity is prevalent, i.e., not limited to the visual modality (Chapter 1.3). I also pointed to the importance of the auditory modality in the origins of language (Chapter 1.4). In the following, I will present evidence on how non-linguistic iconic vocalizations can serve to form symbols and signs.

Iconicity serves as a bridge "for the cognitive system to connect communicative form with experience of the world" (Perniss and Vigliocco, 2014, 2). This connection based on iconicity – the motivation of the sign – is essential to ground meaning (e.g., Cuskley and Kirby, 2013; Nölle et al., 2020; Perniss et al., 2010; Żywiczyński et al., 2021). Scholars have argued about the dominant modality in the process of symbol creation for years. While some accounts suggest that gestures bear prime in transporting the meaning (Fay et al., 2013, 2014), other sources signal the potential of vocalizations (Ćwiek et al., 2021; Edmiston et al., 2018; Erben Johansson et al., 2021; Perlman, 2011; Perlman and Lupyan, 2018).

The dominant role of gestures is echoed in the work of Fay et al. (2013, 2014). Using a referential communication task, Fay et al. (2013, 2014) tested the effectiveness

[4]Italics by the author.

and efficiency of novel signals in three modality conditions. Participants could either only use gestures, only vocalize, or use both modalities. In contrast to Fay et al. (2013), Fay et al. (2014) changed the role of the participants (signaller vs. perceiver) between the rounds to additionally test the alignment of the signals. Regarding alignment, Fay et al. (2014) found that it is the strongest in the gesture condition, and it also trumps the communication effectiveness. Overall, both studies suggest that gesture is both more effective and more efficient than vocalization. The multimodal condition yields no advantage in Fay et al.'s (2013; 2014) studies. However, a recent experiment by Macuch Silva et al. (2020) demonstrates that multimodality makes the communication of novel signals more efficient (cf. Chapter 1.4).

However, numerous accounts also demonstrate the iconic capacity of vocalizations. In a 2021 study, Erben Johansson et al. showed how iconicity in vocalizations can emerge over imitation. In the beginning, the researchers designed one identical seed word. With this word, they started a vocal imitation game towards four different meanings, which were revealed to the participants in different conditions. The participants imitated each others word one after the other, knowing that it means *big*, *small*, *pointy*, or *round*. No instructions about the meaning were given in the control condition. The final words, after 15 rounds of imitations, were very different from each other and complied with the predictions of size and shape sound symbolism (Erben Johansson et al., 2021, 11, also cf. Chapter 2).

In another study, Edmiston et al. (2018) used real-word sounds, rather than made-up words, as seed sounds for their imitation study. Participants were asked to imitate the sounds of *glass (breaking)*, *paper (tearing)*, *water (splashing)*, and *zipper (moving)*. The seed sounds were imitated for up to eight generations. In general, Edmiston et al. (2018) show that vocal imitations become more stable, thus, more word-like, generation by generation. After eight generations, the imitations lost their individual characteristics, i.e., differentiating one water sound from the other water sound; however, they maintained their category traits. Hence, vocalizations preserve the motivation, the connection to the real world and may play a crucial role in creating new words (Edmiston et al., 2018, 8).

The above findings find resonance in Perlman et al. (2015b), where the authors report that people can create ad hoc vocalizations to communicate an array of adjectival concepts. In their experiments, vocalizations became more stable over repeated rounds, which was echoed by higher guessing accuracy in the comprehension experiment. Finally, Perlman et al. (2015b) reveal that enhanced iconicity trumps the guessing rate.

In 2018, Perlman and Lupyan administered a contest for participants to submit their vocalizations communicating 30 paleolithic-relevant meanings spanning nouns, actions, and properties. The winner of the contest was chosen based on the ability of native speakers of English to guess the intended meaning. The vocalizations of the winning team were guessed with an overall 56% accuracy. From among the meanings, 27 out of 30 were guessed with an accuracy above the chance-level

(Perlman and Lupyan, 2018, 4). Thus, humans can create and comprehend vocalizations to communicate a myriad of concepts. However, because both vocalizers and participants were shared a common culture, as they were all at least proficient in English, a question arises, how generalizable are those findings.

Addressing this issue, Ćwiek et al. (2021) conducted a cross-linguistic comprehension study based on Perlman and Lupyan (2018). The authors run two experiments, one online and one in the field. The online experiment conveyed 25 languages from nine language families, and the field experiment six languages (seven groups, including American and British English speakers) from four language families. In the online version, participants listened to 90 vocalizations comprising 30 meanings earlier tested by Perlman and Lupyan (2018). Each meaning was assigned the three best-guessed recordings, as found in the contest by Perlman and Lupyan (2018). Participants were asked to guess the intended meaning of the recording by choosing one of the six alternatives, one of which was correct, and five chosen randomly from the pool of incorrect answers. In the field version, Ćwiek et al. (2021) constructed a set-up appropriate for illiterate participants. Therefore, only the noun concepts were tested. Here, the participants saw 12 pictures in front of them, each corresponding to one noun concept, and they were asked to choose the intended meaning from among the pictures. The overall guessing accuracy in the online experiment was equal to 64.6%, with all of the meanings guessed correctly above chance level across the board. The results of word-class analyses corroborate earlier research (Perlman and Lupyan, 2018; Perry et al., 2015, 2018; Winter et al., 2017) with verbs being guessed most accurately (70.9%), followed by nouns (67.7%), adjectives (58.5%), and demonstratives (44.7%). In the field experiment, the overall guessing accuracy reached 55%, and all meanings were guessed correctly across the board. Here, participants were much more accurate with animate entities (72%) than inanimate entities (39%). The study by Ćwiek et al. (2021) provides evidence that people from disparate cultures can comprehend iconic vocalizations far beyond the expression of emotions. Their results show the potential of vocalizations to be a powerful means of transporting and possibly grounding meaning. Importantly, based on their findings, the authors advocate a multimodal origin of language. As they say, "[b]oth modalities, visual and auditory, could work together in tandem, each better suited for communicating under different environmental conditions (e.g., daylight vs. darkness), social contexts (e.g., whether the audience is positioned to better see or hear the signal), and the meaning to be expressed (e.g., animals, actions, emotions, abstractions)" (Ćwiek et al., 2021, 7).

To conclude the introduction to this thesis, I will now proceed to summarize the main concepts given throughout this chapter.

1.5 Mid Summary: Iconicity and Multimodality are Intertwined

After having reviewed a large amount of the relevant literature – however incomplete the review still is – my conclusion is that both iconicity and multimodality play a crucial role in language origins. The message of this chapter goes back to the thought originally brought by Perlman (2017, 376): "[l]anguage is iconic and multimodal to the core."

As shown above, much of the research to date has focused on the "either-or" perspective. Ergo, that language began either as gestures (e.g., Corballis, 2002; Tomasello, 2008), or as sounds (e.g., Dunbar, 1998; see also Fitch, 2010, for a review). And while we cannot travel back in time yet to know for sure, we can observe the behavior of primates and children during language acquisition. In the behavior of primates, children, and the whole of humankind, gestures and sounds are connected (Gentilucci and Volta, 2007; Imai and Kita, 2014; Wagner et al., 2014). Studies on prosodic prominence – "a property by which linguistic units are perceived as standing out from the environment" (Terken, 1991) – show that prominence can be captured by using manual gestures or tapping (Rathcke et al., 2021; Wagner, 2016; Wagner et al., 2019). Moreover, gestures and speech are temporally coordinated, respecting the different properties of the two motor systems – slower, heavier arms and faster, lighter tongue (Ćwiek and Fuchs, 2021). Even speech production itself is not truly unimodal, as our body receives tactile and proprioceptive feedback through articulatory movements. Language is multimodal now, and it may well have been multimodal since its early days.

The gesture- vs. vocalization-first debate does not only reflect on the topic of the modality. The prime of one modality over the other lies in the potential of the given modality to convey meaning. Without a common understanding – thus, a convention – communication can rely upon resemblance. Then, through its own form, the sign embodies the properties of its referent (Perniss and Vigliocco, 2014). Iconicity is crucial in grounding symbols, as it serves the basic goal of understanding without prior conventions (Fröhlich et al., 2019; Nölle et al., 2020; Zlatev, 2014). As noted by Dingemanse et al. (2015; see also Dingemanse, 2013; Meir, 2010), however, the communicative strength of iconicity is not equal across modalities. Iconicity is constrained by the modality such that the expression of a visual event is easy in the visual modality and consequently hard in the oral–aural modality. In turn, the expression of sound is easy in the oral–aural modality and hard in the visual modality. According to Dingemanse et al. (2015, 608), both modalities perform well in expressing size, repetition, temporal unfolding, or intensity, but the expression of abstract concepts or logical operators is difficult in both the oral–aural and visual modality. This fact is echoed by the results of Ćwiek et al.'s experiment (2021), where concepts that are connected to sound (i.e., *sleep* with snoring, *child* with a high-pitched cry, *water* with splashing) were recognized with a much higher accuracy than concepts without an imminent characteristic sound (i.e., *that*, *gather*, *dull*). Sign languages afford lots of iconicity; however, they, too, fall back upon metaphors "to express a vast range of abstract and concrete concepts using vivid visual imagery" (Taub, 2001, 97, referring here to the American Sign Language).

This modality constraint may at first seem to limit. However, it only does so if we focus on the unimodal (or apparently unimodal, like in the acoustic communication) approach. The limitation dissolves when the perspective shifts to a broader view of multimodality. It then becomes evident that iconicity is not restricted within one modality, but rather it can be shifted from one modality to the other. One modality can do what the other cannot; hence, they co-operate (Pouw et al., 2021, 270). Truly integrated into one system, gestures and vocalizations fulfill the communicative need (cf. Macuch Silva et al., 2020). The multimodality strengthens iconicity, and both iconicity and multimodality have been the forces that helped out ancestors get a communication system off the ground (Ćwiek et al., 2021; Fröhlich et al., 2019; Nölle et al., 2020; Perlman, 2017; Wacewicz and Żywczyński, 2017; Zlatev, 2014). Put short, multimodality is at play to display iconicity, and conversely, iconicity reinforces multimodality.

The following chapters of this dissertation are dedicated to three forms of iconicity – sound symbolism (Chapter 2), iconic prosody (Chapter 3), and ideophones (Chapter 4). Each of these chapters is a stand-alone research project; however, they all investigate a form of expression of iconicity and are, therefore, connected to one another. In Chapter 2, I focus on sound symbolism. I first examine the current literature and subsequently present work on the sound symbolic nature of German Pokémon names. It is followed by research on iconic prosody in Chapter 3. In this chapter, I investigate a novel hypothesis that the iconic prosody of vertical position and size may be rooted in physiology. Then, in Chapter 4, I address the topic of ideophones. I present a database of German ideophones collected from children's books and conduct two studies with the help of the data. The first study tackles the classification of ideophones, and the second study investigates whether the use of ideophones in books changes with age. Finally, in Chapter 5, I discuss the significance of the results, the limitations of this work, and future directions.

Rather than focusing on one specific expression of iconicity, this thesis investigates a multitude of iconic expressions. Each of these expression could truly serve as a basis for a separate thesis (cf. e.g., Akita, 2009; Dingemanse, 2011; Kwon, 2015; Magnus, 2001; Marttila, 2010; Reiter, 2011; Sidhu, 2019). This dissertation, however, was designed that way on purpose. First of all, in order to address the question, whether there is iconicity in the current (vocal) communication. And, by answering this question, in order to expose the pluripotential nature of iconicity in language today (cf. Abralin, 2021).

Chapter 2

Sound Symbolism

2.1 The Scope of Sound-Symbolic Research

Iconicity, generally, is a correspondence between certain properties of linguistic form and meaning (Perniss and Vigliocco, 2014, 1). Sound symbolism, specifically, is concerned with the relationship between certain sounds and their meaning (Hinton et al., 1994, i). Iconicity as a force is present across many different linguistic layers, including morphology, syntax, and discourse. It is particularly prominent at the level of speech sounds. Traditionally, linguists treat speech sounds – whether as phones or phonemes – as units that do not bear meaning. In the best case, and within a particular language, phonemes function as the smallest meaning-distinguishing units (Bussmann, 2006). However, contrary to the view that sounds do not bear meaning, iconicity research has uncovered correspondences between certain sounds and particular meanings or characteristics. This domain is called sound symbolism.

As early as 1929, Sapir compared various vowel pairs in both front-back and close-open dimensions. Most distinctly, he showed that the open vowel /a/ is associated with larger objects, whereas the close vowel /i/ with smaller objects. Sapir asked both English speaking children (aged 11–18), university students, and adults, plus seven Chinese speaking participants, whether the word *mal* symbolized a large or small table, in contrast to the word *mil*. After seeing that the majority of the participants preferred *mal* to represent a large table, Sapir concluded that "certain vowels and certain consonants 'sound bigger' than others" (Sapir, 1929, 235). This pioneering finding brought an avalanche of research on the symbolism of sounds (e.g., Bentley and Varon, 1933; Brown et al., 1955; Firth, 1935; Newman, 1933; Taylor and Taylor, 1962, among the earliest examples).

From this point onward, researchers looked for relationships between sound and a variety of meanings or characteristics, such as size (Berlin, 2006; Ohala, 1997; Sapir, 1929), spatial deixis (Johansson and Zlatev, 2013; Traunmüller, 1996), shape (Köhler, 1929; Ramachandran and Hubbard, 2001), but also touch (Winter, 2016), or taste (Simner et al., 2010). Some of them rely on phonological units, others on acoustic properties of given sounds, like the fundamental frequency (Ohala, 1994, e.g., the "frequency code"). They are referred to as cross-modal correspondences because features of the one sensory domain awaken a correspondence to the other sensory domain – like /a/ is connected to the sensation of "large" (Spence, 2011). The cross-linguistic, or – as sometimes called – universal, nature of those correspondences is disputed, for even in the seemingly robust vowel height and size relationship, there are exceptions. Diffloth (1994, 112) reports that in Bahnar, high vowels express largeness, and low vowels smallness, exactly opposite to the often-cited findings for English and Chinese participants by Sapir (1929). However, Diffloth (1994, 112f.) argues that this does not undermine the iconic nature of the vowel–size correspondence – different cultures come up with different solutions for the same opposition. The iconic nature of the sensory correspondence remains intact.

In a prominent study, Blasi et al. (2016) analyzed over 6000 word lists from 62% of the world's languages for sound and meaning associations. The meanings tested were 40 basic concepts, based on a Swadesh list (Swadesh, 1955). The authors found evidence for preferred and dispreferred sounds (N total = 23) for 30 of the concepts, such as "small", which was positively correlated with /i/ and /ç/, or "breasts", which correlated positively with /u/ and /m/, and negatively with /a/, /h/, and /r/. Blasi et al. excluded phylogenetic persistence and areal distribution as factors contributing to the effect and concluded that mechanisms such as sound symbolism or iconicity are a possible source.

Certain sounds occurring in concepts evoking a related sensory meaning are called phonesthemes (Firth, 1935). Oftentimes, the English onset cluster *gl-*, like in *glimmer, glitter, glow* etc., is given as an example to represent concepts related to light or vision (Bergen, 2004). In the anglocentric world *gl-* may thus seem iconic, yet, looking outside of the English-speaking communities, we find no evidence to support this view (cf. Kwon, 2016). In a Peircean view (1955), such instancies belong to the secondary iconicity. While primary iconicity relies on "the perception of an iconic ground (...)", secondary iconicity is rather a knowledge of the functional relationship between form and meaning (cf. Sonesson, 1994, 741). However, another possibility might be that, rather than iconic, many of the phonesthetic examples are systematic. According to Dingemanse et al. (2015, 604), systematicity is "a statistical relationship between the patterns of sound for a group of words and their usage". Because systematicity relies on phonological cues of a given language, such relationships should be regarded as language-specific and arbitrary – at least of the second level, given the Peircean theory (but see Mompean et al., 2020). However, as the name phonesthetic suggests, the awareness of this aesthetic use of speech sounds may be of help when finding a proper name to depict a certain characteristic (Crystal, 1995).

Scholars have also explored the cross-linguistic character of sound symbolism, i.e., outside of the scope of a given language. Brown et al. (1955) asked native speakers of English to translate the meaning of 21 antonym pairs in an unknown language. The pairs were presented auditorily and visually in Chinese, Czech, or Hindi and comprised meanings connected to sensory experiences, such as warm–cool, or heavy–light. Overall, the participants guessed the meaning of the words correctly above the chance level. Because the words used in the paradigm are by no means onomatopoeic in nature, the authors concluded that the astounding accuracy stems from sound symbolism, common to all humans. Subsequently, corroborating (e.g., D'Anselmo et al., 2019; Gebels, 1969; Maltzman et al., 1956; Weiss, 1966), but also contradicting evidence was brought to light (e.g., Brackbill and Little, 1957). Miron (1961) showed that Japanese and English native speakers judge front vowels and consonants as "pleasant" and "weak", while they perceive back vowels and consonants as more "unpleasant" and "strong". It has been reported that most of the documented world's languages use sound symbolism to express the diminutive (Nuckolls, 1999; Stolarski, 2015; Ultan, 1978) – either with a front vowel (most frequently with /i/) or with consonant palatalization (Alderete and Kochetov, 2017; Jurafsky, 1996; Nichols, 1971). In contrast, West-African languages use low tones and vowels to convey largeness, heaviness (Westermann, 1927). And as mentioned above, Blasi et al. (2016) found that both /i/ and the palatal sibilant /ɕ/ were positively correlated with the notion of "smallness" across languages worldwide. Taken together, this suggests that there are at least parts of sound symbolism that have a cross-linguistic character. In such cases, sound symbolism seems to be based on particular articulatory or acoustic properties of the sounds (see Shinohara and Kawahara 2010, but cf. Itagaki et al. 2020), which are cross-modally correlated with properties in the physical world.

In their book entitled *Sound symbolism*, Hinton et al. propose a taxonomy for different types of sound symbolism. Firstly, they distinguish *corporeal sound symbolism*, which comprises of non-segmentable utterances that are rarely written, such as "aaugh!" (i.e., the sound of disappointment) or "achoo!" (i.e., the sound of sneezing). Those utterances depict the emotional or physical state and not only cross-linguistic but also cross-species (Hinton et al., 1994, 2f.). Next, *imitative sound symbolism* describes instances of onomatopoeia and other imitations of sound and movement, like "knock" or "ding-dong". Those words are more prone to be conventionalized in a given language (Hinton et al., 1994, 3f.). On the other hand, *synesthetic sound symbolism* refers to the tendencies of certain sounds to depict certain properties of objects, like the case illustrated in the above paragraph. Hinton et al. (1994, 4) add that, despite a large amount of work demonstrating the cross-linguistic character of this type of sound symbolism, there also exists contrary evidence. Lastly, Hinton et al. (1994, 5) identify *conventional sound symbolism*, where specific phonemes or clusters are associated with particular meanings within the scope of a given language. The last category concerns cases of phonesthemes described above.

In the section below, I present examples of sound symbolism in names across various categories: brands, animals, and proper and fictitious names. Wandering through, I focus on the connection between certain sounds and the meanings they characterize, as can be found in the existing research. The overview shall serve as a basis to formulate hypotheses regarding the sound symbolism in German Pokémon names. Subsequently, I describe the experiment conducted to test the hypotheses and its results. Section 2.5 discusses the outcome of the current study and is followed by chapter conclusions.

2.2 Sound Symbolism of Names

"[O]ur friend Cratylus has been arguing about names; he says that they are natural and not conventional; not a portion of the human voice which men agree to use; but that there is a truth or correctness in them, which is the same for Hellenes as for barbarians" (Sedley, 2003). This quote, at the very beginning of Plato's dialogue *Cratylus*, introduces the notion of the iconic rather than arbitrary naming of things. There seems, however, to be the limit to how much iconicity there can be in a language (cf. Lupyan and Winter, 2018). Not all words can be iconic; nevertheless, plenty of new words are added to the language every year. Oxford English Dictionary reports having added more than 650 new words in the latest update (*Updates to the OED* 2020, accessed on November 12, 2020), and Merriam-Webster has a total of roughly 470,000 entries (*How many words are there in English? | Merriam-Webster* 2020, accessed on November 12, 2020). Are any of those words iconic? And if so, which? Among those words, there are names for a myriad of entities, states, or characteristics in dozens of categories. In the following, I disentangle a few of those categories and present how sound symbolism might contribute to name creation.

2.2.1 Brand and Product Names

We undoubtedly live in a world of consumption. All forms of media are saturated with advertisements. Companies and manufacturers fight over consumers' attention. In this overwhelming competition landscape, marketing researchers started wondering how to make a brand name more visible. Here, sound symbolism lends itself to be a powerful mechanism that can transmit information about the product characteristics of a particular brand: its size, appearance, or taste (Klink, 2000). Sound symbolism can be additionally enhanced when combined with semantic cues conveyed in a product name (Klink, 2001; Yorkston and Menon, 2004). However, researchers also show that the perception of sound symbolism differs across participants, e.g., due to gender (Klink, 2003), or the country of origin (Athaide and Klink, 2012).

Klink (2000) tested how contrasting phonemes in brand names are perceived concerning a set of adjectives. In the first study, most strikingly, listeners judged products with front vowels, as compared to back vowels, to be smaller, lighter (vs. darker), milder, thinner, softer, faster, colder, more bitter, more feminine, friendlier, weaker, lighter (vs. heavier), and prettier. Klink's findings on fricative–plosive, voiceless–voiced stops, and voiceless–voiced fricative oppositions were mixed. The second study tested a more authentic setting in which brand names were embedded in advertisements. All in all, the author concludes that product names can convey product meaning, such as size, speed, or other characteristics (Klink 2000, 18f. corroborated by Lowrey and Shrum 2007; Pathak and Calvert 2020; Pathak et al. 2020; Shrum and Lowrey 2007). Apart from sound symbolism, sound repetition was shown to impact consumers' choices positively (Argo et al., 2010). Notably, sound repetition is another iconic means frequently used in ideophones (Kentner 2017; cf. also Chapter 4.1).

Klink (2001) demonstrated that combining both semantic information and sound symbolism has an even better impact on brand meaning recognition. He compared novel names for, e.g., a shampoo, where *Polbee* served as a baseline, *Silbee* as a sound symbolic name (due to *s* and *i* being correlated with softness), and *Silsoft* was a combination of semantic and sound symbolism. The semantic and sound symbolic names revealed the strongest association with the relevant attribute and received the highest liking scores, thus indicating an additive effect of semantics and sound symbolism (Klink, 2001, 31). In another study, Yorkston and Menon (2004) showed that sound symbolism conveyed in a name can be influenced by additional – matching or mismatching – information about the product. Nonetheless, they conclude that, unless the product is discredited by alternate information, sound symbolism is an automatic process, contributing to the brand's overall perception.

Previous research also showed that consumers' gender additionally influences the perception of sound-symbolic brand names, thus allowing for specific gender targeting (Klink 2009; see also Moorthy et al. 2018). For example, female participants reacted more positively to brands with front vowels, as compared to male participants (Klink, 2009, 317). Overall, research suggests that, with means of sound symbolism and sound-symbolic cross-modal correspondences, brand marks and brand names can be coherently designed to help transmit brand meaning (Klink, 2003).

Apart from gender, the brand personality – "the set of human characteristics associated with a brand" (Aaker, 1997, 347) – is perceived differently across cultures (Aaker, 1997). The puzzling question is whether sound symbolism in brand names can be perceived similarly by native speakers of different languages. Even if a brand name is not sound-symbolic, its transfer into different languages may be a challenge for marketing executives. For example, *nimm2*, candy manufactured by a German brand Storck, translates to "take two", yet without being translated, the connotation is lost outside of the German-speaking communities. One of the Procter & Gamble brands, known globally as *Vicks*, had to adapt its name to *Wick* before conquering the German market, as both <Vick> [fɪk] and <Wicks> [wɪks] have sexual connotations in German (Schmidt and Neuendorff, 2008).

Based on various linguistic traits, such as sound, meaning, textual, and visual elements, Usunier and Shaner (2002) proposed a framework for linguistic brand name evaluation. The authors discern the difference between East Asian – in particular Chinese – and Western culture, pointing out a greater value of spelling in the West, as compared to a higher importance of the visual value in the East. They suggest that spelling should be kept to a minimum and not contain language-specific phonemes or consonant clusters to design a global brand name. Furthermore, the name should, e.g., be transparent without translation (like Kodak or Nokia) and should not refer to any cognates in a particular language (Usunier and Shaner, 2002, 226). Nevertheless, if the name is to be translated, short and semantically relevant names are the most preferable in Chinese-to-English brand name translations (Fetscherin et al., 2015). In another experiment, Athaide and Klink (2012) tested whether sound symbolism perception is maintained between English and Hindi consumers. Using a set of previously tested sound symbolic relationships in novel brand names (cf. Klink 2000 two paragraphs above), the authors show that 21 of the 31 effects tested in a previous study by Klink (2000) hold for Hindi speakers. For the remaining null effects, Athaide and Klink (2012, 210) report, e.g., some names having, opposite to English, semantic referents in Hindi or neighboring languages.

2.2.2 Animal Names

Some animal names are metaphorical. In such a case, one of the names becomes polysemic – exhibiting multiple meanings – which is ensured by the metaphorical use of the name for other species. For example, in Greek *Diplodus puntazzo* fish is referred to as *hýaina*, as it displays vertical stripes, similar to a hyena, an Afro-Asiatic carnivore (Guasparri, 2019, 81). Although, to the best of my knowledge, there is reasonably less research on symbolism in animal names than in other domains, the existing accounts show that animal names can also be sound-symbolic. Overall, the research suggests that, in certain languages, animal names show sound-symbolic patterns. Those patterns may help to differentiate specimens across species and also within a species.

In an analysis of bird and fish names in Huambisa, an indigenous language in Peru, Berlin (1992) showed that bird names have a much higher probability of containing the vowel [i] in the initial syllable than fish names, which in turn favor the vowel [a]. Also, obstruents with high acoustic frequency, like [p], are more frequent initial segments of bird names, as compared to segments with relatively lower frequency, like [k], which in turn occur initially in fish names. In Huambisa, bird and fish names seem to convey the symbolic meaning of their characteristics: birds fly high and are agile in the air, whereas fish swim low and continuously through the waters. Both high vertical space and speed have been previously noted as correspondences of high-frequency segments, and low vertical space and flowing movements as those of low-frequency segments (Berlin 1992, also cf. Chapter 3 for vertical space). Initial segments are generally perceptually prominent, because they allow for rapid lexical access and, thus, cue the meaning of the word (Ussishkin and Wedel, 2009), therefore, the sound symbolic association to a certain characteristic of the referent is present at the earliest convenience. Furthermore, Berlin (1992, 242ff.) shows that size–sound symbolism is also prevalent in bird names, where small birds more often bear [i] in the initial syllable, not only in Huambisa, but also in Wayampí (Tupian), Apalaí (Cariban), and Tzeltal (Mayan). The size–sound relationship holds for Huambisa fish names, too (Berlin, 1992, 247ff.).

As noted by Westermann (1937), some West African languages exhibit sound-symbolic patterns in animal naming as well. Particularly in Ewe, many animal names are based on vocalizations for those animals. Back vowels /o, u/ are used more frequently than front vowels /i, e, a/ to communicate the size of an animal. Interestingly, in his analyses, Westermann (1927, 1937) considers /a/ to be a front vowel. There is also some evidence for the specificities of consonants in animal names. While Berlin (1992) proposed /r/ and /l/ to depict the characteristic of anurans, Hays's (1994) analysis of New Guinean frog names revealed that /g/ seems to be much more universal in that matter. Hays (1994) concludes that despite /g/ being even more frequent in his sample, both /r/ and /g/ may have "anuran" characteristics, and their co-occurrence is most frequent in frog names, among other words tested.

Finally, Berlin (2006) replicated the sound-symbolic tendencies in animal naming with his students by asking them to name two birds: one larger and one smaller. He showed that they reliably named the larger bird using the vowels /o, u/ and voiced stops, in comparison to the vowels /i, e/ and voiceless stops, which were used more frequently when naming a smaller bird. Therefore, sound symbolism is more an inherent property of humans, which is more likely to come to light in smaller communities than in complex ones (Berlin, 2006, 40).

2.2.3 Proper Names

In many languages, names are still used to mark a person's gender but also to emphasize certain traits that parents would like for their children. Numerous books exist on names, their meaning, and origin, which parents-to-be reach to choose the appropriate name with wishful traits for their offspring. It has been frequently demonstrated that male and female first names have different phonological characteristics (e.g., Cassidy et al., 1999; Cutler et al., 1990). In Polish, the great majority of female given names ends with <-a>, as it is a female gender marker in Polish nouns in general, and there exist only a few exceptions, such as *Beatrycze* or *Naomi*, which are loan names from Latin and Hebrew, respectively. Nevertheless, the question here is whether phonological patterns can provide insight on the characteristic of the referent, such as their gender, and most importantly, whether these patterns are sound symbolic in nature?

Investigations on English given names showed that female names have more syllables than male names, and the latter are more prone to end with a stop consonant, while the former in a vowel (Cassidy et al., 1999, 368). Cutler et al. (1990) corroborate the finding that female names tend to be longer; also, the initial syllable in female names is prosodically weaker, while in male names, it is more often stressed. They also found that female names are more likely to contain a stressed vowel [i]. Pitcher et al. (2013) conducted a corpus analysis of names given to babies across England, Wales, New South Wales (Australia), and the US between 2001–2010. They considered only stressed syllables in each name and grouped /i, ɪ, iː, e, a͜ɪ, ɒ, ʌ/ as "small", and /uː, ə, æ, e͜ɪ, a͜ʊ, aː, o/ as "large" vowels. In their database, male names were significantly more inclined to bear a "large" sounding vowel than a "small" sounding vowel. The situation was the opposite in female names (Pitcher et al., 2013, 2ff.). A similar effect was observed in French first names. Suire et al. (2019) examined a corpus of French names from 1900–2009 and found that in the prominent last syllable, male names tend to contain low-frequency vowels[1], in comparison to female names, which were more prone to encompass high-frequency vowels. The reported phonological patterns also hold for English nicknames, where the symbolism of the vowel [i] in female names is pronounced, as well as of the voicing opposition in plosives between female and male names (Klerk and Bosch, 1997). In the latter, male names have more voiced plosives, while female names more voiceless ones.

The cases mentioned above align with the previously described tendencies for low-frequency sounds to be correlated with large entities and conversely high-frequency sounds with small entities (cf. "frequency code" Ohala 1994). To emphasize the symbolic nature of sounds in names, Sidhu and Pexman (2015) and Sidhu et al. (2016) provide evidence that despite the meaning associated with certain names, sound-symbolic effects persist. In their study, participants correctly matched round-sounding and sharp-sounding names to round and sharp silhouettes, respectively. In that sense, their experiment extends to a sound–shape cross-modal effect, similar to the famous bouba/kiki (originally baluba/takete, later also known as maluma/-takete) paradigm (cf. Chapter 1.4.2; also Köhler 1929; Ramachandran and Hubbard 2001). And even though certain associations with a name, be it semantic or individual, may override sound-symbolic effects to some extent, Sidhu and Pexman (2015, 2019) show that sound symbolism evokes metaphors to personality traits. In their study, a "round" personality is perceived as friendly and easygoing, while a "spiky" personality as determined and rigid (Sidhu and Pexman, 2019, 3).

In summary, previous research shows that proper names display characteristics based on which certain properties of the referent, such as their gender, can be determined. These characteristics employ sound-symbolic patterns found in other domains. Thus, sound symbolism may also allow for references in proper names to further, more metaphorical traits. As Sidhu and Pexman (2019) conclude, "it is possible that if *Rose* were to go by another name, she might not seem as sweet."

2.2.4 Fantasy Names

In the original version of the fairy tale "Snow White", the Grimm brothers did not name the seven dwarfs accompanying the soon-to-be-princess (Grimm and Grimm,

[1]Low-frequency sounds, or high-frequency sounds, respectively, relate to fundamental frequency, as specified by Ohala (1994) in 'frequency code.'

2013, 131ff.). Years later, in the 1937 animation film version produced by Walt Disney, the seven dwarfs were given a more central role in the production, and hence, names (Sharpsteen et al., 1937). Since then, they were known as Grumpy, Bashful, Happy, Doc, Sneezy, Sleepy, and Dopey. Therefore, each name was designed to reflect the character of a respective dwarf semantically. The Smurfs, initially introduced in a Belgian comic series created by Peyo (1958), show a similar semantic pattern in their names. The majority of the Smurfs are named after personality traits or professions: there is Clumsy Smurf, Brainy Smurf, Poet Smurf, and so on. In both examples, iconicity was not the primary concern and – if existent at all – only a by-product of the lexical meaning.

But when inventing a name for a character, an author can let their imagination loose. Some authors go as far as J.R.R. Tolkien and invent not only names but languages for their characters. In *The Lord of the Rings* (Tolkien, 1991), Tolkien applied sound symbolism to convey traits in names, e.g., Tom Bombadil sounds like a "jolly, rumbustious owner" (Smith, 2006, 5). Elsen (2005, 2006, 2014, 2015) reports various sound symbolic aspects connected to the character's personality in fantasy names:

1. Powerful magicians, scholars, and other characters connected to wisdom show Latin/Greek-like phonotactics, e.g.: *Salamir*, *Kalakaman*, *Muntagonus*;

2. Demons and other evil creatures contain back consonants, e.g.: *Ch'tuon* (demon), *An-Rukhbar* (demon), *Brazoragh* (orc);

3. Names of young and attractive female protagonists often end with *-a*, e.g.: *Agwira*, *Meena*, *Shayla*;

4. Young and/or small characters contain "light" vowels, mostly [i], in their names, e.g.: *Brin* (small, good-willed prince), *Elim* (childish prince), *Kelwitt* (good-willed, dolphin-like creature).

Interestingly, people can translate sound-symbolic information in a name to a visual representation, which is essential for fantasy figures. Davis et al. (2019) asked participants to draw creatures for twelve different nonce words, like *ackie*, *cougzer*, or *horgous*, instructing them to base their imagination about the appearance on the sound of the name. The words were previously rated on various traits such as size or masculinity. To check whether sound properties were truly depicted visually in participants' drawings, another group of raters was shown the drawings and had to pick the most fitting name from among four alternatives. In general, the authors show that the created drawing does depict the characteristics provided by the earlier rating study with names only, i.e., *keex* was rated as small and spiky, and in turn drawn smaller and spikier, too. Besides, participants were able to assign the name to the creature correctly. These findings speak for sound symbolism not only to be able to provide insight on a single modality but also to penetrate different modalities.

The above-mentioned studies reveal striking sound-symbolic patterns, yet the instances are scattered across many sources (like in the cases presented by Elsen, 2005, 2006, 2014, 2015). The world of Pokémon (sometimes called Pocket Monsters), a multimedia franchise created by Satoshi Tajiri in the 1990s (Hidaka, 1997; Masuda et al., 1996), presents an extraordinary example of a vast fantasy name universe. Pokémon are fighting creatures inspired by insects. They live freely in the world, in which Pokémon trainers may care for them and train them. The trainers compete against each other in gyms and leagues, and they may use a couple of Pokémon in the match. Curiously, the creatures themselves have vastly different types – Normal, Fire, Water, Electric, among others. Some can also evolve from Baby Pokémon to Stage 1 and subsequently Stage 2 level. Pokémon of the same species, but on different evolution stages have different names. All Pokémon also have size parameters and battle statistics, like hidden power (HP), attack, defense, and speed, which change only with the evolution stage. As a consequence, the Pokémon universe is a unique testbed for studying sound symbolism.

In a 2018 study, Kawahara et al. examined the effect of voiced obstruents, the number of moras, and vowel quality on Pokémon attributes height, weight, strength, and evolution level. The authors found positive correlations among all or almost all Pokémon characteristics with the number of voiced obstruents and the number of moras (see also Kawahara and Kumagai, 2021). They also showed that Pokémon names with high vowels in the initial syllable are comparably smaller and lighter. Furthermore, Japanese participants showed a preference for sibilants in flying Pokémon types, which supports the ancient sound-symbolic assumption worded by Socrates (Kawahara et al., 2020).

Subsequent research extended Kawahara et al.'s findings onto additional languages. Shih et al. (2019) investigated English, Japanese, Mandarin, Cantonese, Russian, and Korean and revealed that although those languages mark identical attributes with sound symbolism, they do so with different sound-symbolic means. Two attributes, which exhibited the most robust correlations across the languages – evolution stage and overall power – are, according to the authors, crucial for game success and, therefore, marked the strongest. Their sound symbolism is alike within a language, e.g., in Japanese both an increase in overall power and evolution leads to an increase in the number of moras. Besides that, the overall power statistic also correlates negatively with the number of labial consonants. Other languages in the sample lack moras and show a similar tendency for either the number of segments (English and Korean for both attributes, and Russian for evolution stage only), the number of syllables (Mandarin for both attributes, and Cantonese for evolution stage only), or the number of reduplications (which reliably decreases with increasing power and/or evolution in both Mandarin and Cantonese). The other attributes scrutinized by Shih et al. (2019), such as height and weight, exhibit more variable sound-symbolic marking across languages. Weight, for example, correlates positively with the number of moras and the number of voiced obstruents in Japanese, whereas in English, it does so reliably with the number of sonorants only. In Mandarin and Cantonese, Pokémon weight correlates positively with the number of syllables and negatively with the number of reduplications. In Korean, the only reliable correlate of weight is the number of central vowels, and in Russian, the number of both alveolar and velar consonants (Shih et al., 2019). Thus, although the attributes are reliably sound-symbolically marked across languages, the way they are marked is modulated by the specificities of a given language.

2.3 Sound Symbolism in Translation

Pokémon gained popularity across the world, and so the names of the creatures oftentimes differ from country to country. Let us take a Pokémon called Charmander in English as an example. In English, Charmander evolves to Charmeleon, and Charmeleon evolves to Charizard. Charmander's original name is ヒトカゲ 'Hitokage', which can be translated as 'salamander' or 'fire lizard'. Charmeleon is called リザード 'Lizardo' in Japanese, which is the English word 'lizard' written in Katakana. Charizard, however, is originally called リザードン 'Lizardon', so his name is the name of the previous stage Pokémon plus a suffix *-n*. In English, the unity between the three Pokémon is even more visible: all three names begin with the prefix *char-*, which means 'burnt.' Charmander is the 'burnt salamander,' Charmeleon – 'burnt chameleon,' and Charizard, the 'burnt lizard.' Like in many other parts of the world, Pokémon have been very successful in Germany, and all Pokémon names were invented for the German market. The identifying prefix *char-* was taken over in the German translation, which stems from English and not from the Japanese original. Glumanda, Glutexo, and Glurak – German translations of Charmander, Charmeleon, and Charizard – all start with the prefix *glu-* which refers to the German word *Glut* 'embers.' Glumanda, the German Charmander, is a *Salamander* [zalaˈmandɐ], 'salamander'. Glutexo – Charmeleon – stands for *Echse* [ˈɛksə] 'lizard'. The name for Glurak, Charizard in German, is the hardest to uncover, but stems most likely from the Latin name for 'dragon' – *draco* (in German: *Drache* [ˈdʁaxə]).

However, in some languages, Pokémon names did not receive a proper translation. As discussed by Shih et al. (2019, 30) regarding Russian, a loanword adaptation may occur in such cases. Russian fans were faced with English names pronounced in Russian, so e.g., Charmander is called [t͜ɕarmanˈdɛr], and Squirtle [skfɪrtl]. Polish Pokémon names are very much alike, so Charmander would be pronounced as [t͜ʃarˈmandɛr], and Squirtle as [skwirtl]. In both cases, English name translations were taken over and adapted to the phonological and phonotactic rules. Hence, <squ-> becomes [skf-] in Russian, and [skw-] in Polish. Consonant clusters are widespread in Slavic languages, e.g., Polish onsets and codas may include up to five segments (Śledziński, 2010; Wierzchowska, 1967). These differences between languages might demonstrate another possibility for sound symbolism to come to light in a given language. It may be the case that in languages that did not receive Pokémon names of their own, other adaptation mechanisms serve to symbolize Pokémon features. Interestingly, however, the Polish Pokémon fanbase proposed unofficial new names for Pokémon. Taking Charmander, Charmeleon, and Charizardas examples again; for Charmander the name Palindra was proposed, which consists of morphemes from the words *palić* 'to burn', and *salamandra* 'salamander'. Charmeleon was given an unofficial name of Palimeleon which is a cluster of *palić* 'to burn', and *kameleon* 'chameleon'. Lastly, for Charizard, two names were proposed: Paliszczur, from *palić* 'to burn', and *jaszczur* 'lizard' (however *szczur* on its own means 'rat'), and Paliran, consisting of *palić* 'to burn' and *waran* 'Varanus lizard'. The inspection of their translation technique repeatedly raises the relevance of the underlying semantics in studying Pokémon names.

Because of its persistence across language boundaries, we might suspect that sound symbolism will be transferred from one language to another. One of the great tests of translator artistry is the last book by James Joyce *Finnegans Wake*. Joyce himself describes the work "(...) it is as in a dream, the style gliding and unreal as the way it is in dreams" (Potts, 1986, 149). The book was first published in 1939, and many professional translators have since fallen victim to its rich sound symbolism. It reads "The fall (bababadalgharaghtakamminarronnkonnbronntonnerronntuonnthunntrovarrhounawnskawntoohoohoordenenthurnuk!) of a once wallstrait oldparr is retaled early in bed and later on life down through all christian minstrelsy" (Joyce, 2012, 3). So far, it was fully translated into eleven languages, but there exist alternative translations in the form of musical or audio-visual events.

A translator is faced with a choice of translating a name – whether it is a character, a place, or a novel product name – into the target language. Fernandes (2006) shows how different translation techniques are used in children's fantasy literature translated from English (EN) into Brazilian Portuguese (BP). Apart from a simple copying (e.g., *Harry Potter* in EN remains *Harry Potter* in BP), a translator may choose a semantically motivated translation (e.g., *cat* in EN becomes *gato* "cat"), substitute a name (e.g., *Harvey* in EN turns to *Ernesto* in BP), leave it or its parts out (e.g., *Polly Plummer* in EN as *Polly* in BP), or adapt it. The adaptation follows phonological and phonotactic rules of the target languages (Fernandes, 2006, 51ff.). In another study, Santoso and Setyaningsih (2020) analyzed how onomatopoeias (cf. Chapter 4.1) are translated from English to Indonesian in Finance Smurfs comic. In the majority of the analyzed data points, iconicity was maintained. However, there were several cases in which the translation iconically diverged from the original. It is possible, as signaled by Jawad (2010, 61), that onomatopoeias – as auditorily symbolic – evoke different connotations across language boundaries. Santoso and Setyaningsih (2020) also mention a difference in the phonological systems of English and Indonesian. Taking that into account, the multitude of degrees of freedom in translation does not allow for assuming a 1:1 transfer of iconicity from the source to the target language. As shown in a longitudinal study by Pischedda (2020), translation strategies within the same universe (here, Disney comics) may change over time.

Pogacar et al. (2017) studied the perception of original character names in Charles Dickens' *Oliver Twist; or, The Parish Boy's Progress* (1838) by Slovene and English native speakers. Both participant groups were able to distinguish positive and negative characters based on their name. The authors stress the importance of keeping original names to maintain sound symbolism. Nevertheless, in their study, both participant groups spoke the language of the name origin and were therefore familiar with the phonological and phonotactic rules. There is no evidence to suspect that with a participant group unfamiliar with the language of origin, the symbolic effect in the names would be retained. It seems that sound symbolism is – despite its abundance – prone to be "lost in translation".

The current study aims to shed light on the sound symbolic nature of German Pokémon names by examining which sound-symbolic patterns can cross language boundaries and in which disguise they do so. The main goal of this research is to expand our knowledge on the possibilities of sound symbolism to employ various linguistic features for similar purposes. It may furthermore serve as a basis to disentangle the preferences of one feature over another in the future.

2.4 Experimental Evidence: Sound Symbolism of German Pokémon Names

Pokémon reached Germany in 1999. The anime series *Pokémon: Indigo-Liga* (original title: *Pokémon: Indigo League;* Hidaka, 1997) was first broadcast on September 1 on RTL II, followed by the release of *Pokémon Rot* and *Pokémon Blau* for Nintendo Game Boy (original title: *Pokémon Red* and *Pokémon Blue* Masuda et al., 1996) on October 8. In the opening episode, the main character Ash – a 10-year-old boy, who is to become a Pokémon trainer – has to pick his very first Pokémon. In the German version, he may choose Bisasam, Glumanda, or Schiggy. In the English version the creatures are called Bulbasaur, Charmander, or Squirtle, respectively. Ash ends up with yet another Pokémon called Pikachu, whose name remains intact through all language versions.

The focus of this analysis, as well as those conducted by Kawahara et al. (2020), Kawahara and Kumagai (2019), Kawahara and Moore (2021), and Kawahara et al. (2018, n.d., 2019) and Shih et al. (2019, 2018), are the sound symbolic patterns in Pokémon names with regard to their attributes. Some predictions are formulated in the following section.

2.4.1 Method

In the following, I will first introduce the data – how it was collected, annotated, and structured. Then, I will describe the statistical methods applied for the analysis.

2.4.1.1 Data and Predictions

To investigate the effects of sound symbolism in German Pokémon names in the most comparable manner, the methodology used in previous research served as a basis. A data set of German Pokémon names, including the power statistics, was created and annotated according to German phonology. The annotated linguistic features comprise of (partly based on the analyses for English by Shih et al., 2019):

I **Prosodic features:** number of (N) syllables, N segments

II **Vowels:** N front vowels, N back vowels, N high vowels, N low vowels, N round vowels

III **Consonants:** N sonorants, N labial consonants, N voiced obstruents

The occurrences of each feature within a Pokémon name were calculated. Front vowels include /i, ɪ, y, ʏ, e, ø, ɛ, œ/, back vowels /u, ʊ, o, ɔ/, high vowels /i, ɪ, u, ʊ/, low vowels /a/, and round vowels /y, ʏ, ø, œ, u, ʊ, o, ɔ/. As for consonants, the sonorants include /m, n, ŋ, l, j/, the labial consonants /m, p, b, f, v/, and the voiced obstruents /b, d, g, v, z, ʁ/.

Like in Kawahara et al. (2018, 6), the above variables are treated as predictors for Pokémon attributes, which are considered outcome variables. The Pokémon attributes analyzed here include: (1) EVOLUTION, (2) HEIGHT, (3) WEIGHT, (4) HIDDEN POWER (HP), (5) ATTACK, (6) DEFENSE, (7) SPECIAL ATTACK, (8) SPECIAL DEFENSE, (9) SPEED, and (10) GENDER. EVOLUTION was coded as 0 for a Baby Pokémon and 1 or 2 for Stage 1 or Stage 2 Pokémon, respectively. GENDER of the Pokémon consists of male and female percentage values. Genderless Pokémon were removed from the analysis of GENDER. The rest of the attributes are numeric values, all were obtained from the Pokémon encyclopedia (Bulbagarden, 2004). Data was collected for a subset of the 893 released Pokémon – only Pokémon from Generation 1 were analyzed, leaving $N = 151$ Pokémon in total and $N = 140$ for the analysis of GENDER.

Based on the sound-symbolic literature reviewed above and previous research (Kawahara et al., 2018; Shih et al., 2019) several predictions were made:

(1) Larger (WEIGHT and HEIGHT) and stronger (HP) Pokémon have longer names (N syllables or N segments).

(2) Larger (WEIGHT and HEIGHT) and stronger (HP) Pokémon names contain more low and back vowels.

(3) Larger (WEIGHT and HEIGHT) and stronger (HP) Pokémon names contain more voiced obstruents.

The remaining analyses performed here are exploratory.

TABLE 2.1: Evidence categories for the Bayes factor (BF) taken from Wetzels and Wagenmakers (2012), after Jeffreys (1998).

Bayes factor			**Interpretation**
	>	100	Decisive evidence for H1
30	–	100	Very strong evidence for H1
10	–	30	Strong evidence for H1
3	–	10	Substantial evidence for H1
1	–	3	Anecdotal evidence for H1
	1		No evidence
1/3	–	1	Anecdotal evidence for H0
1/10	–	1/3	Substantial evidence for H0
1/30	–	1/10	Strong evidence for H0
1/100	–	1/30	Very strong evidence for H0
	<	1/100	Decisive evidence for H0

2.4.1.2 Statistical Analysis

A two-step statistical analysis was performed to ensure the reliability of the effects in question. The analysis included correlation analysis and subsequent data modelling (following Shih et al., 2019).

Similarly to Kawahara et al. (2018) and Shih et al. (2019), the initial analyses were based on calculating all possible correlations between the linguistic features and Pokémon attributes. Bayes factor was chosen as the most reliable method for calculating the correlations. The motivation for this is that a frequentist correlation allows for limited statements regarding the hypothesis. A correlation test like Pearson's or Spearman's compares a null hypothesis (H0, no correlation) with an alternative hypothesis (H1, a non-null correlation). Along with the r correlation coefficient, we obtain a p-value, which expresses the likelihood of an alternative, given that the null hypothesis is true. Thus, H0 can be rejected or not, but nothing can be said about the probability of the H1 (Wetzels and Wagenmakers, 2012; Winter, 2020). Bayesian alternatives allow for making those statements. Bayes factor (BF) estimates whether the data is more probable to occur given the null or the alternative hypothesis (Wetzels and Wagenmakers, 2012, 1059). The evidence categories for BF are given in Table 2.1. Based on the Bayes factor, the probability of the data under H0 or H1 can be calculated as well. For example, if $BF_{10} = 0.2$, the data is $p = 1/0.2$ (or $BF_{01} = 1/BF_{10}$), ergo five times more probable under the H0 compared to H1.

Before obtaining the correlations between the linguistic features and Pokémon attributes, the latter were checked for normal distribution. The values for HEIGHT, WEIGHT, HP, DEFENSE, and SPECIAL ATTACK were not normally distributed, therefore they were log-transformed prior to the correlation testing. Also, the factor GENDER was not normally distributed and log-transformed, but because some instances contained values equal to zero, e.g., when a Pokémon is 100% female or male, no

BF could be calculated. Therefore, for GENDER only the second step of the analysis was performed. In the second step, Bayesian models were computed for each of the Pokémon attributes with linguistic features as predictor variables. A thorough explanation of the logic behind the Bayesian inferential framework is given in Chapter 3.5.1. All variables were scaled before modeling. For EVOLUTION, an ordinal model was computed; for the other attributes, multiple linear regression models were calculated for each attribute separately with all linguistic features as predictors. As GENDER consist of Male and Female percentages, the Male parameter was used, consistent with Shih et al. (2019, 10). Default priors of the `brms` package were used for all models. All models included four chains with 4,000 iterations each and a total of 8,000 post-warm-up samples. There were no divergent transitions, and all $\hat{R}$ values were = 1.0, indicating successful convergence for all models.

The analyses were performed in `R` (R Core Team, 2019). The package `tidyverse` was used for data wrangling (Wickham, 2017), Bayes factor was determined with the `BayesFactor` package (Morey et al., 2018), Bayesian linear regression models were computed using the `brms` package (Bürkner, 2017, 2018), and the plots were generated with `ggplot2` (Wickham, 2016).

2.4.2 Results

In the subsequent section, I report the results of the correlation analyses performed for each group of linguistic features: prosodic, vowels, and consonants. Following, I will describe the models for each of the Pokémon attributes.

2.4.2.1 Correlations

As mentioned above, the correlations were calculated using the Bayes factor. The interpretation of its value is based on Wetzels and Wagenmakers (2012) and Table 2.1. Because Bayes factor values were calculated separately for all linguistic features and Pokémon attribute pairs, they are additionally listed in Table 2.2 for better readability.

Prosodic Factors The number of syllables shows only one correlation that gives evidence for H1 – a positive correlation with the attribute SPECIAL ATTACK, see Figure 2.1a. The Bayes factor value of H1 against H0 was $BF_{10} = 1.866$, thus the data is roughly two times less likely to occur under the null hypothesis. All of the other Pokémon attributes suggest evidence for H0 in their correlation scores with the number of syllables, therefore no effect. The correlation with EVOLUTION provides anecdotal evidence for H0, with $BF_{10} = 0.530$. The rest show substantial evidence for H0: DEFENSE, $BF_{10} = 0.266$; HP, $BF_{10} = 0.242$; HEIGHT, $BF_{10} = 0.216$; WEIGHT, $BF_{10} = 0.202$; ATTACK, $BF_{10} = 0.193$; SPECIAL DEFENSE and SPEED with each $BF_{10} = 0.190$.

TABLE 2.2: Bayes factor values for all linguistic features (horizontal) and Pokémon attribute (vertical) combinations. While red symbolizes evidence for the null hypothesis, green signals evidence for the alternative hypothesis. Yellow represents no evidence. The lighter shade stands for anecdotal, and the darker shade for substantial evidence. In the case of green, the strongest shade represents strong evidence.

	Syllables	Segments	Front vowels	Back vowels	High vowels	Low vowels	Round vowels	Sonorants	Labials	Voiced obstruents
Evolution	0.530	1.204	0.191	0.374	0.270	0.235	0.492	0.238	0.245	1.354
Height	0.216	0.467	0.275	0.289	5.136	0.191	0.283	0.194	0.724	0.241
Weight	0.202	0.229	0.189	0.190	1.070	0.189	0.189	0.217	11.145	0.189
HP	0.242	0.236	0.259	0.570	0.253	0.269	0.338	0.600	0.189	0.679
Attack	0.193	0.194	0.245	0.243	0.192	0.221	0.350	0.213	0.899	0.218
Defense	0.266	0.199	0.211	1.713	0.415	0.224	0.875	0.205	0.391	0.310
Special attack	1.866	0.612	0.193	0.220	0.877	0.268	0.206	0.669	1.059	0.199
Special defense	0.190	0.233	0.190	0.255	1.581	0.250	0.215	0.202	0.191	0.448
Speed	0.190	0.245	0.215	0.425	0.192	0.329	0.585	0.224	0.293	0.321

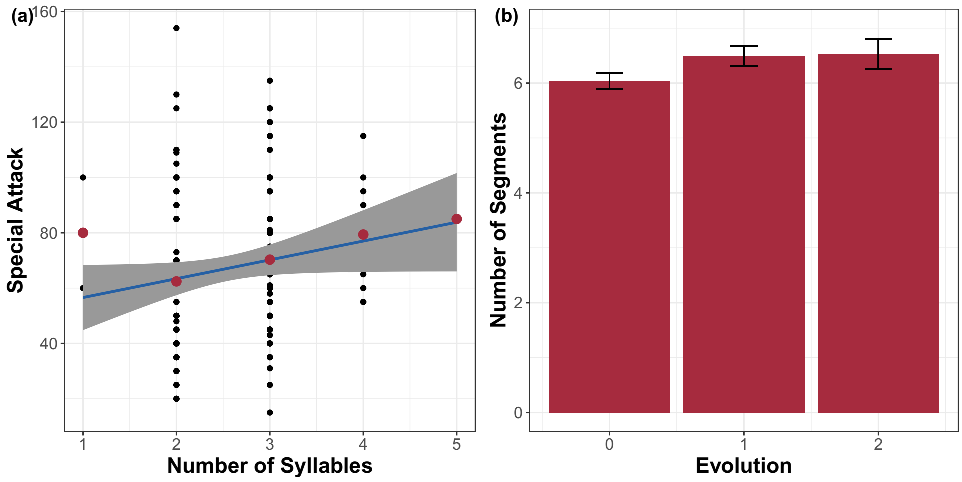


FIGURE 2.1: Relationship of prosodic factors with Pokémon attributes. (a) The number of syllables and SPECIAL ATTACK; the red dots indicate the group means; the linear regression is shown as a blue line, with the 95% confidence interval in gray. (b) The number of segments and EVOLUTION; the standard errors are shown as black bars for each Evolution level.

Similarly, the number of segments also shows a single correlation with evidence for H1. It correlates positively with EVOLUTION, as depicted in Figure 2.1b. The Bayes factor here is $BF_{10} = 1.204$, suggesting that the data is 1.2 times more likely to occur given the H1. This value is relatively lower, thus the correlation weaker, than the one between the number of syllables and SPECIAL ATTACK. This finding partly corroborates the results by Kawahara et al. (2018) and Shih et al. (2019). Apart from that, other attributes correlated with the number of segments provide evidence for H0, hence a correlation is not determinable, with the only anecdotal evidence for SPECIAL ATTACK with $BF_{10} = 0.612$. Substantial evidence for H0 is present in: HEIGHT, $BF_{10} = 0.467$; SPEED, $BF_{10} = 0.245$; HP, $BF_{10} = 0.236$; SPECIAL DEFENSE, $BF_{10} = 0.233$; WEIGHT, $BF_{10} = 0.229$; DEFENSE, $BF_{10} = 0.199$; and ATTACK with $BF_{10} = 0.194$.

The current analysis of the prosodic factors of Pokémon names shows no support for prediction (1) formulated earlier that larger (WEIGHT and HEIGHT) and stronger (HP) Pokémon have longer names.

Vowels The Bayes factor values for each of the attributes and the number of front vowels give substantial evidence for H0, rather than H1. As previously, in the descending order, the attributes received: HEIGHT, $BF_{10} = 0.275$; HP, $BF_{10} = 0.259$; ATTACK, $BF_{10} = 0.245$; SPEED, $BF_{10} = 0.215$; DEFENSE, $BF_{10} = 0.211$; SPECIAL ATTACK, $BF_{10} = 0.193$; EVOLUTION, $BF_{10} = 0.191$; SPECIAL DEFENSE, $BF_{10} = 0.190$; and WEIGHT with $BF_{10} = 0.189$.

The number of back vowels provides anecdotal evidence for H1 with the attribute DEFENSE with $BF_{10} = 1.713$. The correlation is positive, as shown in Figure

2.2a, and the data is 1.72 times more likely to occur, given the H1. The remaining attributes deliver evidence for H0, with anecdotal values for HP, $BF_{10} = 0.570$; SPEED, $BF_{10} = 0.425$; and EVOLUTION, $BF_{10} = 0.374$. Substantial evidence for H0 is given for: HEIGHT, $BF_{10} = 0.289$; SPECIAL DEFENSE, $BF_{10} = 0.255$; ATTACK, $BF_{10} = 0.243$; SPECIAL ATTACK, $BF_{10} = 0.220$; and WEIGHT, $BF_{10} = 0.190$.

As for the number of high vowels, substantial evidence for H1, according to Bayes factor, is given for the attribute HEIGHT, with $BF_{10} = 5.136$. Hence, the data is five times less likely to occur under the null hypothesis. As can be seen in Figure 2.2b, the number of high vowels and Pokémon height are negatively correlated, meaning that the more high vowels a Pokémon has in the name, the smaller it is. Next, the attribute SPECIAL DEFENSE yielded anecdotal evidence for H1 with $BF_{10} = 1.581$, so the data is 1.6 times more likely to occur under the alternative. The correlation is also negative, as shown in Figure 2.2c. No evidence could be obtained for the attribute WEIGHT, which returned $BF_{10} = 1.070$. The relationship between the number of high vowels and Pokémon WEIGHT is therefore inconclusive. The null hypothesis received anecdotal support with attributes SPECIAL ATTACK, $BF_{10} = 0.877$, and DEFENSE, $BF_{10} = 0.415$. The attributes: EVOLUTION, $BF_{10} = 0.270$; HP, $BF_{10} = 0.253$; SPEED, $BF_{10} = 0.192$; and ATTACK, $BF_{10} = 0.192$ provide substantial evidence for H0.

None of the Pokémon attributes correlates with the number of low vowels. All comparisons present substantial evidence for H0: SPEED, $BF_{10} = 0.329$; HP, $BF_{10} = 0.269$; SPECIAL ATTACK, $BF_{10} = 0.268$; SPECIAL DEFENSE, $BF_{10} = 0.250$; EVOLUTION, $BF_{10} = 0.235$; DEFENSE, $BF_{10} = 0.224$; ATTACK, $BF_{10} = 0.221$; HEIGHT, $BF_{10} = 0.191$; and WEIGHT with $BF_{10} = 0.189$.

The number of round vowels also provides only evidence for the null hypothesis. Anecdotal support is given for: DEFENSE, $BF_{10} = 0.875$; SPEED, $BF_{10} = 0.585$; EVOLUTION, $BF_{10} = 0.492$; ATTACK, $BF_{10} = 0.350$; and HP, $BF_{10} = 0.338$. In turn, the attributes: HEIGHT, $BF_{10} = 0.283$; SPECIAL DEFENSE, $BF_{10} = 0.215$; SPECIAL ATTACK, $BF_{10} = 0.206$; and WEIGHT, $BF_{10} = 0.189$ show substantial evidence for the null hypotheses.

Overall, the prediction (2) made earlier – that larger (WEIGHT and HEIGHT) and stronger (HP) Pokémon have more low and back vowels in their names – cannot be sustained based on the analysis so far.

Consonants The number of sonorant consonants yields no evidence for the H1 with any of the Pokémon attributes. Anecdotal evidence for H0 are present for attributes SPECIAL ATTACK, $BF_{10} = 0.669$, and HP, $BF_{10} = 0.600$. The remaining Pokémon attributes show substantial evidence for H0, therefore lacking a correlation with the number of sonorants. In the descending order these include: EVOLUTION, $BF_{10} = 0.238$; SPEED, $BF_{10} = 0.224$; WEIGHT, $BF_{10} = 0.217$; ATTACK, $BF_{10} = 0.213$; DEFENSE, $BF_{10} = 0.205$; SPECIAL DEFENSE, $BF_{10} = 0.202$; and HEIGHT, $BF_{10} = 0.194$.

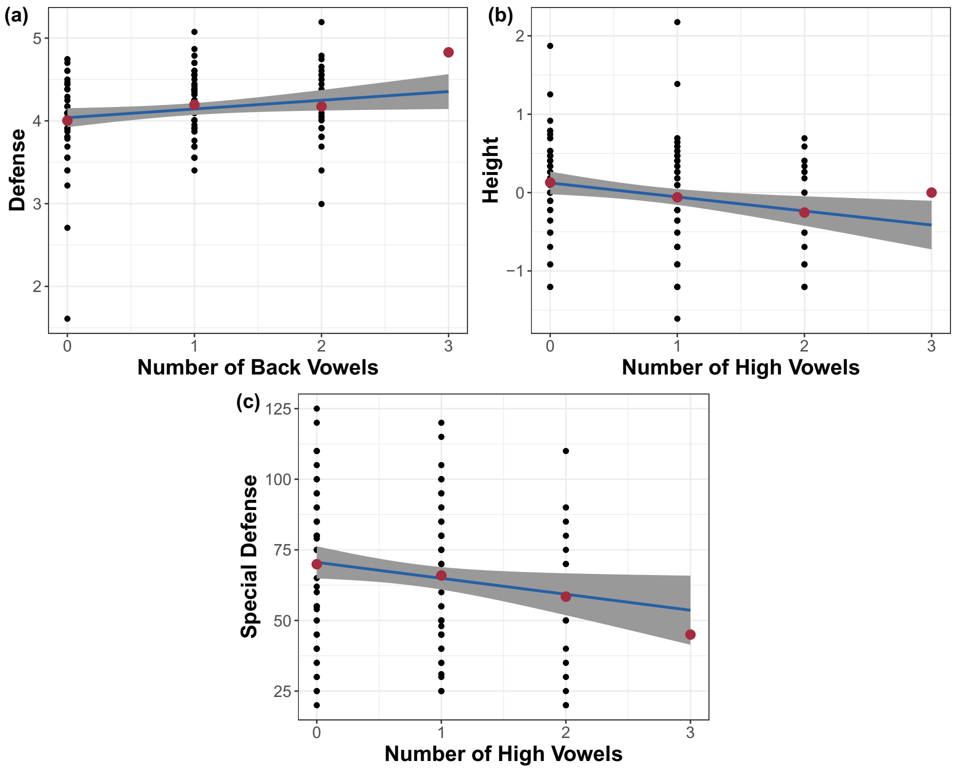


FIGURE 2.2: Relationship of the number of vowels with Pokémon attributes. (a) the number of back vowels and DEFENSE, (b) the number of high vowels and HEIGHT, and (c) the number of high vowels and SPECIAL DEFENSE. The red dots indicate the group means; the linear regression is shown as a blue line, with the 95% confidence interval in gray.

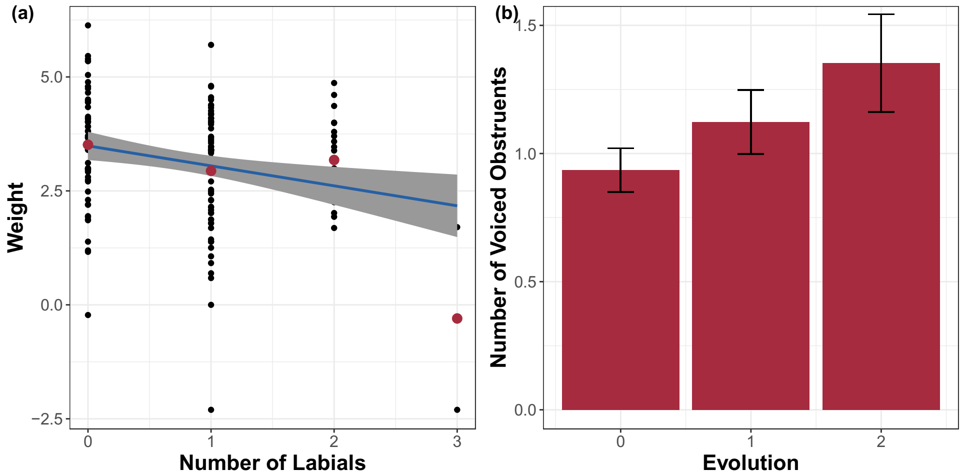


FIGURE 2.3: Relationship of the number of consonants with Pokémon attributes. (a) The number of labials and WEIGHT, (b) the number of obstruents and EVOLUTION. For plot (a), the red dots indicate the group means; the linear regression is shown as a blue line, with the 95% confidence interval in gray. For plot (b), the standard errors are shown as black bars for each Evolution level.

Further, the number of labial consonants shows strong evidence for H1 with the attribute WEIGHT. As given in Figure 2.3a, the correlation is negative and has a $BF_{10} = 11.145$, suggesting that the probability of the data occurring under the null equals 0.09. This finding is in line with the results obtained by Shih et al. (2019) for English Pokémon names, who also reported a negative correlation between weight and the number of labial consonants. No evidence could be obtained for the correlation of labial consonants and SPECIAL ATTACK, with $BF_{10} = 1.059$. The attributes ATTACK, $BF_{10} = 0.899$; HEIGHT, $BF_{10} = 0.724$; and DEFENSE, $BF_{10} = 0.391$ provide anecdotal evidence for H0, thus neither is correlated with the number of labial consonants in our data set. Substantial evidence for H0 is given for the attributes: SPEED, $BF_{10} = 0.293$; EVOLUTION, $BF_{10} = 0.245$; SPECIAL DEFENSE, $BF_{10} = 0.191$; and HP, $BF_{10} = 0.189$.

Finally, the number of voiced obstruents yields one anecdotal evidence for H1 – in its correlation with the attribute EVOLUTION. The Bayes factor here is $BF_{10} = 1.354$, meaning that the data is roughly 25% less likely to occur under the null hypothesis. The correlation, presented in Figure 2.3b, is positive, hence the more evolved the Pokémon, the more voiced obstruents it has in a name. Nevertheless, the rest of the Pokémon attributes do not show a correlation with the number of voiced obstruents in our data set. Anecdotal evidence for H0 is given for HP, $BF_{10} = 0.679$, and SPECIAL DEFENSE, $BF_{10} = 0.448$. Substantial evidence for H0 is present for: SPEED, $BF_{10} = 0.321$; DEFENSE, $BF_{10} = 0.310$; HEIGHT, $BF_{10} = 0.241$; ATTACK, $BF_{10} = 0.218$; SPECIAL ATTACK, $BF_{10} = 0.199$; and WEIGHT, $BF_{10} = 0.189$.

2.4.2.2 Bayesian Models

As mentioned above, the second step of the analysis involves computing a model for each of the Pokémon attributes. The Bayesian inferential framework was used for a more reliable testing in the case of the small number of data points. A detailed introduction to the advantages of Bayesian models and their interpretation is given in Chapter 3.5.1, though a few essential pieces of information are still needed at this point. Based on the real-world data, the model priors, and the model itself, certain predictions are made about how the data is distributed (posterior distributions). The model data reported below comprises four basic values: the estimate value, the 95% credible interval (CrI), and two posterior probabilities – of the posterior being above, and below zero. The estimate refers to the central point of the posterior probability. The 95% CrI shows where – with a 95% probability – an effect could be found in the posterior data. The posterior probability itself expresses the probability of the posterior distribution to be above or below zero, where zero expresses no difference made by the parameter. The two values for the posterior probability should be contradictory to one another. They signify the direction of the effect, similar to a positive or negative correlation, i.e., the posterior probability above zero shows a positive effect, whereas the posterior probability below zero – a negative effect of the tested parameter. Here, I assume that the posterior probability equal to or greater than 0.95 signifies an existing effect. The 95% CrI is an additional indicator that the tested parameter influences the outcome variable, in which case the 95% CrI should not include the threshold 0.

Evolution Table 2.3 and Figure 2.4 summarize the results of the model. Given the data, the priors, and the model, none of the linguistic features, except the number of segments, show an effect on Pokémon EVOLUTION; yet this, too, is unreliable. The number of segments yields a posterior probability $\beta > 0$ of 0.95, suggesting that a posterior sample falls above the threshold 0 with a 95% chance. However, the 95% CrI for the number of segments contains 0, $\beta = 0.44[-0.08, 0.95]$. This effect was previously signaled by Shih et al. (2019) for English. A correlation between the number of segments and EVOLUTION was also established in the previous step of the analysis. However, the analysis with Bayes factor also showed a relationship between the number of voiced obstruents and evolution, which is not mirrored here, given the data, the model, and the priors. The posterior probability $\beta > 0$ and $\beta < 0$ of the rest of the linguistic features does not meet the required 95% or more. Also, the 95% CrI of those features contains the threshold 0, meaning that the effect of these features is not determined as positive or negative.

Height Table 2.4 and Figure 2.5 summarize the results of the model for Pokémon HEIGHT. Given the current data, the model, and the priors, the number of labial consonants yields a negative effect on Pokémon HEIGHT ($\beta = -0.19[-0.36, -0.02]$) with a posterior probability $\beta < 0$ of 0.99. The results suggest that having more

TABLE 2.3: Posterior distributions of the linguistic features for Pokémon EVOLUTION. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability of the posterior being larger than 0, and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.35	(−1.05, 0.33)	0.16	0.84
Segments	0.44	(−0.08, 0.95)	**0.95**	0.05
Front vowels	0.44	(−0.24, 1.13)	0.90	0.10
Back vowels	0.50	(−0.65, 1.59)	0.81	0.19
High vowels	−0.30	(−0.71, 0.10)	0.07	0.93
Low vowels	0.31	(−0.35, 0.99)	0.82	0.18
Round vowels	0.07	(−0.89, 1.09)	0.55	0.45
Sonorants	−0.29	(−0.68, 0.09)	0.07	0.93
Labials	0.17	(−0.18, 0.51)	0.83	0.17
Voiced obstruents	0.21	(−0.15, 0.58)	0.87	0.13

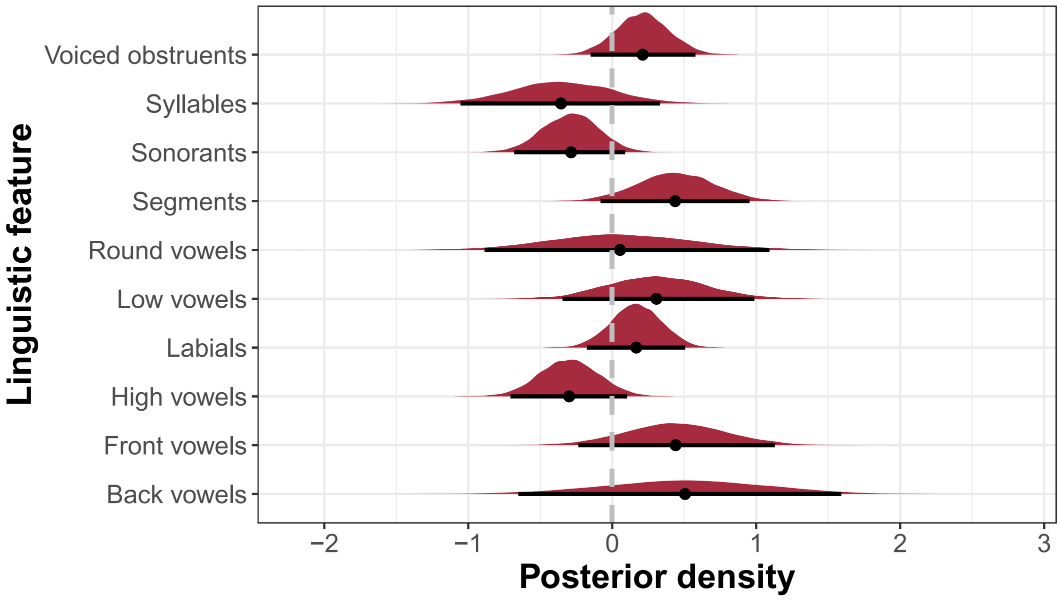


FIGURE 2.4: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute EVOLUTION. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.4: Posterior distributions of the linguistic features for Pokémon HEIGHT. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.07	(−0.42, 0.26)	0.33	0.67
Segments	0.12	(−0.15, 0.39)	0.81	0.19
Front vowels	0.02	(−0.33, 0.35)	0.51	0.49
Back vowels	0.16	(−0.35, 0.68)	0.73	0.27
High vowels	−0.16	(−0.36, 0.04)	0.06	0.94
Low vowels	0.00	(−0.31, 0.33)	0.51	0.49
Round vowels	−0.09	(−0.54, 0.35)	0.34	0.66
Sonorants	−0.08	(−0.27, 0.11)	0.20	0.80
Labials	−0.19	**(−0.36, −0.02)**	0.01	**0.99**
Voiced obstruents	0.01	(−0.17, 0.20)	0.54	0.46

labial consonants in a name has a negative impact on the height of a Pokémon. This effect was not identified in the correlation analysis. It was, however, present in the study by Shih et al. (2019) for English, another Germanic language. The remaining linguistic features do not exhibit an effect on Pokémon HEIGHT. Among those features is the number of high vowels, which showed substantial evidence for a negative correlation with HEIGHT in the Bayes factor analysis.

Weight The results of the model for Pokémon WEIGHT are summarized in Table 2.5 and Figure 2.6. Two of ten linguistic features built into the model present an effect on WEIGHT. The number of labial consonants shows a negative influence on WEIGHT ($\beta = -0.23[-0.39, -0.06]$), with a posterior probability $\beta < 0$ equal to 1.00. The 95% CrI does not include 0, and all posterior samples are below 0, given the data, the model, and the priors. This result is in line with the strong evidence found for a negative correlation between WEIGHT and the number of labial consonants shown above, as well as with the result by Shih et al. (2019) for English. The model estimates also suggest an existing effect of the number of high vowels on WEIGHT ($\beta = -0.23[-0.42, -0.04]$). The posterior probability $\beta < 0$ is 0.99, implying a negative effect. Thus, the more high vowels a Pokémon has in its name, the less it weighs. In the previous step of the analysis, the relationship between the number of high vowels and WEIGHT proved inconclusive, as no evidence could be obtained with the Bayes factor. The remaining linguistic features show no effect on WEIGHT (cf. Table 2.5).

Hidden Power (HP) The model outcome for Pokémon HP is shown in Table 2.6 and Figure 2.7. Firstly, the number of syllables has a positive effect on HP

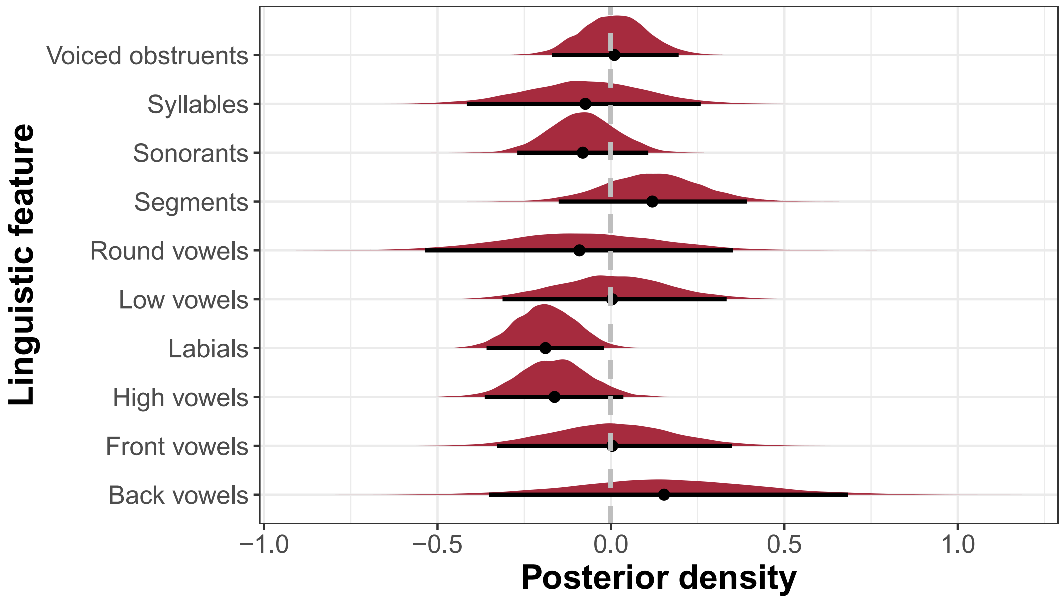


FIGURE 2.5: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute HEIGHT. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.5: Posterior distributions of the linguistic features for Pokémon WEIGHT. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.13	(−0.45, 0.20)	0.22	0.78
Segments	0.10	(−0.17, 0.37)	0.78	0.22
Front vowels	0.16	(−0.17, 0.49)	0.83	0.17
Back vowels	0.38	(−0.11, 0.87)	0.94	0.06
High vowels	−0.23	**(−0.42, −0.04)**	0.01	**0.99**
Low vowels	0.15	(−0.16, 0.46)	0.82	0.18
Round vowels	−0.19	(−0.63, 0.24)	0.19	0.81
Sonorants	−0.09	(−0.28, 0.10)	0.18	0.22
Labials	−0.23	**(−0.39, −0.06)**	0.00	**1.00**
Voiced obstruents	−0.07	(−0.24, 0.11)	0.23	0.77

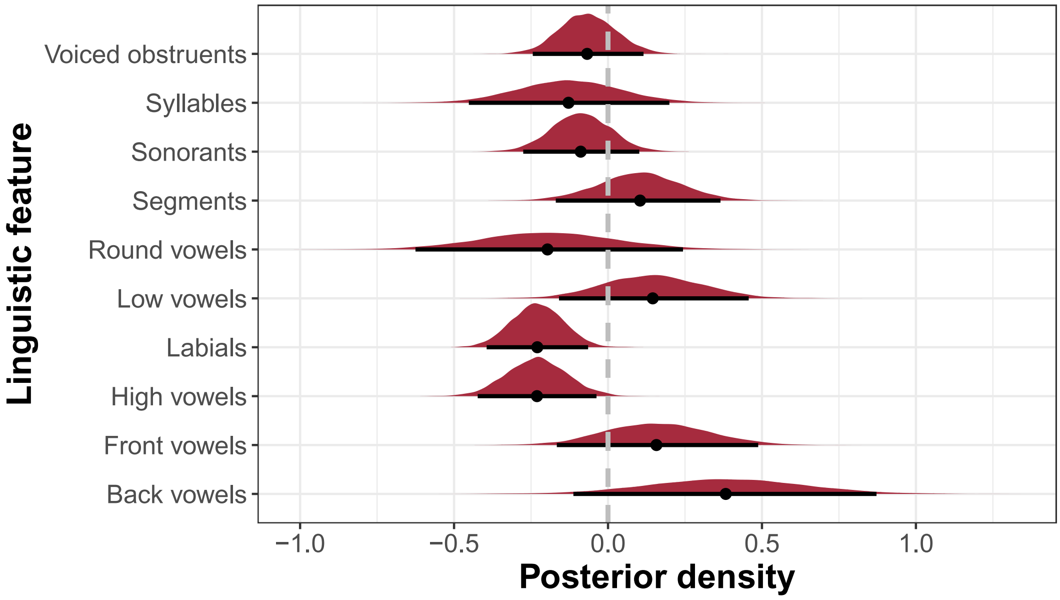


FIGURE 2.6: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute WEIGHT. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

($\beta = 0.41[0.09, 0.75]$), given the data, the model, and the priors, and yields a posterior probability $\beta > 0$ of 0.99. It suggests that the more syllables there are in a Pokémon name, the more HP it has. This result was not obtained using the Bayes factor analysis. On the contrary, BF_{10} provided substantial evidence for the lack of a correlation between the number of syllables and HP. It is, however, in line with the results for Mandarin Chinese obtained by Shih et al. (2019). For English, another Germanic language, Shih et al. (2019) found a positive effect of the number of segments on overall power. Also, the number of front vowels shows a negative impact on HP given the model, the data, and the priors ($\beta = -0.42[-0.75, -0.09]$), with a posterior probability $\beta < 0$ equal 0.99. The effect implies that the more front vowels there are in a Pokémon name, the smaller its HP. This result is novel given the first step of the analysis, which demonstrated substantial evidence for the lack of a correlation between the number of front vowels and HP. Similarly, the number of low vowels has a negative effect on the modeled data, while the Bayes factor analysis suggested a lack of effect. Given the model, the data, and the priors, a higher number of low vowels leads to lower HP ($\beta = -0.44[-0.75, -0.12]$). The posterior probability $\beta < 0$ for the effect amounts to 1.00. The remaining seven linguistic features, which were not mentioned, did not show an effect on Pokémon HP (cf. Table 2.6).

TABLE 2.6: Posterior distributions of the linguistic features for Pokémon HP. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	0.42	**(0.09, 0.75)**	**0.99**	0.01
Segments	0.00	(−0.28, 0.27)	0.51	0.49
Front vowels	−0.42	**(−0.75, −0.09)**	0.01	**0.99**
Back vowels	−0.05	(−0.54, 0.44)	0.42	0.58
High vowels	−0.05	(−0.24, 0.14)	0.31	0.69
Low vowels	−0.44	**(−0.75, −0.12)**	0.00	**1.00**
Round vowels	−0.28	(−0.71, 0.17)	0.10	0.90
Sonorants	0.03	(−0.16, 0.21)	0.61	0.39
Labials	−0.07	(−0.24, 0.09)	0.20	0.80
Voiced obstruents	−0.15	(−0.33, 0.04)	0.06	0.94

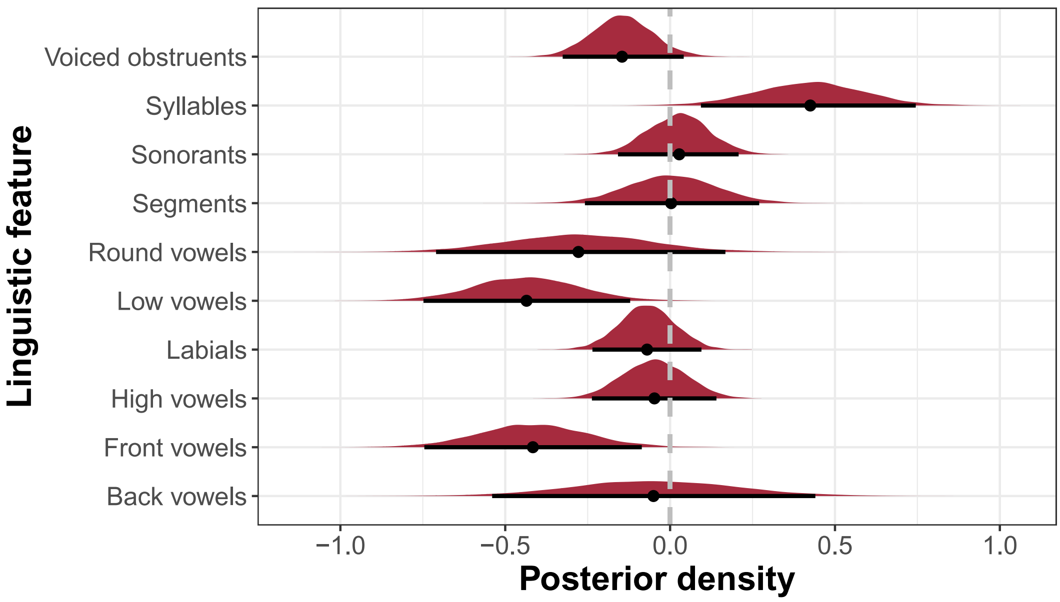


FIGURE 2.7: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute HP. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.7: Posterior distributions of the linguistic features for Pokémon ATTACK. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.25	(−0.60, 0.09)	0.08	0.92
Segments	0.10	(−0.15, 0.40)	0.80	0.20
Front vowels	0.26	(−0.08, 0.61)	0.93	0.07
Back vowels	0.41	(−0.11, 0.93)	0.94	0.06
High vowels	−0.07	(−0.27, 0.13)	0.24	0.76
Low vowels	0.24	(−0.08, 0.56)	0.93	0.07
Round vowels	−0.27	(−0.73, 0.19)	0.12	0.88
Sonorants	−0.06	(−0.25, 0.14)	0.29	0.71
Labials	−0.13	(−0.30, 0.04)	0.07	0.93
Voiced obstruents	0.00	(−0.17, 0.18)	0.49	0.51

Attack Table 2.7 and Figure 2.8 summarize the outcome of the model for Pokémon ATTACK. None of the linguistic features show an impact on this Pokémon attribute. This reflects the previous step of the analysis, in which no correlations with ATTACK were found. Kawahara et al. (2018), however, did find a relationship between both the number of voiced obstruents, as well as the number of moras and ATTACK in Japanese Pokémon names.

Defense The results of the model for Pokémon DEFENSE are given in Table 2.8 and Figure 2.9. The number of syllables shows a negative impact on DEFENSE, given the data, the model, and the priors ($\beta = -0.35[-0.67, -0.03]$), with a posterior probability $\beta < 0$ of 0.98. Therefore, the more syllables there are in a Pokémon name, the smaller its defense parameter is. The first step of the current analysis does not reflect this result, as it provided evidence for lack of correlation between the number of syllables and DEFENSE. Also, Kawahara et al., 2018 note a positive correlation for the number of moras and Pokémon defense in the Japanese data, which is contradictory to the current result. Further, the number of front vowels in the Pokémon name has a positive effect on DEFENSE, given the current data, the model, and the priors. The posterior probability $\beta > 0$ equals 0.95, which is sufficient to recognize the effect, however the credible interval spreads across the zero point ($\beta = 0.27[-0.05, 0.59]$). Hence, the effect is inconclusive. Besides, the Bayes factor analysis of the number of front vowels and DEFENSE yielded substantial evidence for the H0. The number of back vowels also shows a positive effect on DEFENSE in the current model ($\beta = 0.60[0.12, 1.07]$). Here, the posterior probability $\beta > 0$ is 0.99, and the credible interval does not span across 0. The previous analysis with the Bayes factor showed anecdotal evidence for the existing positive correlation between the number of back

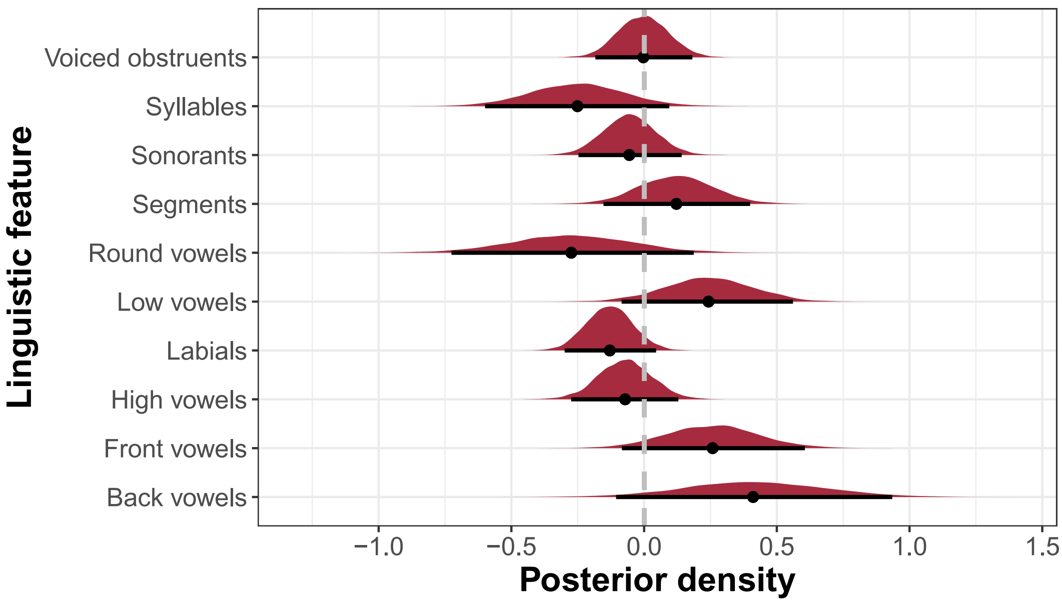


FIGURE 2.8: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute ATTACK. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

vowels and DEFENSE. Given the data, the model, and the priors, the number of high vowels impacts DEFENSE negatively with a posterior probability $\beta < 0$ of 0.99, and $\beta = -0.23[-0.43, -0.04]$. The effect suggests that the more high vowels are present in the Pokémon name, the lower its defense parameter. Nevertheless, the previous Bayes factor analysis provided anecdotal evidence for the lack of correlation between the two parameters in question. Lastly, the number of labial consonants shows a negative effect on DEFENSE given the data, the model, and the priors. The posterior probability $\beta > 0$ amounts to 0.98, and the 95% credible interval does not encompass zero $\beta = -0.17[-0.34, -0.00]$ (the exact point of the upper boundary is -0.001078456). This result, however, was not present in the Bayes factor analysis. Other linguistic features do not yield any effect on DEFENSE, given the current data, the model, and the priors.

Special attack Table 2.9 and Figure 2.10 summarize the results for SPECIAL ATTACK. Firstly, the number of high vowels show a negative effect on SPECIAL ATTACK ($\beta = -0.18[-0.37, 0.01]$). Although the posterior probability $\beta < 0$ reaches the value of 0.97, the effect is inconclusive due to the 95% credible interval spreading across zero from -0.37 to 0.01. The finding does not corroborate the results of the Bayes factor, which showed anecdotal evidence for the lack of correlation between the number of high vowels and SPECIAL ATTACK. Secondly, the number of round vowels yields a negative effect on SPECIAL ATTACK, given the data, the priors,

TABLE 2.8: Posterior distributions of the linguistic features for Pokémon DEFENSE. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.35	**(−0.67, −0.03)**	0.02	**0.98**
Segments	0.20	(−0.06, 0.47)	0.93	0.07
Front vowels	0.27	(−0.05, 0.59)	**0.95**	0.05
Back vowels	0.60	**(0.12, 1.07)**	**0.99**	0.01
High vowels	−0.23	**(−0.43, −0.04)**	0.01	**0.99**
Low vowels	0.22	(−0.09, 0.52)	0.92	0.08
Round vowels	−0.21	(−0.63, 0.21)	0.17	0.83
Sonorants	0.06	(−0.24, 0.12)	0.25	0.75
Labials	−0.17	**(−0.34, −0.00)**	0.02	**0.98**
Voiced obstruents	−0.15	(−0.32, 0.03)	0.06	0.94

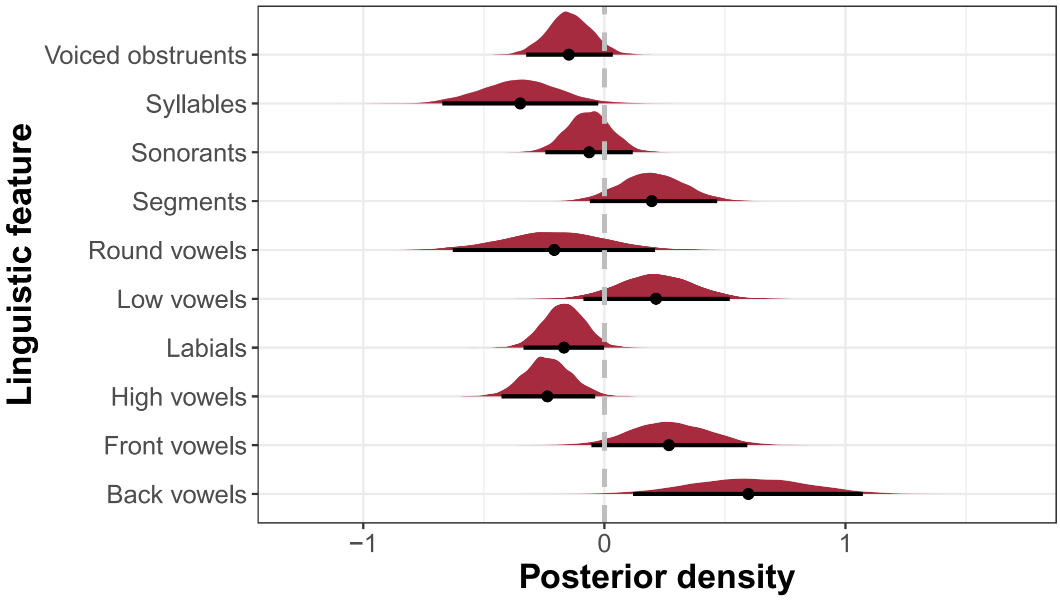


FIGURE 2.9: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute DEFENSE. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.9: Posterior distributions of the linguistic features for Pokémon SPECIAL ATTACK. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	0.23	$(-0.11, 0.56)$	0.91	0.09
Segments	-0.04	$(-0.31, 0.23)$	0.40	0.60
Front vowels	-0.01	$(-0.35, 0.32)$	0.48	0.52
Back vowels	0.37	$(-0.12, 0.86)$	0.93	0.07
High vowels	-0.18	$(-0.37, 0.01)$	0.03	**0.97**
Low vowels	-0.03	$(-0.34, 0.29)$	0.43	0.57
Round vowels	-0.41	$(-0.84, 0.02)$	0.03	**0.97**
Sonorants	0.08	$(-0.11, 0.27)$	0.79	0.21
Labials	0.14	$(-0.03, 0.31)$	**0.95**	0.05
Voiced obstruents	-0.01	$(-0.18, 0.17)$	0.47	0.53

and the model ($\beta = -0.41[-0.84, 0.02]$), with a posterior probability $\beta < 0 = 0.97$. However, here as well, the 95% credible interval encompasses zero spanning from -0.84 to 0.02. In the first step of the analysis, the Bayes factor indicated substantial evidence for the lack of a correlation between the number of round vowels and SPECIAL ATTACK. Lastly, given the data, the model, and the priors, the number of labial consonants has a positive effect on SPECIAL ATTACK ($\beta = 0.14[-0.03, 0.31]$). Even though the posterior probability $\beta > 0$ amounts to the minimum 0.95, the 95% CrI includes 0. Thus the effect found in the model is inconclusive. This effect was also inconclusive in the previous step of the analysis with the Bayes factor. The remaining six linguistic features do not exhibit an effect on SPECIAL ATTACK in the current step of the analysis (cf. Table 2.9). This includes the number of syllables, which showed anecdotal evidence for an existing correlation with SPECIAL ATTACK in the Bayes factor analysis.

Special defense Table 2.10 and Figure 2.11 summarize the results for SPECIAL DEFENSE. Based on the model, the data, and the priors, the number of high vowels yields a negative impact on SPECIAL DEFENSE ($\beta = -0.21[-0.40, -0.00]$; the exact point of the upper boundary is -0.004839660), with a posterior probability $\beta < 0$ $= 0.98$. This finding corroborates the result obtained with the Bayes factor and suggests that the more high vowels there are in a Pokémon name, the smaller its special defense parameter. Also, the number of voiced obstruents shows a negative effect on SPECIAL DEFENSE. It yields a posterior probability $\beta < 0$ of 0.97, but the 95% credible interval spans across zero with $\beta = -0.18[-0.36, 0.01]$. The remaining parameters do not show an effect on the parameter SPECIAL DEFENSE (cf. Table 2.10).

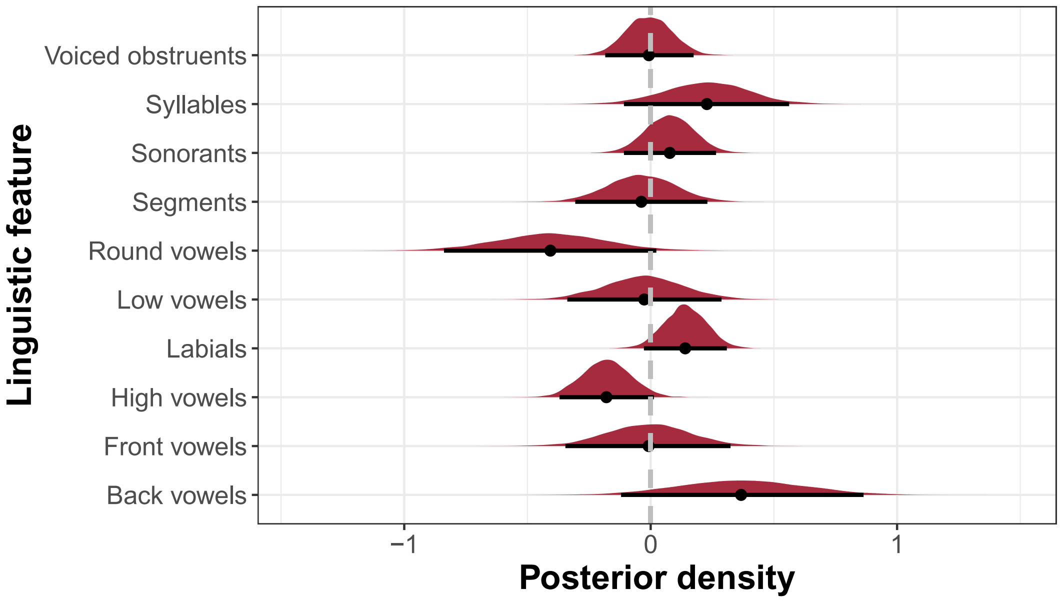


FIGURE 2.10: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute SPECIAL ATTACK. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.10: Posterior distributions of the linguistic features for Pokémon SPECIAL DEFENSE. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.19	(−0.53, 0.15)	0.13	0.87
Segments	0.20	(−0.06, 0.48)	0.93	0.07
Front vowels	0.14	(−0.20, 0.47)	0.79	0.21
Back vowels	−0.03	(−0.54, 0.48)	0.45	0.55
High vowels	−0.20	**(−0.40, −0.00)**	0.02	**0.98**
Low vowels	0.12	(−0.20, 0.43)	0.77	0.23
Round vowels	0.09	(−0.35, 0.54)	0.66	0.34
Sonorants	−0.04	(−0.23, 0.15)	0.33	0.67
Labials	−0.05	(−0.22, 0.12)	0.29	0.71
Voiced obstruents	−0.18	(−0.36, 0.01)	0.03	**0.97**

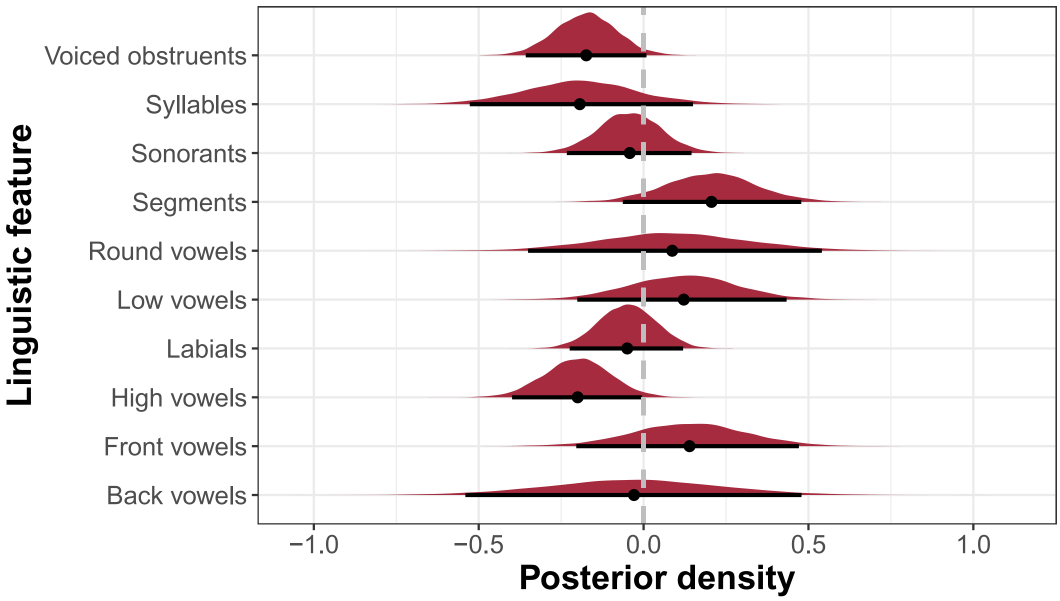


FIGURE 2.11: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute SPECIAL DEFENSE. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

Speed The model results for Pokémon SPEED are given in Table 2.11 and Figure 2.12. None of the linguistic features built into the model shows an impact on the Pokémon attribute SPEED. This result is similar to the analysis with the Bayes factor.

Gender Table 2.12 and Figure 2.13 show the results of the model for Pokémon GENDER, modeled as percentage of `Male`. This Pokémon attribute was not previously examined with the Bayes factor; thus, the results cannot be compared. This part of the analysis, however, reveals no noteworthy effects present for Pokémon GENDER based on the linguistic features implemented in the model.

2.4.2.3 Cumulative Results

The two-step analysis presented above provided some results which align with one another. Still, some of the effects were found only in one of the methods applied. The following paragraphs serve to disentangle the inconsistencies and reveal the informative effects.

Most notably, the strong evidence given for a negative correlation between WEIGHT and the number of labials ($BF_{10} = 11.145$) was reflected by the model of WEIGHT. The posterior probability $\beta < 0$ yielded 1.00 and the 95% CrI was reliably below 0 ($\beta = -0.23[-0.39, -0.06]$). Therefore, with increasing Pokémon weight, the

TABLE 2.11: Posterior distributions of the linguistic features for Pokémon SPEED. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	−0.16	(−0.50, 0.18)	0.18	0.82
Segments	0.16	(−0.12, 0.43)	0.87	0.13
Front vowels	0.11	(−0.24, 0.45)	0.74	0.26
Back vowels	0.22	(−0.29, 0.72)	0.81	0.19
High vowels	−0.03	(−0.23, 0.17)	0.38	0.62
Low vowels	0.14	(−0.19, 0.46)	0.80	0.20
Round vowels	−0.23	(−0.67, 0.23)	0.15	0.85
Sonorants	−0.07	(−0.26, 0.12)	0.23	0.77
Labials	−0.06	(−0.24, 0.11)	0.25	0.75
Voiced obstruents	0.03	(−0.16, 0.22)	0.63	0.37

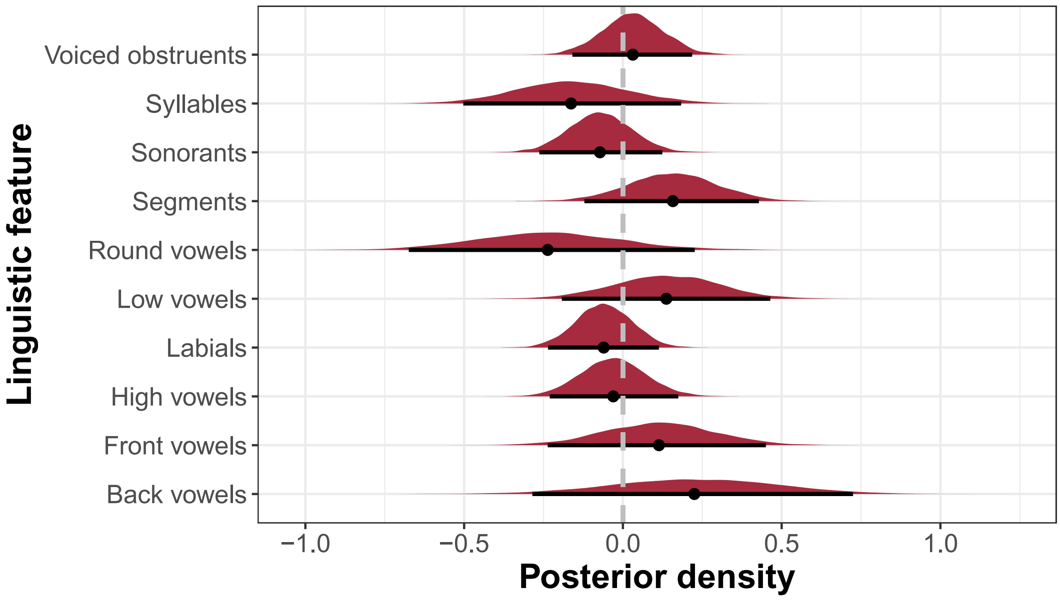


FIGURE 2.12: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute SPEED. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

TABLE 2.12: Posterior distributions of the linguistic features for Pokémon GENDER. The column Estimate shows the posterior means; the 95% credible intervals are given in brackets. The rightmost columns show the probability that the posterior is larger than 0 and smaller than 0. The probabilities written in bold are greater than or equal to 95%.

Parameter	**Estimate (CrI)**		$Pr(\beta > 0)$	$Pr(\beta < 0)$
Syllables	-0.05	$(-0.35, 0.25)$	0.38	0.62
Segments	-0.06	$(-0.30, 0.19)$	0.32	0.68
Front vowels	0.16	$(-0.13, 0.45)$	0.86	0.14
Back vowels	-0.23	$(-0.76, 0.30)$	0.20	0.80
High vowels	-0.02	$(-0.20, 0.15)$	0.39	0.61
Low vowels	0.23	$(-0.06, 0.51)$	0.94	0.06
Round vowels	0.38	$(-0.13, 0.89)$	0.92	0.08
Sonorants	-0.08	$(-0.26, 0.10)$	0.19	0.81
Labials	-0.04	$(-0.19, 0.11)$	0.31	0.69
Voiced obstruents	-0.05	$(-0.21, 0.12)$	0.29	0.71

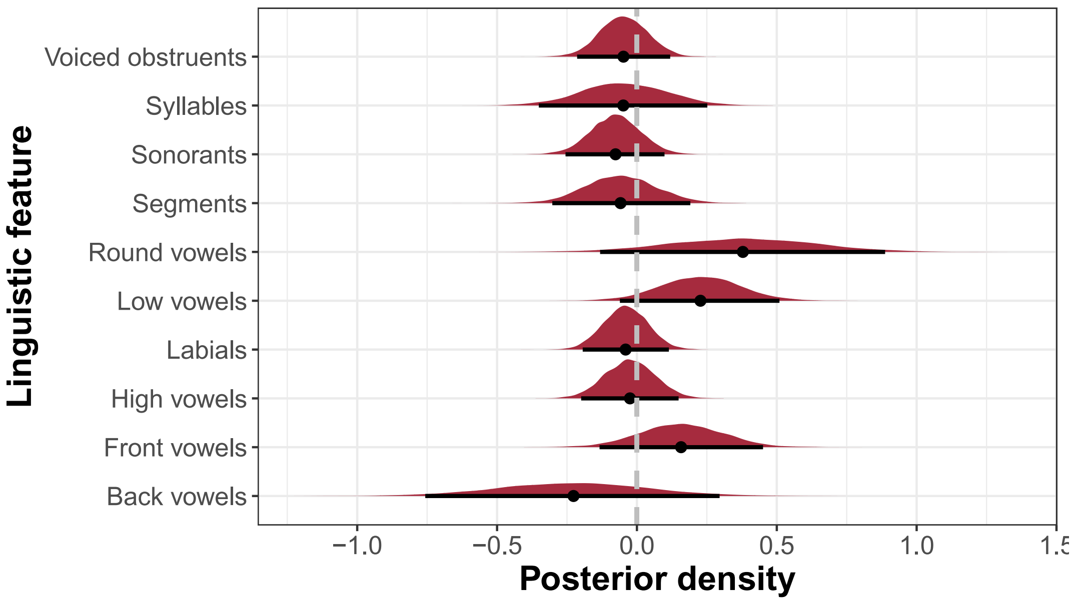


FIGURE 2.13: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the Pokémon attribute GENDER. Each of the linguistic features is given on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

number of labial consonants in a Pokémon name decreases. The Bayes factor analysis also provided anecdotal evidence for a positive correlation between DEFENSE and the number of back vowels ($BF_{10} = 1.713$). The results was corroborated by the Bayesian model indicating that the number of back vowels has a positive impact on DEFENSE ($\beta = 0.60[0.12, 1.07]$) with a posterior probability $\beta > 0$ equal 0.99. This suggests that if more back vowels are present in a Pokémon name, its defense parameter is higher. Furthermore, a negative correlation was identified between the attribute SPECIAL DEFENSE and the number of high vowels with anecdotal evidence ($BF_{10} = 1.581$). This correlation was mirrored by the results of the Bayesian model as well; the number of high vowels showed a negative impact on SPECIAL DEFENSE ($\beta = -0.20[-0.40, -0.00]$), with a posterior probability $\beta < 0 = 0.98$. These findings imply that a higher number of high vowels results in a lower special defense parameter of a Pokémon.

Further effects found during the first and the second step of the analysis show inconsistencies. The anecdotal evidence for a positive correlation between EVOLUTION and the number of segments was only reflected by the posterior probability $\beta > 0$ in the model, as it reached 0.95. The 95% credible interval, however, crossed the zero point ($\beta = 0.44[-0.08, 0.95]$). Also, while the Bayes factor analysis showed no evidence for an existing correlation between WEIGHT and the number of high vowels ($BF_{10} = 1.070$), the Bayesian model demonstrated a robust negative effect of the number of high vowels on Pokémon WEIGHT ($\beta = -0.23[-0.42, -0.04]$), with a posterior probability $\beta < 0 = 0.99$. Because both steps of the analysis do not align with one another, this effect remains inconclusive. The effect is also inconclusive for SPECIAL ATTACK and the number of labials. The parameters showed no evidence in the Bayes factor analysis ($BF_{10} = 1.059$). In the model for SPECIAL ATTACK, the posterior probability $\beta > 0$ for the number of labials was equal 0.95, but the 95% CrI spanned across zero ($\beta = 0.14[-0.03, 0.31]$).

Some correlations established with the Bayes factor were not reflected in the model outputs at all. Most notably, this includes substantial evidence for a negative correlation between HEIGHT and the number of high vowels. Apart from that, two correlations with anecdotal evidence could not be backed up by the model results, including the correlations between SPECIAL ATTACK and the number of syllables, and EVOLUTION and the number of voiced obstruents.

Lastly, various effects in the models, even if robust, were not identified with the Bayes factor. These effects are: HEIGHT and the number of labials, HP and the number of syllables, HP and the number of front vowels, HP and the number of low vowels, DEFENSE and the number of syllables, DEFENSE and the number of front vowels, DEFENSE and the number of high vowels, DEFENSE and the number of labials, SPECIAL ATTACK and the number of high vowels, SPECIAL ATTACK and the number of round vowels, and SPECIAL DEFENSE and the number of voiced obstruents. Given the two-step analysis, only the effects that showed consistent significance across both statistical methods are considered informative.

At this point, I would like to note that the two-step analysis performed for the data was chosen to minimize the risk of finding effects that are unlikely to hold true. Such a situation would be a false positive or a Type I error in a frequentist framework. In Bayesian interference, the null hypothesis significance testing does not take place, and, at least in a two-sample test, the Bayesian framework has been recently shown as more robust against Type I error (Kelter, 2020). However, some factors may influence the output of Bayesian tests, such as the model assumptions or the priors (Kelter, 2020). Acknowledging the probability of a misleading result in such a case, I decided on a very conservative procedure. Of course, it can be argued that this procedure invites another type of error, the so-called false negative, or a Type II error. Although I grant that this might be true, I still maintain that a researcher should remain cautious with overgeneralizing when dealing with a small sample of data, as is the case here.

2.5 Discussion

The current investigation examined the sound-symbolic tendencies in German Pokémon names in a two-step analysis. It revealed one positive relationship between the Pokémon defense parameter and the number of back vowels and two other negative relationships. One between special defense and the number of high vowels, and the other between weight and the number of labial consonants.

In German Pokémon names, the strongest relationship is present between Pokémon weight and the number of labial consonants. This negative correlation is in line with the results provided by Shih et al. (2019) for Japanese and English. As an explanation, Shih et al. (2019, 14) put forth the analysis by Kumagai and Kawahara (2017) in which the authors show that labial sounds are associated with babies in Japanese. Similarly to Japan, a lot of German diaper brands contain labial consonants: *Pampers*, *HiPP*, *babydream*, *babylove*, *Bella Baby Happy*, *Bevola Baby*, etc. However, *Huggies* is a popular diaper brand in Germany, too. Because in English-speaking countries, there also exist diaper brand names lacking labial consonants, Shih et al. (2019) propose that this is why the effect is weaker for English. This weakening, however, does not hold for the German data, as this is the most robust effect found in the current analysis. The fact that labial consonants are among the very first ones which babies can articulate (MacNeilage and Davis, 2000) may serve to strengthen the effect in German. Newborns start using their lips from day one to suck milk from their mother's breasts. Also, as established by Blasi et al. (2016), /m/ is among the sounds cross-linguistically correlated with breasts. It all suggests that, in language perception, labial segments may be connotated with baby-like features. Thus, weight, as one of the main landmarks of perceived size, is sound-symbolically represented by the number of labial consonants in German Pokémon names.

Two other effects established in the course of the analysis – a positive correlation between defense and the number of back vowels, and a negative correlation between special defense and the number of high vowels – are novel with respect to earlier research. They are, however, consistent with the "frequency code" by Ohala (1984, 1994), which suggests that low-frequency sounds – such as back or low vowels – evoke the sense of dominance. Both a higher defense and special defense parameters of a Pokémon testify about its endurance in a fight. As revealed in the analysis, in German, Pokémon that have a higher defense parameter are more likely to have more back vowels in their name, thus signaling dominance. The relationship between special defense and the number of high vowels is negative and, therefore, also goes in the direction predicted by the "frequency code" (Ohala, 1984, 1994). Here, a Pokémon with a higher special defense parameter is more likely to have fewer high vowels in the name, as numerous high vowels would instead give an impression of a small and defenseless Pokémon.

Among the robust effects established in the two-step analysis, none resonated with the three predictions expressed in Chapter 2.4.1. The predictions were based on the previous findings by Kawahara et al. (2018) and Shih et al. (2019) on Pokémon names from various languages. Those studies, however, used different statistical methods, i.e., frequentist statistics, and did not take the same amount of Pokémon attributes and linguistic features, respectively, into consideration. Notably, Kawahara et al. (2018) and Shih et al. (2019) analyzed Pokémon names across generations, while this study focused on one generation only. It is, therefore, possible that some effects are pronounced in a between-generation manner. The incongruence in the findings does not show that Pokémon names are not an optimal testbed for sound-symbolic analyses. On the contrary, it points to the variety of sound-symbolic relationships that languages adopt. Some of those relationships use different features interchangeably; this depends on the language structure, as was shown by Shih et al. (2019, 27) – where Japanese uses the number of moras, other languages employ the number of syllables or segments.

The relationship between Pokémon size (in general) and the number of moras, segments, or syllables, was by far the most robust one cross-linguistically in Shih et al.'s (2019) study, yet, it was not identified as reliable in the current analysis of German. This difference shows that – despite cross-linguistic tendencies – languages have freedom in their sound-symbolic expression. The number of moras, syllables, or segments serves as a signal of size in numerous languages in which Pokémon names were tested. For German Pokémon, the low number of labial consonants expresses a large size – in the form of higher weight. Then, a Pokémon that can defend itself better in a fight will most likely have more back vowels in a name, and even more so, if it also has fewer high vowels in a name, which implies that its special defense parameter would be higher. As in the example of German Pokémon names, this all shows the richness of the expression of size, dominance, and, possibly, danger with sound symbolism.

Each language has its phonological system and phonotactic rules. Those may govern that the sound symbolism on a given feature-attribute combination is strongly present in one language but lacking in another one. Unlike Japanese, German is known for its complex consonant clusters in both syllable onsets and codas (Hoole et al., 2012). Such differences may lead to divergent markedness of sound-symbolic patterns, even within the same universe, like Pokémon. Therefore, if a robustly sound-symbolically marked feature is not found in another language – whether of the same or different family – the language under investigation should not be banned as lacking sound symbolism. On the contrary, it exposes new gateways for sound symbolism to come to light. In a language with a simple syllable structure, in which complex consonant clusters would violate phonotactic rules, syllable structure as a feature may not be useful for sound symbolism. However, in languages such as German, where syllables in themselves have weight due to vowel length or the number of consonants in the coda, an approach that takes syllable structure into account may be more fruitful than a simple analysis of the number of syllables or segments. The comparative analysis of sound symbolism in languages may reveal, as Shih et al. (2019) say, the same dimensions in drastically different expressions. It should, nevertheless, be combined with a prior investigation on informative features in a given language – not across languages. Such investigation might lead to establishing some kind of a "sound-symbolic functional load," whereby certain features are exposed to be the the carrier of sound symbolism within a language. Further comparison of these features between languages would enable uncovering cross-linguistic sound-symbolic tendencies.

There are still some limitations to the current analysis. First and foremost, only a subset of Pokémon created to date was analyzed. There is 893 Pokémon from eight generations, while the database of this analysis encompassed only the first generation with its 151 Pokémon. The insufficient amount of data might have led to some effects to remain undetermined or unpronounced whatsoever. On the other hand, the addition of further Pokémon generations may also cover the effects found here. However, an analysis of a complete data set could unveil the inconsistencies between the Bayes factor and Bayesian linear regression results seen during the present study. Therefore, it would most likely provide more robust results than those seen above. Future experiments could also incorporate an analysis of Pokémon types, which were not examined for their sound symbolism to date.

Most importantly, a semantic analysis of the origin of Pokémon names could serve as a notable addition to any future experiment on their sound symbolism. German Pokémon names were translated from English, rather than from the Japanese original, and, as previous studies suggest, there are numerous strategies for translation of fictional names and sound symbolism (Fernandes, 2006; Jawad, 2010; Pischedda, 2020; Santoso and Setyaningsih, 2020). Disentangeling different translation strategies would allow for examining sound symbolism within the groups and establishing, which of those strategies retains sound symbolism to which extent. The current study shows that even if the studied sound-symbolic expression may seem "lost in translation", it may emerge in another form that is more suitable for the given language, whether these are more voiced obstruents or fewer high vowels that signify strength. Sound symbolism does have universal tendencies, but – cross-linguistically – also many forms in which it can come to light.

Summing up, this study revealed the sound-symbolic patterns in German Pokémon names. It exposed that, across languages, similar dimensions, such as dominance, can be expressed with vastly different means. I suggest that the sound-symbolic markedness is directly related to the phonological and phonotactic structure of a given language. Subsequent investigations should therefore consider the distinctive features within a language to study cross-linguistic sound symbolism. Also, working within a fictional character universe from a cross-linguistic perspective is subject to the risk of a very variable name translation. Therefore, taking translation strategies into account might additionally benefit future research on Pokémonastics, i.e., the sound symbolism of Pokémon names.

Chapter 3

Iconic Prosody

3.1 Introduction to Iconic Prosody

Speech consists of segments – vowels and consonants – and suprasegmental phenomena. The iconicity of the former was described in the previous chapter. Here, I will focus on the iconicity of the suprasegmental parts of speech and refer to it as iconic prosody, a term used in the past by, e.g., Perlman et al. (2015a). Prosody has been put on a par with intonation in the past, though different authors use these terms in different ways (cf. Hirst and Di Cristo, 1998, for a discussion). Here, prosody is used as an umbrella term for suprasegmental parts of speech, such as stress, intonation, quantity, and rhythm (Bussmann, 2006; Trask, 2004). Prosody can be regarded from the perspective of perception or production. While pitch is a perceptual or auditory correlate, fundamental frequency (also called F0) is an acoustic property of sound in speech production. Therefore, I will use pitch for speech perception context and fundamental frequency or F0 in the context of speech production.

Prosody is a mix of various suprasegmental elements, and due to its broad spectrum, it has vast possibilities for iconicity. One of the examples is the intonation of questions and statements. In cross-linguistic studies, it has been shown that questions are often expressed with rising or high intonation of one of the elements, in contrast to statement sentences, which correlate with falling or low intonation (Bolinger, 1978; Hermann, 1942; Ultan, 1969). This simple yet remarkably widespread phenomenon occurs in most of the examined languages and cultures, with only very few exceptions. In his study, Ultan (1969) found the tendency for rising contours in yes/no-questions in languages genetically as disparate as Japanese (Japonic), Russian (Indo-European), and Tagalog (Austronesian), among others. In speech perception, Gussenhoven and Chen (2000) found that native speakers of Standard Chinese (Sino-Tibetan), Dutch (Indo-European), and Hungarian (Ugro-Finnic) exhibit the same preference and recognize rising intonation at the sentence offset as question intonation. A linguistic or cultural similarity for these language groups can be ruled out, thus the underlying mechanism must be of a different sort. Ohala (1982; 1984) observed that in the animal kingdom, high-pitched vocalizations are associated with submission, compared to low-pitched vocalizations that are associated with dominance. He proposed that this innate force is responsible for the cross-linguistic question vs. statement intonation phenomenon (Ohala, 1994). Socially, as Ohala states, high intonation indicates "deference, politeness, submission, lack of confidence" and low intonation signals "assertiveness, authority, aggression, confidence, threat" (Ohala, 1994, 327). Thus, the rising intonation contour of questions iconically transmits a message of uncertainty, while the falling contour of statements iconically expresses certainty. This is further supported in Merin and Bartels's (1997) theory, which says that an intonational rise transfers the choice to the interlocutor, while an intonational fall allots it to the speaker (cf. Merin and Bartels, 1997, 7).

In this chapter, I will focus on the iconic prosody of two established cross-modal relationships. Firstly, on the relationship between vertical space and fundamental frequency. It has been shown in the past that high-pitched sounds are perceived as coming from higher vertical space, regardless of their actual source (e.g., Mudd, 1963; Pratt, 1930; Roffler and Butler, 1968). Similarly, in speech production, a higher fundamental frequency was demonstrated to be used when talking about "higher" concepts (Clark et al., 2014; Nygaard et al., 2009b). Secondly, I will investigate the relationship between size and voice frequencies. Past research shows that high-pitched vocalizations are perceived as coming from a smaller source (Ohala, 1994; Pisanski et al., 2014a), yet some studies suggest that the perception of size relies rather on formant frequencies than on the fundamental frequency (Pisanski et al., 2014b; Pisanski and Rendall, 2011). In vocal production, animals were shown to signal size using fundamental frequency (Bee et al., 2000).

Despite its cross-modal character, iconicity is often taken as both explanans and explanandum of an effect in question. So far, only very few studies have suggested the potential origin of iconic cross-modal relationships. It has been proposed that the origin lies either in the biological constraints ("frequency code" by Ohala, 1994), natural environment (Parise et al., 2014), or in the multi-sensory perception, associated with synesthesia (e.g., Marks, 1975; Ramachandran and Hubbard, 2001) or with sensorimotor processes (Perniss and Vigliocco, 2014). The current investigation of the two cross-modal relationships mentioned above follows the view that iconic prosody is rooted in biological sensorimotor processes. Specifically, I put forward a hypothesis that these relationships are grounded in physiological processes.

3.2 Motivation for the Empirical Investigation

In the following, the background and the motivation will be outlined, giving evidence to both vocal iconicity and other potential influences of frequency mechanisms constrained by the body. The latter point is crucial to illustrate the possible physiological grounding of prosodic iconicity.

3.2.1 Previous Work on Iconicity of Vocal Frequencies

Vertical Space

Imagine you are walking down the street, and someone shouts your name. Most likely, you would immediately turn in the right direction. But would you recognize if the person shouting your name was on a balcony or the ground level? Which cues lie behind the vertical dimension in sound perception? For over a century, researchers have been interested in the localization of sound in the vertical dimension and the cognitive processes behind it (Seashore, 1899). Studies show that there is an iconic relationship between vertical space and fundamental frequency (F0) – high elevation is connected to a high sound and low elevation to a low sound.

In a pioneering study, Pratt (1930) asked participants to localize sounds of different fundamental frequencies vertically. The sounds were played randomly at five equidistant height levels. Regardless of the actual position of the source, the participants vertically aligned the tones according to their fundamental frequency: 4096 Hz at the highest position, subsequently followed by 2048, 1024, and 512 Hz, and 256 Hz at the lowest position. The preference to vertically localize higher sounds in the higher plane and lower sounds in the lower plane is thus sometimes referred to as Pratt effect (cf. Pratt, 1930). And although the frequencies used by Pratt differ from one another by a whole octave on a musical scale (from the highest at approx. C8 to the lowest at approx. C4), the effect has been later successfully replicated in further experiments with less divergent frequencies (Mudd, 1963; Roffler and Butler, 1968; Trimble, 1934).

In his study, Trimble (1934) used a fixed sound source and controlled for the absolute intensity of the stimuli, which had the following frequencies: 500, 600, 700, 850, 1200, 1750, 1900, 2250, and 3950 Hz. All but one participant were consistent in their judgment and localized higher tones in higher positions. Trimble extended the experiment with a task on an apparent displacement of a phantom sound using ascending and descending frequencies. Participants were asked to draw the displacement. Consistent with previous findings, higher tones were localized higher in vertical space. Mudd (1963) implemented a similar method in his investigation. Participants listened to pairs of sounds differing in either F0, intensity, duration, or direction, and were asked to represent the localization of the first and the second sound on a pegpanel. The stimuli were presented over headphones. Among other results, fundamental frequency yielded a significant effect in both horizontal and vertical dimensions, the latter being far stronger. In their experiment, Roffler and Butler (1968) played the stimuli in nine different frequencies (250, 400, 600, 900, 1400, 2000, 3200, 4800, and 7200 Hz) from vertically varying sound sources. The participants were asked to localize the sound source on the vertical plane. As in previously mentioned studies, the pitch was matched to the vertical position such that higher sounds were localized above the lower sounds. The authors also showed that the effect persists for congenitally blind people and young children, who presumably did not know that concepts high and low may refer to sound.

Why this effect is so persistent may have to do with innate sound perception mechanisms. In a highly controlled and widely cited study, Blauert (1969) demonstrated that in the median plane (front to back), the localization of sound relies purely on spectral characteristics. The participants sat in a darkened room, hence lacking any visual cues. Recordings of participants' ear canals were made, and the noise was presented either from the back or the front of the participants. Also, noises mimicking the opposite sides were played; for example, a spectrally "back" sound was played from the front loudspeaker. The results show that even if played from a front speaker, spectrally "back" sounds were identified as coming from behind, and spectrally "front" sounds played from a back speaker as coming from the front. It can be concluded that human ears are fine-tuned to spectral differences between sound source locations rather than are to other signal properties.

Recently, Parise et al. (2014) investigated sound mapping in the vertical plane. The authors recorded natural environment sounds with two directional microphones that were mounted on the head of a human, who was moving freely through urban and rural areas. Both animate and inanimate sound sources were recorded, like birds or rustling trees. The data show a strong mapping sound frequency and elevation (frequency–elevation mapping or FEM for short). The revealed effect was the strongest in the middle spectrum range between 1–6 kHz. Thus, in the natural environment, high-frequency sounds tend to come from higher locations, compared to the low-frequency sounds that originate from lower positions. Apart from that, the authors looked at which impact do the filtering properties of the ear have on the effect. They found that the outer ear enhances elevated sounds – as a result, these have more energy at high frequencies. They furthermore controlled for the proximity of the sound source by separating proximal and distal stimuli. The data reveal that FEM is given in both near and far sound sources and conclude: "This might suggest that human spatial hearing is so finely tuned to the environment that even the filtering properties of the outer ear, and hence its convoluted anatomy, evolved to mirror the statistics of natural auditory scenes" (Parise et al., 2014, 6105). Therefore, the statistical probability in the auditive environment behind the FEM may have grounded listeners' expectations towards sound source localization in the vertical space. But how is this relevant to human speech? In their results, Parise et al. (2014) highlighted the strongest effect within the spectral range of 1–6 kHz. This range is only partly relevant with regard to speech, which spans from 0–10 kHz and occurs mostly in low-frequency ranges: "[a]lmost half of the auditory frequency scale covers frequencies below 1,500 Hz" (Johnson, 2011, 55). It is also known that due to the structure of the basilar membrane located in the inner ear, the human ear is more prone to detect small changes occurring below 1 kHz, rather than on other frequency bands (Johnson, 2011, 52f.).

Such small changes have been found in past studies investigating the cross-modal correspondence between vertical space and F0 in speech production. In a series of experiments, Shintel et al. (2006) presented participants with animations of a dot moving up or down (Experiment 1) and left or right (Experiments 2 and 3). In Experiment 1, the participants were asked to describe the direction by saying "It is going up" or, in a control condition, to say nonce words embedded in a carrier sentence "Please say the word ___". The analysis revealed a significant effect of direction, with a mean F0 difference of 6.7 Hz. Overall, the results suggest that speakers communicate information using congruent prosodic means. With regard to the vertical plane, speakers use higher F0 when talking about objects moving upwards and lower F0 when talking about objects moving downwards.

More recently, Clark et al. (2014) conducted an experiment, in which they asked participants to read stories each containing metaphors on either positive or negative emotions (Bolinger, 1986, after Clark et al., 2014, 5), words on high or low vertical space, and high or low sounds. They expected higher vs. lower fundamental frequency to be used in up vs. down stories, respectively. Apart from discriminating between the three concepts, the analysis differentiated between phrases that were shared and contrasting in the stories. Overall, the analysis revealed no effect in stories on emotions and sounds. However, the main effect of semantic valence (up vs. down stories) was found for stories about vertical space. As the authors did not find a correlation effect of semantic valence with contrastiveness, they concluded that the difference in fundamental frequency between up and down stories was spread across the whole recording. The mean F0 difference between the up and down stories about vertical space was equal to 5.4 Hz.

In another study, Nygaard et al. (2009b) compared the F0 of novel words that represented adjectives from the following adjective pairs: happy/sad, hot/cold, big/small, yummy/yucky, tall/short, and strong/weak. The novel words were assigned randomly to represent each meaning and were counterbalanced across the dimensions so that each participant used the same novel word for both adjectives of the pair. The novel word was embedded in a carrier sentence "Can you get the (novel word) one?" and intended meanings were presented as pictures depicting the concept. In addition, the speakers were asked to use infant-directed speech and were reported as having experience with young children. The analyses revealed that speakers employ acoustic parameters to convey the meaning of the individual semantic domain (i.e., positive and negative adjectives). Additionally, in follow-up perception experiments, the authors found that listeners, when confronted with picture pairs matching the auditory stimulus (e.g., "big" and "small" pictures for "big" recording), were able to confer the intended meaning based on the prosody. When confronted with a mismatching picture pair (e.g., "big" and "small" pictures for "happy" recording), listeners performed significantly worse than when pictures were matching, but the statistical evaluation also indicates a main effect of prosody. Here, mean F0 differences in the production of novel words spanned from

25 (strong/weak) to 142 Hz (happy/sad), yet, it has to be noted that the difference is enhanced due to infant-directed speech.

Two conclusions should be made at this point. Firstly, in speech perception, listeners associate high-frequency sounds with elevated positions and low-frequency sounds with lower positions. The effect is bi-directional, as it also occurs in speech production – when talking about literally or figuratively elevated objects, speakers use higher fundamental frequency than when referring to lower objects. In perception experiments, the differences between high and low objects tend to be expressed with large hertz differences, while in production experiments, the effects expressed in hertz are much smaller (see Clark et al. 2014, but cf. Nygaard et al. 2009b). Secondly, one possible source of this cognitively established cross-modal correspondence may be natural environment auditory statistics, as in the world, higher frequencies tend to come from elevated sources, compared to lower frequencies that come from lower sources (Parise et al., 2014).

Size of an Entity

Imagine now you are walking in the park and see a stray dog barking at you. How would you feel if the dog was large? And how would it feel if it was small? Both sensations would most likely evoke a different emotional response in yourself, but why is that?

Larger organisms can produce lower sounds than small organisms because vocal fold length is physically correlated with both body size and F0 (Riede and Brown, 2013). Longer vocal folds have more mass, thus they vibrate slower (Gick et al., 2013, 86). Nevertheless, many animals can produce sounds of both lower and higher F0, like male frogs that lower the F0 to defend their territories (cf. Bee et al., 2000). As found by Riede and Brown (2013), two factors served as a mechanism to uncouple the linear relationship between body size and F0: the tension of the vocal folds – as less tense vocal folds become shorter and heavier (Gick et al., 2013, 88) – and properties of one of the vocal folds' tissue layer's, *lamina propria*. The uncoupling most likely enabled animals to communicate with a broader spectrum of sound frequencies, thus making fundamental frequency a meaning-bearing unit. "Simply stated, birds and mammals use harsh, relatively low-frequency sounds when hostile and higher frequency, more pure tonelike sounds when frightened, appeasing, or approaching in a friendly manner. Thus, there appears to be a general relationship between the physical structures of sounds and the motivation underlying their use" (Morton, 1977, 855). In his article, Morton (1977) summarized numerous aggressive and nonaggressive animal vocalizations and concluded that the former are characterized by a harsh low-frequency sound and the latter by a pure high-frequency sound. He suggested the evolutionary significance of this relationship to lie in the ancient correlation of body size and F0. Larger animals are dominant within and between a species. Morton suggests that growing body size requires more energy, hence is less effective. As a result of the economy, evolving in the communication domain is favored over excessive size growth. On the other hand, high tones are helpful in establishing a parental role. An offspring is always physically smaller than a parent and, as a result, naturally produces high-frequency vocalizations. Both sides – an offspring and a parent – attract one another to establish a relationship using vocalizations that are designed in such a way not to induce fear in the offspring (Morton, 1977, 864ff.).

The relationship between body size and fundamental frequency of vocalizations was pursued by Ohala (1980; 1982), who suggested that the articulatory gesture when producing high-frequency vocalizations – retracting the corners of the mouth – is a potential origin of a smile. On the other side of the spectrum, constriction or protrusion of the lips suggests aggression or dominance.[1] In acoustics, the longer the vocal tract, the lower the F1 (first formant frequency, i.e., the lowest amplified frequency due to the resonance of the vocal tract), and the effect of lip protrusion on the vocal tract length has been shown in the past (Riordan, 1977). In subsequent work, Ohala (1984; 1994) explores the idea further and compiles various tendencies in prosodic behavior to justify the relationship that he refers to as "frequency code". Ohala points out regularity in the use of F0 across different domains. He departs from the most likely universal tendencies of raising F0 in questions and falling F0 in statements (as mentioned in Chapter 3.1). Ohala shows evidence for the affective use of F0 to show dominance – in his experiment, listeners judged recordings with lower F0 as more dominant or self-confident (Ohala, 1984, 3; replicated by Fitch, 1994 with synthesized speech). He then indicates the sound symbolic use of tone in tone languages, like Ewe, where high tones represent the notion of "small" and low tones the notion of "large". Moreover, he mentions the sound symbolic use of vowels and consonants: high front vowels, with relatively higher intrinsic F0, represent the notion of "small" and low back vowels the notion of "large" (like I discussed in Chapter 2.1).

Ohala (1984) also discusses the correlation between facial expressions and F0 and points out primates that retract their mouth corners when producing high-frequency vocalizations, compared to protruded lips when producing low-frequency vocalizations. Relying on Morton's (1977) analyses, Ohala hypothesizes that animals use different frequencies in vocalizations to appear larger or smaller, which is possible when the visual domain is lacking. Both body mass and size of an opponent are crucial to estimate a potential threat, which in the animal kingdom is a matter of life and death. The lower the pitch and the harsher the vocalization of a given animal, the more threatening, dominant, or aggressive the animal is assumed to be. And conversely, the higher the pitch and the purer the vocalization emitted by the sound source, the smaller its size is estimated to be, thus the source itself is interpreted as less threatening, dominant, and aggressive. Ohala (1984) calls this the "frequency code".

[1] One of the famous figures who is known to frequently use protruded lips is the former US President Donald Trump.

Finally, Ohala suggests that the "frequency code" is innate for humans and possibly other species (Ohala, 1984, 10). Due to sexual dimorphism, there are anatomical differences in male and female larynx. A larynx of an adult male is approx. 50% larger in the anterior-posterior dimension than a larynx of an adult female (Negus 1949, after Ohala 1984, 10). This results in males having longer, more massive vocal folds than females. Apart from that, the male larynx is about 15–20% longer than the female larynx. This all contributes to males having generally lower F0. As found by Fitch and Giedd (1999, 1519), the most noticeable changes contributing to sexual dimorphism in humans occur around puberty – a period crucial to establishing sexual dimorphism.

Past research has shown that not only perceived pitch influences body size estimation in human listeners, but also formant frequencies have a significant impact. In an experiment with synthesized speech, Fitch (1994, 28ff.) found F0 to be negatively – lower pitch implied a larger speaker – and formant dispersion to be positively correlated – longer vocal tract implied a larger speaker – with body size rating. Based on the results, Fitch hypothesizes that formant perception in humans evolved to enable body size estimation. Vocal tract length is positively correlated with body size (Fitch and Giedd, 1999, 1518). Given that formant frequencies are tightly coupled with vocal tract length, their relationship with body size estimation is evident. This finding has been corroborated in further studies, which found that formant cues override F0 cues in body size estimation (Pisanski et al., 2014b; Pisanski and Rendall, 2011), as well as in studies with other species (e.g., rhesus macaques: Fitch, 1997). More importantly, body size estimation based on auditory cues does not seem to rely on visual cues, as was demonstrated in an experiment with blind individuals (Pisanski et al., 2017a, where body size was measured as body height). Not only were the accuracy scores indifferent between early blind, late blind, and sighted listeners, but also participant groups made similar errors. Apart from the general and possibly innate (Ohala, 1994) relationship between body size and F0, Pisanski et al. (2017a) suggest that blind individuals learn to judge body size (here: height) using vocal cues by estimating sound source elevation.

The review thus far suggests: (1) elevated sources to be correlated with high-frequency sounds (e.g., Parise et al., 2014) and (2) large body size (height and/or weight) to be correlated with low-frequency sounds (e.g., Ohala, 1994). In other species, increasing body size is not as pronounced vertically as in humans – who, due to the erect posture, vary substantially in height. Despite this, it has been shown that listeners associate low positions with a large body size of both males and females when listening to low-frequency stimuli, but not when hearing high-frequency stimuli (Pisanski et al., 2017b, 1243ff.). It has been further shown that head elevation has an effect on body size perception via F0. Both while sitting and standing, participants correctly estimated low-frequency sounds played from low spatial positions to be coming from a larger source, yet only in a standing condition listeners rated male voices to be "larger" than female voices (Pisanski et al., 2017b, 1246f.). Overall, Pisanski et al. (2017b) show that, in perception, auditory cues override spatial cues as large body size was estimated for low-frequency sounds, but not high-frequency sounds in a low spatial position. Compared with the key role that elevation played in Parise et al.'s (2014) frequency-elevation mapping (of both animate and inanimate sources), Pisanski et al.'s (2017) results indicate that it is essential to consider animacy's effect in further experiments.

3.2.2 Previous Work on the Influence of Physiology on Vocal Behavior

The evidence above shows how vocal behavior iconically indicates real-world conditions – either the vertical position of an entity or its size. Because F0 patterns are strongly determined by the morphology of speech organs, below, I will focus on the physiological influence on vocal behavior.

In general, just as speech production itself, F0 control is anatomically complex. The prerequisite to both life and speech is respiration, with the former being its primary purpose and the latter an overlaid function (Perkins and Kent, 1986). Although respiration is very much like a reflex (Anokhin, 1974, 238), it also includes voluntary and subconscious but learned behavior (cf. Shea, 1996, 2). Yet, speech breathing is different from quiet breathing, required to sustain life functions by exchanging gases. For speech, air must be under pressure, for example, in order to make the vocal folds vibrate at a certain rate. The speech begins with air being directed out of the lungs with a set of intercostal and abdominal muscles (Gick et al., 2013, 61). With the additional muscular effort required for speech breathing, the subglottal air pressure builds up below the vocal folds in the trachea. Given that we want to produce a voiced sound, such as a vowel, the vocal folds form a kind of a dam resisting the pressure. Here, Bernoulli's Principle comes into play. It states that a decrease in its velocity must accompany a rise in pressure of a flowing fluid or gas, and conversely, a decrease in pressure must be accompanied by a rise in velocity. While the pressure above the vocal folds stays the same, the pressure below the vocal folds builds up, thus, at some point, becoming much greater. This causes the transglottal pressure difference to grow. Just like a dam cannot hold the water after a flood that is exerting too much pressure on it, the vocal folds must at some point give way to release the subglottal pressure. According to Bernoulli's Principle, in a narrow area, such as the glottis, the fluid or gas is susceptible to less pressure but gains speed. Therefore, the expiratory air stream flows through the glottis with much greater velocity and creates a perpendicular suction causing a movement of the vocal folds. The glottis opens due to raising air pressure, then the vocal folds close again due to the Bernoulli Principle, and the cycle repeats. In turn, the vocal folds vibrate at a certain frequency that is determined by the activity of both intrinsic and extrinsic laryngeal muscles (Erickson et al., 1982).

The activity of the muscles in and around the larynx has the largest impact on vocal frequency. The internal muscles of the larynx are shown in Figure 3.1. The

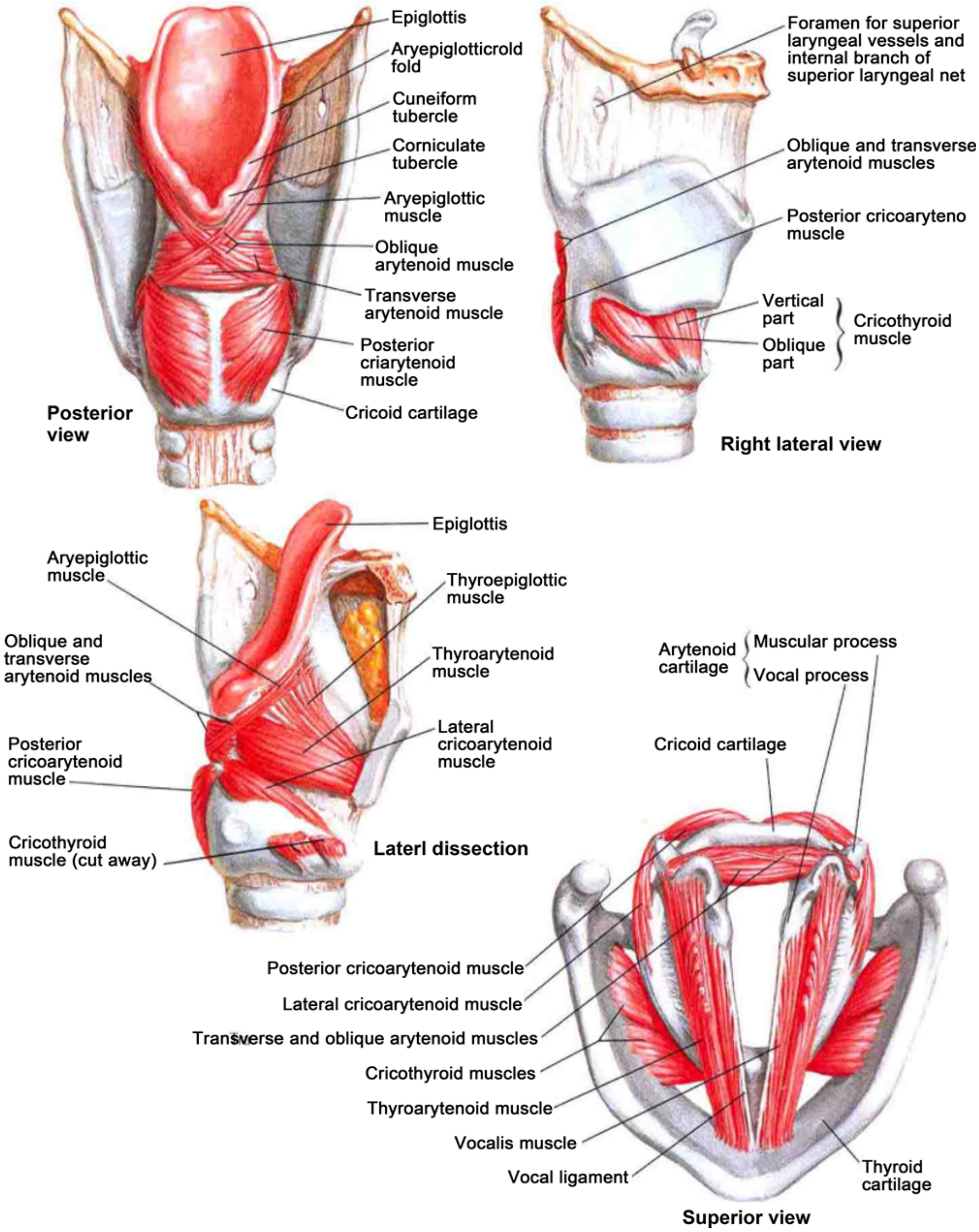


FIGURE 3.1: Internal muscles of the larynx. The image is taken from Arrangoiz et al. 2018, 172, available under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/).

main intralaryngeal muscle to control the F0 is the cricothyroid (CT) muscle (Gick et al., 2013, 77). It is attached to the cricoid and the thyroid cartilage, and, when activated, it pulls on the latter forward and downward. This movement does to the vocal fold what tensing a guitar string does – same as a tense guitar string produces higher frequency sound, so do tense vocal folds. Additionally, the thyroarytenoid (TA) muscles along the vocalis muscles (vocal folds) can be activated (Gick et al. 2013, 87f.; Honda 1995). Also, the lateral cricoarytenoid (LCA) muscle has been shown to exhibit a high correlation with F0, as it adjusts the effective length of the vocal fold vibration (Honda, 1988). The posterior cricoarytenoid (PCA) muscle – responsible for opening the glottis and the activation of which leads to voicelessness – may also play a role for F0 control by pulling the arytenoid cartilage backward (Honda, 1995, 221). From the external muscles, the cricopharyngeus (CP) was shown to have an inverse correlation with F0 (Honda, 1988). Most importantly, Erickson et al. (1982) showed that strap muscles act together with the internal CT muscle to lower F0. The role of the strap muscles in F0 control has been documented in further experiments (Honda, 1996; Hwan Hong et al., 1997). As their primary function is to change the position of the larynx itself, it is another step to ask, what influence may external movement, such as head rotation, have on F0?

Using a sentence-by-sentence multiple regression analysis, Munhall et al. (2004) found that 63% of the variation in the fundamental frequency could be explained by the speaker's head movement during speech production. In their data, the head movement followed F0, thus head elevation correlated with F0 rise, and conversely, head lowering with F0 fall. The authors discuss their results in the light of visual prosody, concluding that it contributes to the utterance in a way that the acoustic prosody does (Munhall et al., 2004, 136). However, given the influence of the muscular activity around the larynx on F0, their results imply that head movement alone can affect F0.

When a wolf howls, it raises the head upwards. It is, on the one hand, for sure to make the sound spread further, but on the other hand, the movement may allow the wolf to achieve a higher vocal frequency. Similarly in humans, sometimes head position seems to be used by popular artists to reach especially high notes and hold them[2] (also noted for untrained singers by Erickson et al., 1982, 276). Interestingly, the effect holds in perception as well. Chen and Massaro (2008) found that the visible information of the neck and head movements improves the ability of Mandarin speakers to recognize the lexical tone.

The fundamental frequency is yet only one part of the coin. As suggested above, the size of an entity has frequently been said to correlate rather with formant frequencies. The source–filter theory of speech production represents the process of

[2] A few examples are: Beyonce performing Listen, live at Oprah (`https://youtu.be/pai1C2dsd3M`, 2:05–2:20, 2:44–3:15), Demi Lovato's performance of Stone Cold, live at Billboard's Women In Music 2015 `https://youtu.be/B5qULV6x_cE`, 1:05–1:12, 1:48–1:52, 2:18–2:24, 2:38–2:46), or Tiffany Mosley's Total Praise, where she reaches notes as high as C#6 (`https://youtu.be/UW7tfjP2Eho`, 1:17–1:23, 1:58–2:01, 2:55–3:03, 4:45–5:10). All videos were last accessed on April, 6, 2021.

speech production "as a filtering process" (Fant, 1970, 16). There, the vocal folds, which are responsible for F0, are a part of the sound source. Supra-glottal elements, such as pharyngeal, oral, and nasal cavities, are the filter. Each of those cavities can be constricted in its own way and at very precise positions. Constrictions form "blockages" on the way out of the source sound and filter it. This filtering process results in changing the sound spectrum, i.e., the amount of vibration at a given frequency of the spectrum. The spectral maxima at certain frequencies of the spectrum are called formants (Fant, 1970, 20). The first two formants give rise to vowel differences, so that for German speakers /i/ has F1 of ~300 Hz and F2 of ~2200 Hz, and /a/ F1 of ~760 Hz and F2 of ~1400 Hz (cf. Figure 4 in Mooshammer and Geng, 2008, 129). Articulatorily, the first formant is correlated with tongue height, ergo the degree of jaw opening – the more open the jaw is, the higher the first formant. The second formant is correlated with the horizontal tongue position – the more frontal the tongue constriction, the higher the second formant. Therefore, the jaw opening and tongue constriction can be identified as physiological correlates of formant frequencies, respectively.

The research described in the previous chapter suggests that body size is best estimated with formant frequencies (in humans: Pisanski et al. 2014b; Pisanski and Rendall 2011; and in rhesus macaques: Fitch 1997). Thus far, to the best of my knowledge, no experiments exist that test this effect in vocal production for inanimate objects. In vocal iconicity, as described in Chapter 2.1, previous accounts reveal participants' preference for large objects to contain an open vowel, such as /a/, opposed to small objects to contain a closed vowel, such as /i/ (Sapir, 1929). The physiological ground may lie in the degree of jaw opening. The vowel /a/ has the largest degree of jaw opening, as opposed to the vowel /i/, where the jaw is almost closed. Within one vowel segment, due to coarticulation, there exists a tolerance for formants (cf. e.g., Öhman, 1966; Strange and Bohn, 1998). Given a preference for open vowels for larger objects, the degree of freedom within the formants of a single vowel has the potential to be used as an iconic signal of the size of a referred-to object. Physiologically this means that the degree of jaw opening would be correlated with object size.

3.3 Theoretical Hypotheses

In the previous sections, we have explored the background and the motivation for further research on iconic prosody. There is one further possible explanation for iconic prosody – synesthesia – yet this idea has not been tested in speech production. For this reason, the current study aims to systematically test for another plausible explanation for iconic prosody. Ohala's (1994) hypothesis that the frequency code is an innate mechanism, present in all humans, has physiological roots – the larger the body of the vocalizer, the lower their fundamental frequency. In the following, I will also motivate how both cross-modal correspondences described above may have physiological roots. Ultimately, the goal is to test the physiological parameters in a setting that evokes iconic speech production. This will allow us to see whether the physiological parameters influence what has so far been both explained by and called iconicity.

The first relationship illustrated above is the one between vertical space and fundamental frequency. It has been shown that a high sound is perceived as coming from an elevated source, as well as when describing an upward movement, higher F0 is used. Hand and facial expressions are known for their communicative character, but head gestures have also been recognized to play a role in face-to-face interaction (Morency et al., 2005). Gaze is a crucial cue given to infants, which they can recognize as early as at six months of age (Morales et al., 1998). When referring to an object, we frequently turn the gaze towards it and point at it with the head (Biguer et al., 1982), the movement of which has been shown to correlate with hand movement (Pelz et al., 2001). Hence, when turning towards an object placed higher, external laryngeal muscles become activated as well. As the existing evidence shows, the muscular activity of external laryngeal muscles does affect F0 (e.g., Erickson et al., 1982; Honda, 1995); thus, a change in F0 could be induced physiologically by a vertical head movement. The open question is, how much does head movement influence fundamental frequency. It has to be noted that vertical head movement here means an upward or downward rotation of the head on the vertical axis. The same is implied further with head position.

The other cross-modal correspondence reviewed above concerns the correlation between the perceived size of the vocalizer and formant frequencies. Here, to the best of my knowledge, production studies are lacking; therefore, I will rely on the evidence provided in the perceptual studies. So far, it has been shown that the perceived body size correlates acoustically with formants (formant dispersion: Fitch 1997; Fitch 1994, and formant frequencies: Pisanski et al. 2014b; Pisanski and Rendall 2011).

The following hypotheses have been derived from the reviewed literature, combining the evidence found thus far for iconic prosody with anatomical knowledge about sound production in the speech apparatus. They test the idea that iconicity is rooted in the human body: it is predicted that changes in muscle activation, e.g., produced by turning the head upwards to look at an object above, give rise to acoustic changes in the voice that reflect cross-modal iconic relationships. The hypotheses are the following:

- H1 Fundamental frequency is influenced by head position – the more upward the head is rotated, the higher the fundamental frequency.
- H2 The degree of jaw opening is influenced by the size of an object being named – the larger the object, the larger the degree of jaw opening.

H1 tackles the cross-modal iconic relationship between vertical position and fundamental frequency for which no underlying mechanism has been found so far. It aims to investigate whether the missing link between the vertical position of an object and F0 is the result of an upwards-looking head position. We expect to find a change in F0 due to head movement, because moving the head upwards engages the muscles around the larynx. These muscles are also responsible for F0 change, therefore, a change in head position should cause a change in the F0 and thus drive the cross-modal iconic effect in question.

H2 deals with the relationship between formant frequencies (especially F1 as suggested by Pisanski et al., 2017b) and object size estimation. So far, this relationship has only been shown in speech perception, with a higher F1 perceived as coming from a larger source. Here, it is tested from the speech production perspective. In speech production, F1 is highly correlated with jaw opening. Thus, jaw opening is a physical factor that affects F1 apart from vertical tongue position. As here it is the goal to assess the nature of the anatomical grounds for iconicity, we use jaw opening as our primary measurement.

3.4 Methodology

Experimental Design

A new paradigm was developed to test what role does the anatomy of the vertical head movement and jaw opening play for iconicity. The task consisted of a game scenario: participants were asked to "shoot" cans that were projected onto the wall in front of them. To shoot a can, the participants were asked to use a laser pointer to point at the can and to say the word written on the can. To avoid learning effects, one of the two different words could appear on a can, either *piff* [pɪf] or *paff* [paf], which are German onomatopoeic words for shooting, similar to English 'bang'. To measure the effect of size on F1, i.e., the jaw opening, small- and large-sized cans were included in the experiment, with the large can being two times as big as the small can (cf. Figure 3.2). The cans were positioned in five equal equidistant vertical and horizontal levels and projected across an area of 1.30 x 1.30 m (low boundary at 1.30 m). In order to elicit different vertical head positions for the different vertical can positions, participants were positioned 1 m away from it. This distance was chosen during a pilot study, optimal for participants not to cast a shadow on the projection surface. In total there were 100 items per participant: 5 vertical x 5 horizontal positions x 2 words x 2 can sizes.

As soon as the participant pointed at the can with the laser pointer and said *piff* or *paff* to shoot it, the drawing of the can appeared to "fall." This was followed by a blank white screen and then the next can was presented. The presentation order was pre-randomized and operated manually by the experimenter. This allowed for more unpredictability in the timing, e.g., by incorporating pauses, so that the participants stay alert. In total, five data sets with pre-randomized presentation order were created to avoid order effects. For technical reasons, the presentation of the items was

FIGURE 3.2: Large and small cans in comparison. A large can was proportionally twice the size of the small can, as measured by the height of the can. Both cans displayed either *piff* [pɪf] or *paff* [paf].

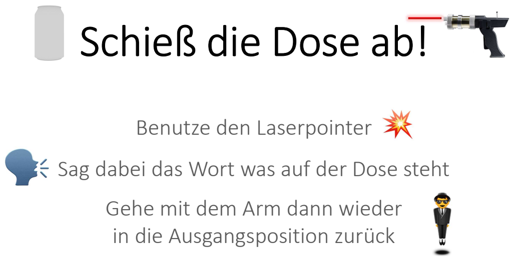


FIGURE 3.3: Instruction slide shown to the participants. The participants were asked to move their arm back to the onset position after each "shot".

divided into two sections, each consisting of 50 items. The main task was preceded by a short familiarization phase that consisted of five cans bearing words other than *piff* or *paff*. The whole experiment took 15-20 minutes, though the experimental task itself lasted no longer than 5-6 minutes in total. The length of the experiment was kept to a minimum to avoid boredom and its potential effect on voice quality. The instructions shown to the participants at the beginning of the procedure are given in Figure 3.3. They were explicitly asked to move their arm back to the onset position after every trial. The participants were given no other specific instructions regarding the arm movement towards the target and its alignment with saying the words. If one of them asked about it, they were told to act naturally, in a way similar to pointing a laser pointer at a particular word or image while giving an oral presentation.

We base our set-up on two iconic relationships: between vertical space and F0 or pitch, and between size and formant frequencies. However, we add the physiological component: head movement as a component of F0, and jaw opening as a component of F1. First, we evoke head movement by projecting the target on a large surface that is fairly close to the participant. We hope that such an arrangement will minimize the possibility to only move the eyes towards the target. Second, we use two visibly different target sizes to evoke different jaw opening.

Data Recording

During the experiment, the voice of the participants was recorded with a Sennheiser ME 64 cardioid microphone. The acoustic data were recorded at a 44.1 kHz sampling rate. In addition, the movements were tracked via an Optitrack motion capture system (software used: Motive, version 1.9.0) with 12 cameras (Prime 13). The precision after the calibration was 0.3 mm. The motion was captured with a 120

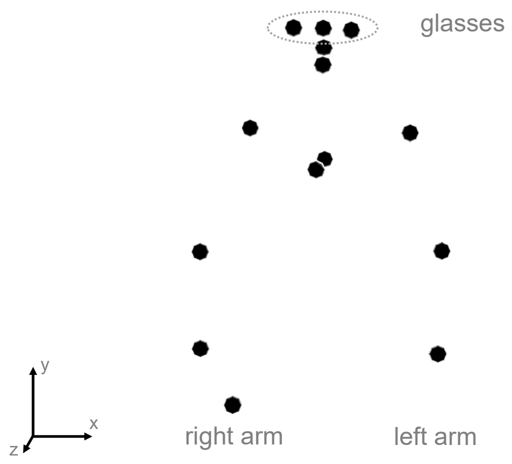


FIGURE 3.4: The placement of the motion capture markers. Here, it can be seen that the participant is holding the laser pointer in her right hand.

Hz sampling frequency. To test H1 and H2, following the motion of the head and the jaw was the most important. Nevertheless, to allow for additional measurements, such as head angle and arm movement, a total of 14 markers were placed on the upper body parts of the participants: three on a pair of purpose-built lensless glasses (left corner, center, and right corner); one each on the upper lip and the lower lip (jaw); one at the position of the sternum; one approximately at the location of the fourth thoracic spine vertebra; one on the laser pointer; and symmetrically two on the shoulders, elbows, and wrists. The placement of all markers is illustrated in Figure 3.4.

Participants

Since men generally have a lower fundamental frequency than women, including both genders would have resulted in yet another factor in the statistical analysis. Therefore, only women participated in the study. A total of 31 German native speakers took part in the study (mean age = 27.84 years, with min = 19 and max = 48 years; mean height = 167.7 cm, with min = 152 cm and max = 183 cm). Twenty-six participants were monolingual, and five reported being bilingual; all apart from one were right-handed. The participants were all recruited using a participant database of the Leibniz-Centre General Linguistics. Before the experiment, they were given basic information about the task and were asked to sign a consent form. They received monetary compensation for their participation. The project was approved by the ethical board of the German Research Foundation (DFG), and the data is available on Open Science Framework[3].

[3]The OSF repository can be visited under the following address: `https://osf.io/ysr75/`

Data Pre-processing

The acoustic data were automatically labeled at the phoneme level using WebMAUS (Kisler et al., 2017) and subsequently manually corrected using Praat (Boersma and Weenik, 2018). All parameters: F0-F2 were extracted automatically with a Praat script. The onset and offset of the vowel were defined as the onset and offset of vocal fold oscillations, respectively. The mean fundamental frequency in hertz was calculated for the whole vowel interval. For formants, median values in hertz were extracted, also in the whole vowel interval. During the automatic parameter extraction, the fundamental frequency range was set to 150–400 Hz to avoid octave jumps of the pitch tracker due to creaky voice.

The motion capture data were first extracted and processed with Mokka (version 0.6.2; Barre and Armand, 2014), and then converted for further processing with MATLAB (version R2017b). The position of the markers was measured within the three-dimensional space with reference to the ground level as a bottom for the y-axis. The maximal vertical position of the center of the glasses marker was extracted within the time frame of the vowel interval, which was provided by the annotated acoustic data.

Two participants were excluded from the analysis. First, speaker 16 was excluded due to a technical problem. In her case, the wrong microphone channel was recorded and the data was of very poor and possibly unreliable quality. After the initial inspection of the acoustic data, it was noticed that speaker 7 exhibited a very high mean F0 and mean F0 variance (mean = 314.86, SD = 39.95) in comparison to the overall mean and standard deviation (mean = 230.40, SD = 31). She was visibly excited during the recording, which is reflected by a high mean F0 of her voice. Her behavior did not reflect the participant's typical speech behavior and was thus excluded from the study. Also, the automatic extraction of all of the parameters yielded some missing values. Excluding the two speakers mentioned above, for F0, there were 2.24% missing data points, for F1 and F2 there were 0.38% and 0.03 % NA values, respectively, and for jaw opening, 1.03% data points could not be extracted.

For H1, outliers were removed for F0 mean and head position based on the Tukey rules on quartiles (i.e., +/- 1.5 IQR). For the former, 40 outliers were identified from among 2835 observations (equal to 1.41%). The mean F0 of the outliers was = 168.88 Hz, the mean F0 before outlier removal = 230.40 Hz, and after = 231.23 Hz. As for head position, one value from among 2900 was identified as an outlier (0.03%), its value was = 1.22 m. Before outlier removal the mean head position was = 160.31 cm and after = 160.33 cm. To test the H2, outliers were removed for formant values (F1 and F2) and jaw opening based on the Tukey rules on quartiles (i.e., +/- 1.5 IQR). For F1, 1 outlier was identified from among 2889 data points (0.03%), with the value = 1647.96 Hz. The mean F1 changed after removal from = 653.72 Hz to = 653.37 Hz. For F2, 13 outliers were identified from among 2899 observations (0.44%), with the mean = 705.46 Hz. The mean F2 before removal was = 1607.33 Hz, and after

= 1611.40 Hz. As for jaw opening, one outlier was identified from 2870 observations (0.03%), with the value = 5.49 cm. The mean jaw opening before and after outlier removal was = 4.26 cm.

3.5 Experimental Hypothesis Testing: Fundamental Frequency and Head Position

The following subsection deals with the statistical analysis of the data to test the H1: Fundamental frequency is influenced by head position – the more upward the head is rotated, the higher the fundamental frequency. In further subsections, the results are discussed in the light of current research.

3.5.1 Statistical Data Analysis

The data analysis was carried out using `R` (R Core Team, 2019, version 3.6.3). We used `tidyverse` package for data processing (Wickham, 2017) and `ggplot2` for data visualiztion (Wickham, 2016). We used the `brms` package (Bürkner, 2017) for Bayesian hierarchical modeling. This statistical approach was chosen for two main reasons.

First, the frequentist modeling approach, most often using the `lme4` package for `R` (Bates et al., 2014), has recently led to a discussion on the appropriate modeling for reliable hypothesis testing. In frequentist linear mixed-effects modeling, random effects are used to account for the variability in the data. It results, for example, from the individual speaker behavior or different influence of syllables that the participants are producing. Thus, random effects are justified by the study design – e.g., if, for some reason, we expect different syllables in the study design to yield a different effect on the outcome variable, we may want to include it as a random effect. It has been suggested that in order to account for such individual behavior, it is necessary to fit a maximal random structure justified by the design (Barr et al., 2013). However, as the design of the study grows in complexity, so does the random effects structure, which in turn leads to convergence issues. A solution has been proposed to use parsimonious mixed-models (Bates et al., 2015). In this paradigm, an iterative reduction of the random slopes is performed to establish the minimal random structure that is still maximally informative for the given data (know as "stripping" the model). But even a stripped version of the model may fail to converge, thus making it impossible to reliably decide which random slope to disregard. This discussion is irrelevant for the Bayesian inferential framework as a maximally complex model can be fitted without convergence issues.

Additionally, there are theoretical issues to consider. While the frequentist framework asks "How plausible is the data given the hypothesis?", the Bayesian approach asks about "How plausible is the hypothesis given the data?". Thus, here the data, and not the hypothesis, is given. In the Bayesian framework, data are the starting point to test how plausible the hypothesis is. The model is given the data and prior assumptions based on the previous research, and it computes posterior samples. Then, the plausibility of the hypothesis is estimated not via an arbitrary $\alpha = 0.05$ cut-off point but with an estimated certainty using probability values.

The hierarchical model to test H1 included the mean fundamental frequency as a function of vowel segment (two levels: /a/ and /ɪ/), can size (two levels: large and small), horizontal position (five levels: 1 leftmost and 5 rightmost), vertical position (five levels: 1 top and 5 bottom), height, and head position. All parameters except for head position are control parameters, as they were also manipulated within the study design. The model also controlled for by-speaker variation within vowel, can size, horizontal position, vertical position, and head position parameters by fitting a random intercept for speaker.

No priors were specified. The Student's t-distribution priors estimated for the model consisted of df = 3, $\mu = 235, \sigma = 29$. Six sampling chains with 6,000 iterations each were run, with a warm-up period of 3,000 iterations.

In the following section, multiple values are reported for each parameter: the 95% credible intervals and the posterior probability that the coefficient parameter is smaller than -1 $Pr(\beta < -1)$. The 95% credible interval (CrI) refers to the posterior probability distribution of the modeled data. On its basis, it can be said that there is a 95% chance that the effect falls within the range of the credible interval. There is no golden measure for the CrI, but 95% has been used in the past based on the frequentist tradition (cf. Kruschke, 2015; McElreath, 2018, for a discussion). In the current analysis, 95% is also taken as a measure. Therefore, I expect -1 not to be included in the 95% CrI and the $Pr(\beta < -1)$ to be close to 1 to constitute compelling evidence of a given parameter having an effect.

The posterior probability expresses the probability of a parameter having a certain value given the data and the model. The values are modeled on the basis of the data. They are subsequently tested against a meaningful threshold (e.g., chance level). Thus, in most cases, the values for every coefficient parameter are tested in comparison to 0 (i.e., how probable is it that it is bigger or smaller than 0). Yet, a difference that is slightly above 0 may be acoustically meaningless, thus, I take a more strict measure of 1 as reference. By using this calculation, we obtain a probability score of how many samples are greater than 1 (or in our case smaller than -1 – this is due to the initial data coding but does not have any influence on the result). To assess the meaningfulness of the posterior probability, I set the reliable score to 95%. While this value is the same as the frequentist cut-off point, it should not be confused with the theoretical meaning of a *p*-value (cf. Chapter 2.4.1).

3.5.2 Results

Table 3.1 and Figure 3.5 summarize the posterior distributions of all parameters tested in the model. As described in the previous subsection, values crucial for the interpretation are the credible intervals – which are based on the data, the priors,

TABLE 3.1: A summary of posterior distributions for all tested parameters. The column Estimate shows the posterior means; the 95% credible intervals are gives in brackets. The probability that the estimate is smaller than -1. The probabilities written in bold are greater than or equal 95%.

Parameter	Estimate (CrI)		$Pr(\beta < -1)$
Vowel	-18.89	**$(-23.34, -14.34)$**	**1.00**
Vertical pos.	-0.87	$(-1.86, 0.13)$	0.40
Horizontal pos.	0.01	$(-0.59, 0.61)$	0
Height	-7.21	$(-18.87, 4.51)$	0.85
Head position	-7.70	$(-16.21, 0.05)$	**0.95**
Can size	-1.50	$(-3.12, 0.13)$	0.73

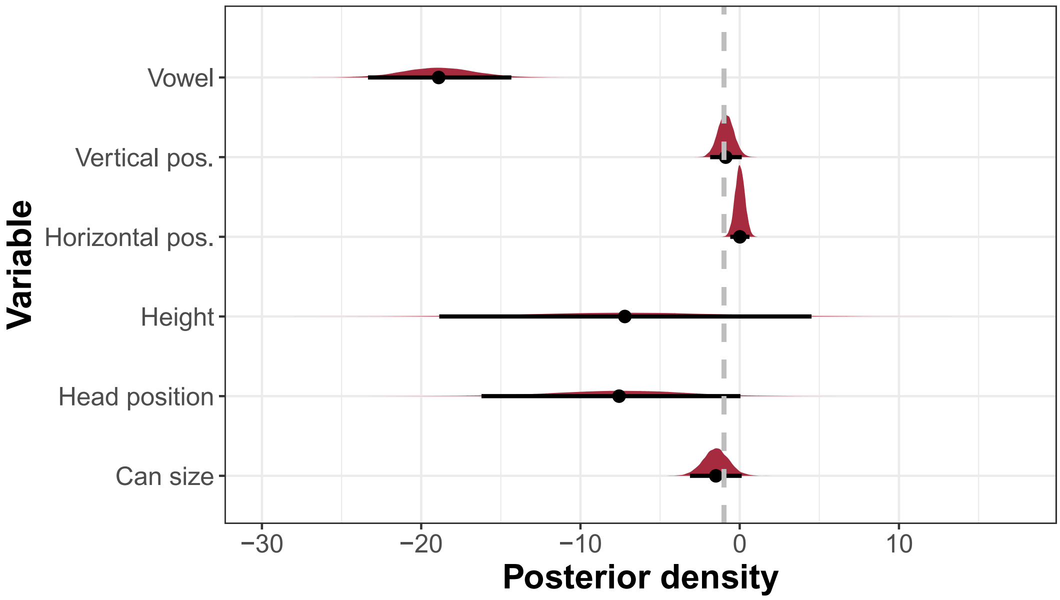


FIGURE 3.5: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the results of the H1 testing. Each of the variables is depicted on the y-axis. The x-axis shows the value of the predicted effect. The distribution of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value $= -1$.

and the model – and the probability that F0 change is caused by the head position change being above 1 Hz.

The effect of the parameter that is crucial for the H1, head position, is inconclusive looking at the 95% credible interval ($\beta = -7.70[-16.21, 0.05]$) since it includes the reference level -1. The posterior probability of head position reaching lower than -1 is at 95%. Thus, for the head position, the 95% credible interval does contain the reference level, as it spans from -16.21 up to 0.05. The posterior probability of the value is lower than the reference level 95% and therefore reaches the threshold that was set for the parameter having a reliable effect. The effect, given the data, the priors, and the model, can be interpreted as a change equal to one standard deviation, for head position $SD = 7.49$ cm, yields a change of -7.70 Hz. The lower the head position, the lower the fundamental frequency. The rest of the parameters serve as control variables that were manipulated within the experimental design. Vowel exhibits a robust effect on fundamental frequency ($\beta = -18.89[-23.34, -14.34]$), as the CrI does not include the reference point -1 and the posterior probability is at ceiling. Given the data, the priors, and the model, the estimated difference between the fundamental frequency of /ɪ/ and /a/ is -18.89 Hz. Due to its high variability, the effect of body height is inconclusive ($\beta = -7.21[-18.87, 4.51]$). Here, the credible interval contains the reference point and goes far beyond it, signaling an opposite effect for some cases. Nevertheless, the probability of the height sample being lower than -1 is 85%. Given the data, the priors, and the model, the estimated change of height equal to one standard deviation, ergo = 7.4 cm, yields a change of -7.21 Hz. The effect of can size on F0 is also ambiguous ($\beta = -1.50[-3.21, 0.13]$) due to the reference level also laying within the CrI. The posterior probability for can size being below the -1 threshold given this data reached 73%. Given the data, the priors, and the model, the estimated difference between the fundamental frequency when referring to small vs. to large cans is -1.5 Hz. For vertical position, no effect was found ($\beta = -0.87[-1.86, 0.13]$). Here the posterior probability of the estimate being lower than -1 equals 40%. There was no effect for horizontal position as well ($\beta = 0.01[-0.59, 0.61]$), with the lowest probability of all $= 0$. Thus, in our experimental setup, an estimate change equal to one SD in vertical ($SD = 1.42$ steps, from top to bottom) or horizontal ($SD = 1.41$ steps, from left to right) position, given the data and the model, yields an F0 change of respectively -0.89 and -0.01 Hz.

The results suggest that a possibly communicatively relevant influence on F0 is the most reliable for vowel. The parameter in question, head position, yields a reliable posterior probability of 95%, meaning that 95% of the samples generated based on the data, the priors, and the model do, in fact, lay below the threshold of 1 Hz. However, given that the sample spread expressed by the 95% credible interval is broad and entails the threshold itself, the effect found here is not entirely reliable. For the rest of the parameters, the impact on F0 is inconclusive, or there is no effect at all. Especially the can position parameters – vertical and horizontal position – yield no effect.

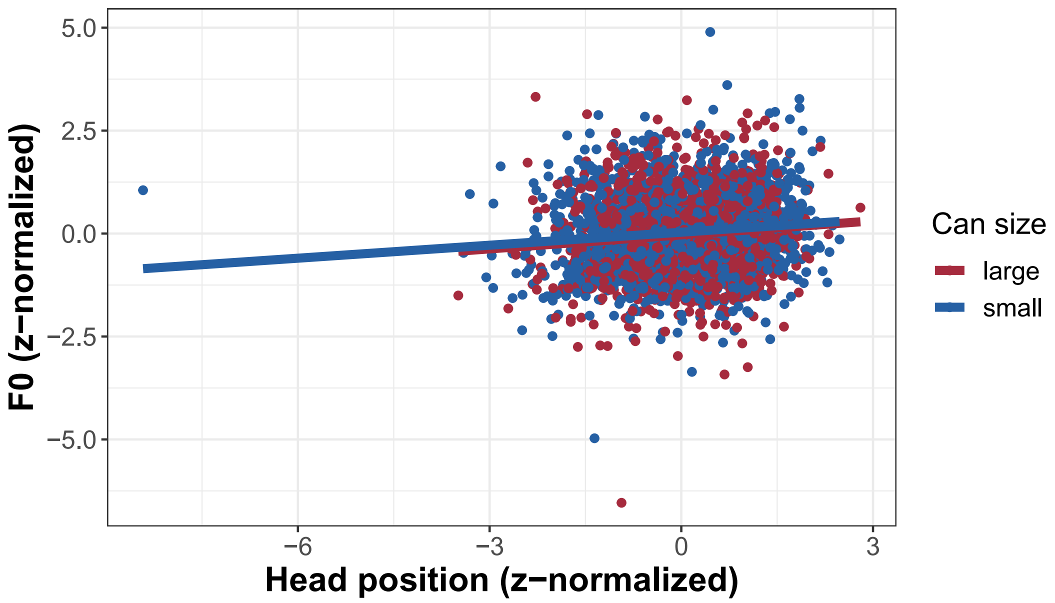


FIGURE 3.6: The relationship between F0 (y-axis) and head position (x-axis). The two colors represent large- and small-sized cans that yield no difference in the F0 here. A slight trend can be observed – raising the head position has a positive impact on the F0.

3.5.3 Discussion

The analysis described above tested the relationship between F0 and head position. It was hypothesized that the change in vertical head position influences F0 such that a higher head position yields higher F0. The results towards the hypothesis are inconclusive, and given the data, the priors, and the model, it can only be supported to a certain extent.

The effect is demonstrated in Figure 3.6. The slope of the plotted values is positive, suggesting a positive relationship between F0 and head position. The slope is not steep, suggesting a rather subtle effect of head position on F0. The posterior samples of head position are dispersed, suggesting high variability in participants' behavior ($\beta = -7.70[-16.21, 0.05]$). In some cases, the effect is evident, but there is a number of cases that show no effect at all. These cases reach a value smaller than -1. The variability may have several reasons. One of them is the individual movement behavior of a speaker. To assess how many people were prone to head movement, I set a cut-off point for "movers" at head position $SD > 2.5$. Thus, if the standard deviation of head movement was less than 2.5, the person was classified as a "non-mover". In the current data, there were 12 movers and 17 non-movers. Therefore, less than half of the speakers are prone to movement. Figure 3.7 shows that this parameter does not influence the effect between head position and F0. This means that even if the head is raised very little, the effect is visible. This can also signal that even when the movement is not radical enough to cause the F0 to raise due to the muscle tension, there cognitive forces may be at play to influence F0.

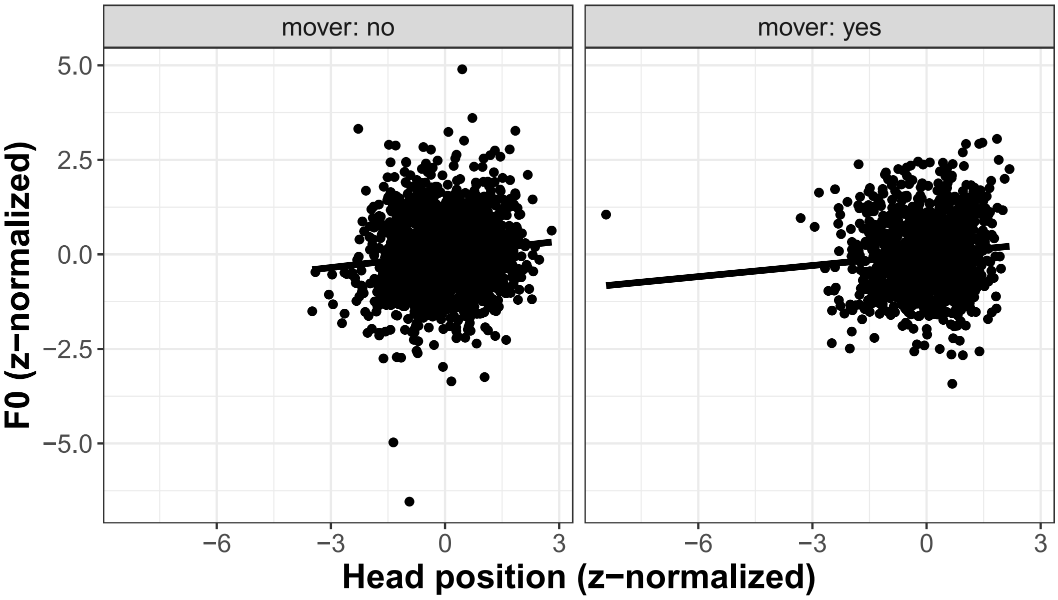


FIGURE 3.7: The relationship between F0 (y-axis) and head position (x-axis) divided by mover. Whether a speaker is a mover (head position $SD > 2.5$) or not (head position $SD < 2.5$) has no influence on the slope.

As mentioned, the effect shown here is based on the connection between the muscles responsible for both head movement and voice production. The results show some evidence that one can influence the other – if the head is raised, the F0 cannot be as low as it would be at a normal or low head position. Recent results by Liu et al. (2020) support the idea that head movement influences F0. The authors found a positive correlation between head movement and F0 in both congenitally blind and sighted speakers. For the latter, head movements were more pronounced for each corresponding change in F0, which suggests that, for sighted speakers, the head movement has a visual prosodic function. Liu et al. also found that the correlation between head movement and F0 was weaker for sighted speakers. The authors note that this may have social- (gaze control or eye contact) or discourse-related reasons. Biomechanical factors are only to some extent obstructed by movement, and there is a lot of freedom to raise or lower the voice at any head position. It has to be pointed out that there are many mechanisms that strongly influence F0 that have nothing to do with movement or physiology but rather with social expression.

F0 is to a large extent learned, therefore we may not have a purely physiological effect here. It has been shown that voice pitch can be used to express dominance (Stel et al., 2012) and can be interpreted as such, in turn being preferred when choosing a male partner (Apicella et al., 2007; O'Connor et al., 2014). Pitch has also been shown to assess trustworthiness – high female voices are perceived as more trustworthy in economic and mate poaching contexts, but in general low-pitched female voices are perceived are more trustworthy (O'Connor and Barclay, 2017). For males, high-pitched voices are perceived as more trustworthy in all tested contexts. Pitch and other linguistic markers can be used to show an affiliation to a certain group or social class. In his 1982 song "Valley Girl", Frank Zappa mocks the linguistic behavior of young upper-middle-class females who come from a certain region in California. One characteristic of their speech is the use of creaky voice (also known as vocal fry or glottal fry), a modality caused by lowering the voice to the extent where the vocal folds are loose and vibrate irregularly. The phenomenon is known for women across the US (Hinton et al., 1987; Yuasa, 2010) and in further English-speaking countries (e.g., Hornibrook et al., 2018). It has been shown that even two-thirds of young adult Standard American speaking females may be susceptible to vocal fry (Wolk et al., 2012). The discussion on how vocal fry is perceived is ongoing, but recent findings suggest that women using vocal fry are perceived as "less competent, less educated, less trustworthy, less attractive, and less hireable" (Anderson et al., 2014).

Fundamental frequency is, therefore, a powerful tool, and, as such, it can be manipulated for different purposes, e.g., to seem more important, more likable, or to identify with a group. The physiological contsraints are not the only ones at play.

Methodological Issues

As noted in Chapter 3.5.1, the Bayesian interference was used to perform the analysis. Initially, the data were analyzed using linear mixed-effect models. A preliminary analysis performed on a subset of participants (15 of 29) showed that head position has a robust effect on F0 (Ćwiek and Fuchs, 2019). The maximal model consisted of the following fixed effects: vowel in the uttered word, participant's height, participant's head position, can's vertical position, can size, and the interaction of vowel and can size (the latter was a suggestion from a reviewer). All apart of participant's height were included as random slopes for the random intercept of subject. Such a model reported convergence issues, therefore, a parsimonious mixed models approach was used (Bates et al., 2015). After the iterative model reduction, random slopes were: vowel and the interaction of vowel and can size.

Subsequently, the analysis was to be performed on all data, yet using the same approach repeatedly led to convergence issues. Knowing that this is the case with LMER due to a complex random structure, I assumed this is due to variance between the participants – the algorithm was unable to calculate a linear relationship. I grouped the participants into two groups: movers and non-movers since this was the parameter in question (cf. H1). As can be seen in Figure 3.7, both groups are very similar. Computing the two models for movers and for non-movers was unsuccessful as these too would not converge, even in their parsimonious variants. I then re-performed the preliminary analysis from Ćwiek and Fuchs (2019) by simply running the code. To my surprise, the old model described in the article would not converge. It has to be noted here that the `lmer` package is constantly developed and updated, so that some old models may not be reliable after a new algorithm update.

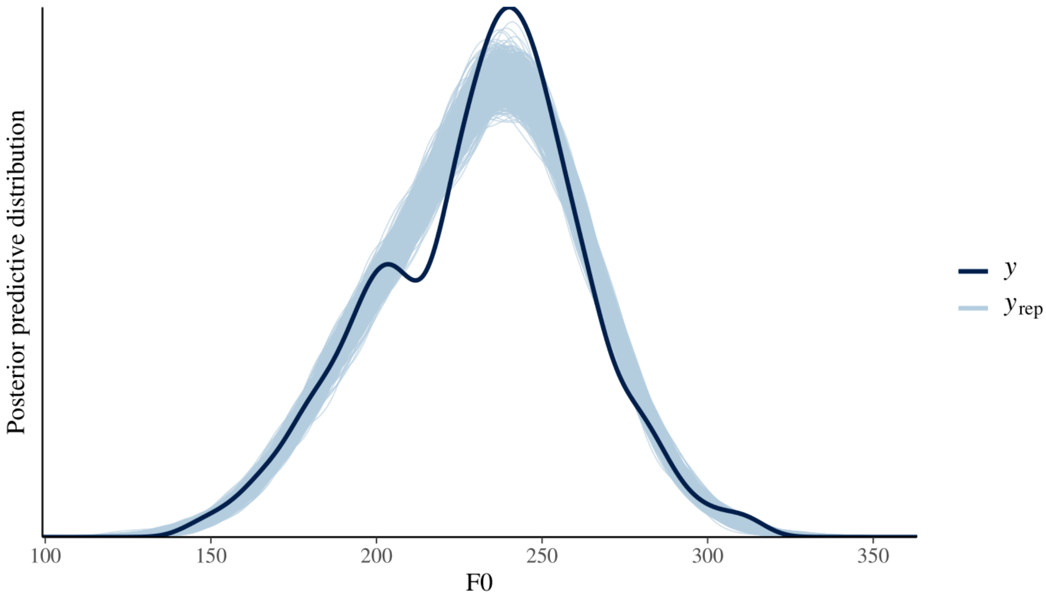


FIGURE 3.8: Posterior predictive checks for head position model. The y-axis here shows the hertz scale, and the x-axis depicts the density of the values. The thick dark line shows the collected data points, here stemming from the recordings. Lighter blue lines (N = 1000) are values predicted based on the priors, the data, and the model.

I then used an alternative method for solving convergence issues with LMER. This method, similarly to Bates et al.'s (2015) parsimonious mixed models, involves model stripping. The package `LMERConvenienceFunctions` (Tremblain and Ransijn, 2015) allows for fitting both random and fixed effects to maximally reduce the model. The models computed with this package, despite their reduced structure, were identified as singular. Singularity refers to a situation when one or more variances in the model are (close to) zero. Singular models are regarded as too complex, consequently have less power, and bear a higher risk to misconverge.

After having experienced problems with modeling the data with more conventional methods, I decided to use the Bayesian approach that has recently gained attention in linguistics (Roettger et al., 2019). The final Bayesian interference model is complete and includes all predictors and a full random effects structure that were initially established. Bayesian modeling with the `brms` package allows for graphic examination of the model fit using posterior predictive checks. This function compares the data density of the collected data points with the predicted data points based on the priors, the data, and the model. As can be seen in Figure 3.8, the actual data depicted by the thick dark line shows a bump at around 200 Hz, which is not shown in the data simulated by the model (light blue lines). This suggests that there is an additional factor responsible for the variance in the data that the current model does not cover. To my knowledge, this observation could not have been made when using frequentist analysis methods. Possible explanations for the bump are given in the Discussion section above.

Is Iconic Pitch Vertical?

In English, we use spatial terms to refer to vocal height – a voice can be "high" or "low". This metaphor is used in many languages, such as German *hoch* (high) and *tief* (low), Spanish *alto* (high) and *bajo* (low), Japanese 高い (high) and 低い (low), or Polish *wysoki* (tall, high) and *niski* (low). However, as variable as languages of the world are, there are also vastly different metaphors for pitch. A contradictory example can be found within the Indo-European language family – Persian speakers describe pitch as *nāzok* (thin) and *koloft* (thick). This relationship is also used in Turkish and Zapotec (Dolscheid et al., 2013). Further examples includes: light and heavy (Kpelle people), young and old (Suyá people), or weak and strong (Bashi people; Dolscheid et al. 2013). For pitches of musical instruments, it has been shown that metaphors related to age, gender, kinship, and social roles are used by African musicians (Ashley, 2004). For example, low frequency manuals of a mbira, a type of plucked idiophone instrument, can be referred to as "old men's voices", medial as "young men's voices", and high frequency ones are "women's voices".

In a series of experiments, Dolscheid et al. (2013) demonstrated that Dutch speakers – who natively use high and low metaphors for pitch – are insensitive to Farsi-like metaphor of thin and thick unless they are trained to do so. After the training, Dutch speakers exhibited the same pattern as Farsi speakers. Based on this results, Dolscheid et al. (2013, 620) conclude that "language-specific metaphors shape people's nonlinguistic representations of musical pitch". Therefore, regardless of the metaphor, the experience of pitch is the same across cultures. However, a mere opposition, even when it is marked, is not sufficient to represent pitch, e.g., speakers of English show no congruity effect when asked to judge pitch as front or back (Dolscheid and Casasanto, 2015). Interestingly, neither do they for terms "small" and "big" – a marked opposition – yet this might be due to the contradictory effects on F0 discussed earlier: big size evokes a low pitch, but the big size is also correlated with height in humans, that in turn evokes high pitch (cf. Ohala, 1994; Pisanski et al., 2017b). Dolscheid and Casasanto (2015) show, however that English native speakers exhibit congruency effects when judging pitch with terms "tall" and "short". Tall and short are words clearly depicting a vertical relation, just as high and low are. In this vein, they belong to the same conceptual category, thus the verticality of pitch may be conceptually engraved in humans (Casasanto et al., 2003).

The high conceptualization of the effect is further supported by experiments with infants (Dolscheid et al., 2014a; Tham et al., 2019). Dolscheid et al. (2014a) tested 4-moth-old Dutch infants' perception of two different metaphors of pitch – high and low, and thin and thick. For both pairs, the participants looked significantly longer at congruent pairs of space-sound stimuli (i.e., raising point or thinning line accompanied by a rising tone) than at incongruent pairs. The results suggest that the metaphor for pitch is present before it can be acquired with the language. Interestingly, both high and low, and thin and thick metaphors were interpreted for pitch equally well, differently as with adult participants (Dolscheid et al., 2013).

Studies are showing, unlike, e.g., Casasanto et al. (2003), that a horizontal domain may play a role in pitch representation as well. Most notably, Rusconi et al. (2006) have shown evidence for a relationship between musical pitch and vertical or horizontal position, calling it a SMARC effect (based on SNARC effect by Dehaene et al. 1993). The authors tested both naïve participants and expert musicians and demonstrated that the SMARC effect – high frequencies favored for spatially higher responses and low frequencies for lower responses – persisted even when the pitch was task-irrelevant. A horizontal mapping – where the left space is correlated with lower and the right with higher sound – was found for musicians only, yet the authors speculate it might be due to a remapping from vertical to horizontal domain (Rusconi et al., 2006, 127). In another study, it has been shown that musical training may activate a "music spatial line" that makes musicians perform better in congruity tasks between pitch and space (Lega et al., 2014). For horizontal association, especially pianist experience seems to be of importance (Timmers and Li, 2016).

Embodied cognition, a view that "cognitive processes are deeply rooted in the body's interactions with the world" (Wilson, 2002, 625), has also been quoted as a potential source of the vertical spatial metaphor for pitch. Zbikowski (1998) notices that low-frequency sounds are located lower in the body as they come from the chest, while high-frequency sounds are produced higher toward the head (e.g., singing head voice). It suggests that this relationship might, in fact, be embodied, as supported by unimodal activation of brain areas (Dolscheid et al., 2014b).

3.6 Experimental Hypothesis Testing: Jaw Opening and Object Size

The subsequent section reports the statistical analysis of the data to test H2: The degree of jaw opening is influenced by the size of an object – the larger the object, the larger the degree of jaw opening. In the following subsections, the results are outlined and discussed.

3.6.1 Statistical Data Analysis

Just like for H1, the data analysis was carried out using `R` (R Core Team, 2019, version 3.6.3) and the `brms` package (Bürkner, 2017) for the Bayesian hierarchical model. The reasons for choosing this approach over the frequentist one are clarified in Chapter 3.5.1.

The correlation between F1 and jaw opening in the data set was moderate at $r = 0.49$. Two models were calculated and compared to estimate whether jaw opening is a good physiological correlate of F1 in our data. The first model included F1 as a function of vowel segment (two levels: /a/ and /ɪ/) and can size (two levels: large and small). It controlled for by-speaker variation within vowel and can size by

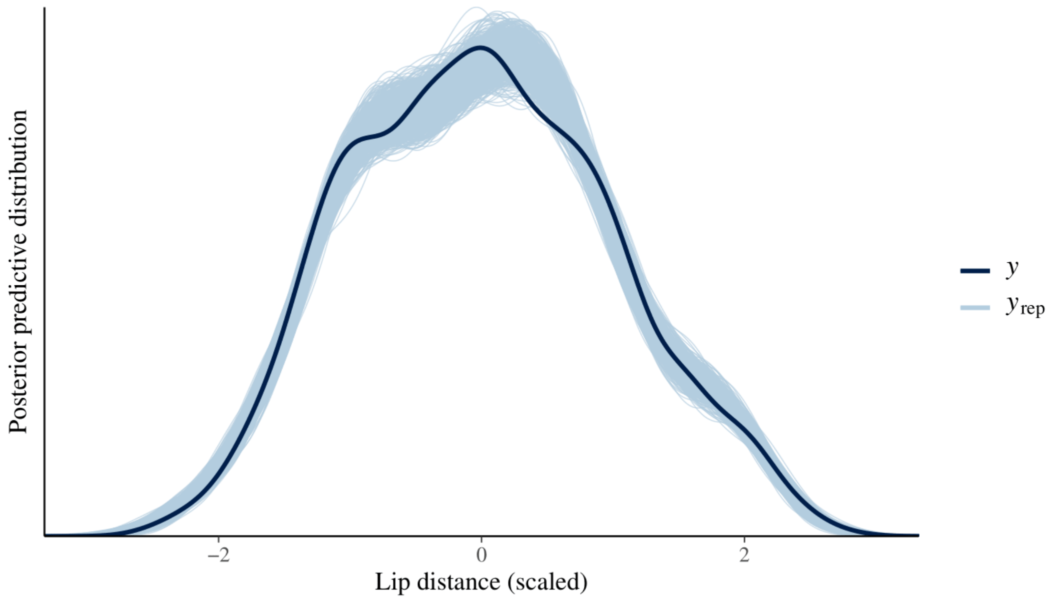


FIGURE 3.9: Posterior predictive checks for jaw opening model. The y-axis here shows scaled values of the degree of jaw opening, and the x-axis depicts the density of these values. The thick dark line shows the collected data points, here stemming from the recordings. Lighter blue lines (N = 1000) are values predicted based on the priors, the data, and the model.

fitting a random intercept for speaker. The second model estimated jaw opening as a function of identical parameters. The continuous variables, F1 and jaw opening, were scaled, and the categorical variables, vowel and can size, were contrast-coded for modeling.

Four sampling chains, each with 6,000 iterations, were run for each model, with a warm-up period of 3,000 samples. For both models, no priors were specified. The Student's t-distribution priors estimated for the both models were identical and consisted of df $= 3, \mu = 0, \sigma = 10$.

Model comparison was done with `brms` function `bayes_factor`. It revealed the Bayes factor in favor of jaw opening over F1 at 0.17. Thus, jaw opening seems to be predicted even better by vowel and can size. In addition, a model with vowel x can size interaction was calculated and compared to the jaw opening model. The Bayes factor in favor of the model without an interaction term over the one with an interaction term was 1603.65, implying a very strong preference for the model without the interaction.

The final model reported below thus includes jaw opening as a dependent variable. The model fit, shown in Figure 3.9, suggests that the modeled data (thin light-blue lines) fit the actual data (thick dark-blue line) very well, as their trajectory follows the same path. As for the H1, multiple values are reported for each parameter: the 95% CrI and the posterior probability distribution of the modeled data. For H2, all posterior samples were tested to be greater than 0, $Pr(\beta > 0)$, i.e., whether there

TABLE 3.2: A summary of posterior distributions for the parameters tested in H2. The column Estimate shows the posterior means; the 95% credible intervals are gives in brackets. The probability that the estimate is greater than 0 is given in the column $Pr(\beta > 0)$. The probabilities written in bold are greater than or equal 95%.

Parameter	**Estimate (CrI)**		$\boldsymbol{Pr(\beta > 0)}$
Vowel	1.06	$\mathbf{(0.94, 1.18)}$	**1.00**
Can size	0.01	$(-0.01, 0.04)$	0.83

was any difference made by the parameter, given the data and the model. Differently to H1, 0 was chosen for H2 because all the parameters in this model were scaled or contrast coded. The posterior probability was set at 95% to assess a meaningful difference (cf. Chapter 3.5.1). Thus if at least 95% of the posterior samples are greater than 0 and the 95% CrI does not contain the 0 cut-off point, the parameter will be regarded as meaningful.

3.6.2 Results

The results in form of the posterior distributions for vowel and can size parameters are given in Table 3.2 and Figure 3.10. The results are interpreted according to two crucial parameters: the credible intervals and the probability that the can size causes a larger than 0 change in the jaw opening. The 95% credible interval of can size, the parameter relevant in H2, shows that the effect of can size has on jaw opening is inconclusive ($\beta = 0.01[-0.01, 0.04]$) because it includes the reference level 0. The posterior probability of can size being higher than 0 is 83%, a number lower than would be expected in a significant influence on jaw opening, set at 95%. It means that, given the data, the priors, and the model, 83% of the posterior samples for the large can size yield a larger jaw opening than the small can size does.

The effect of vowel, a control parameter, is evident looking at the 95% credible interval ($\beta = 1.06[0.94, 1.18]$) since it lies far beyond the reference level 0. The posterior probability of vowel being larger than 0 is 100%. Thus, given the data, the model, and the priors, 100% of the posterior samples for the vowel /a/ result in a larger jaw opening than for the vowel /i/.

Overall, the results suggest that the vowel segment has a robust effect on the jaw opening, with jaw opening being significantly larger in /a/, in comparison to /i/. The effect of can size on jaw opening, as hypothesized in H2, was not found.

The lack of an effect of can size on jaw opening is demonstrated in Figure 3.11. Large-sized cans are represented in red, on the left side for the given vowel, and small-sized cans in blue, on the right side for the given vowel. Within vowel segments, /a/ and /i/, the medians and interquartile ranges do not differ between large and small cans, similarly to the whole distribution of the data points, as depicted by

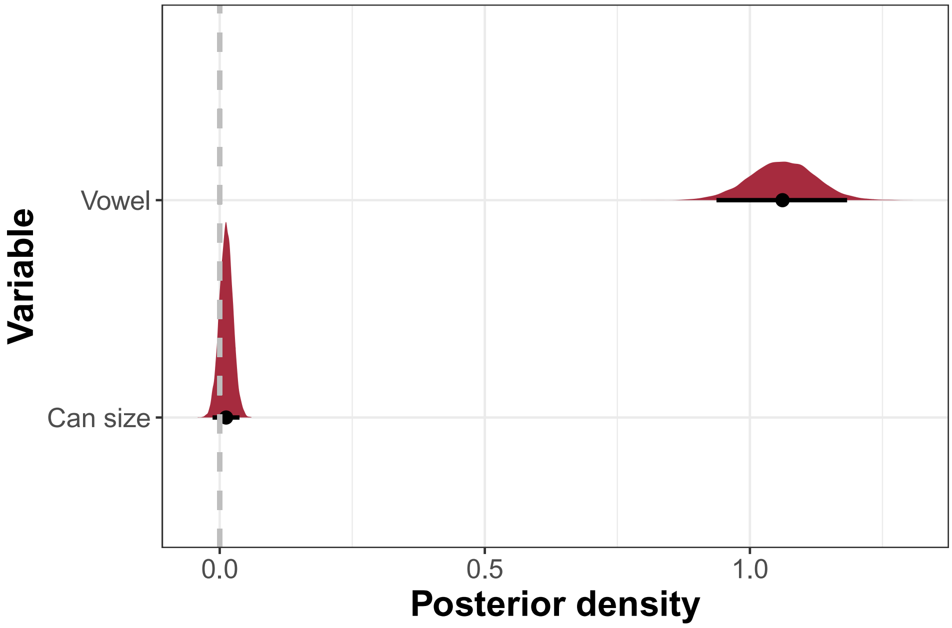


FIGURE 3.10: Posterior densities with posterior means (black dots) and 95% credible intervals (black lines) for the results of the H2 testing. Each of the variables is depicted on the y-axis. The x-axis shows the value of the predicted effect. The distributions of all values for the given variable were combined to a density plot. The gray dashed line is drawn at the x-axis value = 0.

the violin shape. The robust effect of vowel on jaw opening is visible when comparing large- and/or small-sized cans across the vowels. The difference between the medians for /a/ and for /i/ is approximately 0.5 cm, with mean jaw opening for /a/ = 4.5 cm and for /i/ ≈ 4 cm.

3.6.3 Discussion

The analysis outlined above tested whether the degree of jaw opening is influenced by the object of reference, ergo whether larger objects result in a larger jaw opening than small objects. Based on the data, the model, and the priors, the hypothesis was refuted.

Figure 3.12 presents the formant areas of both vowels in the collected data. Additionally, large and small can sizes were color-coded for each vowel. The areas for both sizes are almost identical. The plot reveals a lack of difference in the formant distribution with regard to large and small cans. To double-check the visual impression mathematically, the areas of the ellipses were calculated. First, the coordinates of each ellipse were extracted, then its center, and the distances to the center from each point on the ellipse. The area was then calculated using the equation $A = \pi \times a \times b$, where a is the semi-minor and b is the semi-major axis, i.e.,

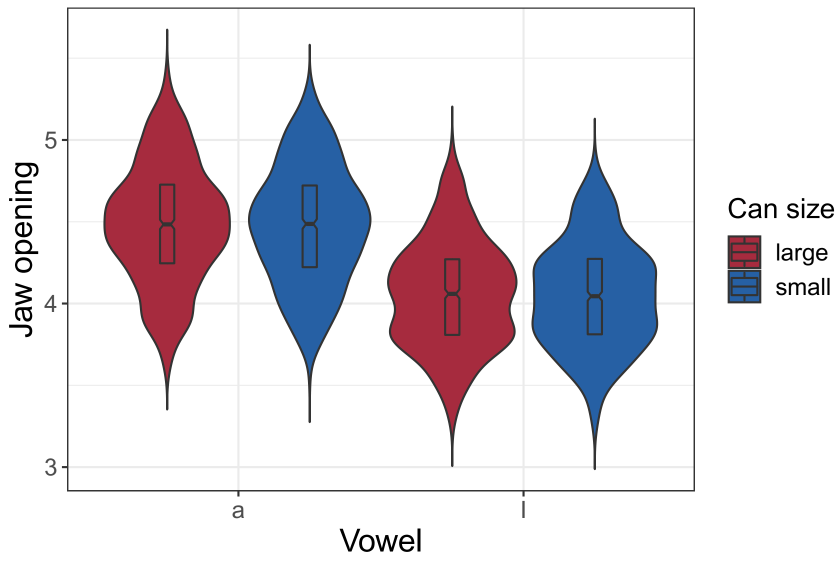


FIGURE 3.11: The relationship between jaw opening (y-axis; in centimeters) and vowel (x-axis; binary /a/ and /i/). The color-coding on the x-axis represents two can sizes: red depicts the large- and blue the small-sized cans. The violin plot visualizes the distribution of the data – the wider the shape, the more data points are included. The median and the interquartile range are depicted by the darker vertical bar within the violin plot.

the smallest and largest distance of a value to the center. The paired *t*-test comparing large and small areas for /a/ and /ɪ/ revealed no difference between the areas, $t(1) = 0.9, p = 0.5$.

The strong effect found for vowel on jaw opening is in line with basic articulatory phonology, where /a/ is classified as an open vowel and /ɪ/ as a closed vowel. In our data, the difference of jaw opening between the two vowels was ≈ 0.5 cm. This effect is nevertheless very robust, as it is anchored in the articulatory-acoustic properties of the vowels. The physiological ground for the expression of size iconicity, i.e., larger jaw opening for larger objects, was not found in our data, yet a few factors may be at play here. Firstly, the expression of size iconicity within a segment may be more subtle than initially anticipated. Secondly, the iconicity of size may apply only to animate entities and not to inanimate objects. Also, it may be the case that the communication of size is limited to communicating own size by the source.

Subtle Expression of Size Iconicity within a Vowel

The clear-cut difference found in jaw opening for the vowel segment, as briefly mentioned above, has its ground in the articulatory-acoustic properties of vowels. Phonetically, openness is one of the crucial parameters to distinguish vowels as it directly influences the F1 – the more open the jaw and the lower the tongue, the higher the F1. The tongue covers relatively small distances within the oral cavity to emphasize the differences between the vowels. For both jaw and tongue movement, there

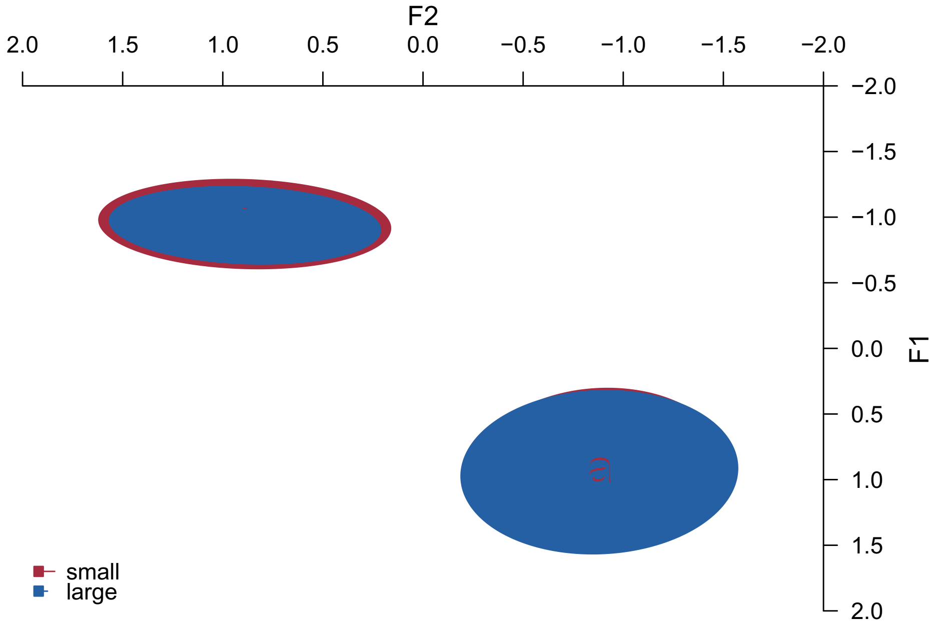


FIGURE 3.12: Formant areas (z-normalized) for /a/ and /ɪ/ with relation to can size. The values for large-sized cans are colored red and for small-sized cans blue. It can be observed that can size does not affect the formant distribution.

might just be not much more room for variance within a single vowel that would become visible with the method applied in the current study. In this line, it might be the case for size iconicity that the distinction between the vowel segments itself is sufficient (cf. Chapter 3 on sound symbolism).

In the current study, two lax vowels /ɪ/ and /a/ were tested. In German, there are both tense and lax vowels, with lax vowels being more prone to coarticulation than their tense counterparts (Hoole and Mooshammer, 2002). Tenseness involves more muscular effort, plus, in German, tense vowels are usually long. In contrast, lax vowels are often short (with some exceptions) and articulated with less effort. Because of less muscle tension in lax vowels, the target of the vowel is not clearly reached as the preparatory movement for the subsequent segment takes over. Interestingly, Hoole and Mooshammer (2002, 136) observed an inter-speaker variance in the jaw movement with regard to tense-lax vowel opposition.

The jaw is principally made of a heavy bone structure compared to, e.g., the tongue made of much lighter soft tissue. The lower jaw movement involves the mandible bone itself, a set of joints (temporomandibular joints), and four different muscles (masseter, temporalis, and the medial and lateral pterygoids), leading to a rather effortful process. According to economy theory, such as the hyper- and hypospeech (Lindblom, 1990), speakers adapt their speech by varying the effort of speech production to listeners' needs in order to achieve a sufficient discriminability of speech segments. The jaw opening is naturally limited to a few centimeters, but there is also a high degree of freedom to how far speakers open their jaw when speaking. Economically, the smaller the jaw opening, the less effort in articulation. As the discriminability becomes harder to achieve, more articulatory effort is used. This can be the case, for example, in a noisy environment or during listener-endorsed repetitions. The task in the current study did not involve a communicative situation, as the outspoken words were not directed to any listener. And as iconicity serves the purpose of depiction for the listener (Dingemanse, 2015, 950), when a listener is not present, the depictive purpose might become irrelevant. Therefore, it can be assumed that, with a lacking communicative goal and the lacking need for depiction, participants were behaving economically, with rather less effort, thus less jaw opening, than in a conversation scenario.

Animacy Effect for Size Iconicity

The second possibility, why the effect was not found, is the inanimacy of the target in the current experiment. Here, the participants aimed at cans and "shot" them with their voice and a laser pointer. As mentioned above, it was far from a communicative situation, in which the participants would have to act listener-oriented as well. Past research revealed an influence of the fundamental frequency and/or formant frequencies on human body size estimation (e.g., Pisanski et al., 2014b, 2017b; Pisanski and Rendall, 2011). Similarly, vocal frequencies correlate with measured body size of other species, such as dogs (Riede and Fitch, 1999) or frogs (Gingras et al., 2013). A subspecies of the latter has been explicitly shown to lower their F0 in defense territories (Bee et al., 2000) and many non-human vertebrates have different vocalizations for hostile and friendly signals (Morton, 1977, 1994). The manipulation of the voice has a clear communicative goal – it should evoke a certain reaction in a recipient. Lowered F0 serves as a marker of largeness and/or hostility (Morton, 1994; Ohala, 1994), a trait that is highly important in an animal's territory. In humans, the interplay of perceived pitch and formant frequencies has also been found to influence the perceived attractiveness of males (Feinberg et al., 2011, 2005; O'Connor et al., 2014; Pisanski and Rendall, 2011). Applying all of these vocal associations wins on importance only in the case where the signal is going to be actually received by the target. Thus, having an animate recipient may be of the highest priority when it comes to iconically marking the size of an entity.

Notably, the experimental task was designed to communicate the size of the referent and not the speaker. However, all of the studies mentioned above report the opposite. The communication of size regards the source – by vocalizing, an animal communicates its own size, and a human their own size. This does not imply that the mechanism can be applied to communicate iconically about the size of other entities.

3.7 Chapter Discussion: Iconicity is (also) Sensory

Involving various acoustic dimensions – such as intensity, intonation, rhythm, and spectral phenomena – prosody is an ideal means to express iconicity. Saying 'loooong' iconically portrays and intensifies the meaning in comparison with the regular 'long' (cf. Fuchs et al., 2019, for evidence in written corpora). The experiment described in this chapter investigated a potential physiological origin of two iconic relationships: (1) between vertical position and fundamental frequency, and (2) between object size and formant frequencies. The hypothesized physiological roots for the relationships were head position and jaw opening, respectively.

The results of a Bayesian hierarchical model for (1) were inconclusive, as the credible interval includes the reference level. Yet, the posterior probability of head position at 95% suggests that head position might affect F0. For (2), no effect of jaw opening with respect to size was found. As reported, e.g., by Hoole and Mooshammer (2002), also in our data set, there were differences in participants' behavior – whether they moved their head or not, and how far they opened the jaw, respectively.

As Figure 3.8 illustrating the model fit reveals, there is some inconsistency between the modeled and actual data. The bump of the thicker blue line at around 200 Hz could not be explained by any of the factors in the model, as it is not followed by light-blue lines – representing the modeled data. Therefore, a question arises, what other factor is missing that would explain the variance in the data? Maybe there is another force driving the iconic relationship under investigation?

Cross-modal correspondences – such as those between pitch and elevation, or pitch (evidence for both pitch, e.g., Pisanski et al. 2014a and formants, e.g., Fitch 1997; Pisanski and Rendall 2011) and size – are everywhere (Spence, 2011, 975). Both English- and Hindi-speaking consumers perceive a beer as lighter when it has front vowels or voiceless stops in the name, as compared to dark beer with back vowels and voiced stops (Athaide and Klink, 2012; Klink, 2000, among other preferences). Therefore, at least some cross-modal correspondences, similarly to relationships based on iconicity, seem not to be anchored in a single phonological system and go beyond that. Such correspondences, which tightly couple two different modalities and are non-arbitrary (Melara and O'Brien, 1987, 323f.), can be identified as synesthetic. Synesthesia is a perceptual phenomenon – one perceives a stimulus in a certain modality and experiences another one simultaneously (e.g., hearing C-sharp may induce a sensation of a blue color, Ramachandran and Hubbard 2001). Similarly, cross-modal correspondences are a perceptual phenomenon – the majority, if not all, of the studies investigate the perception. While cross-modal correspondences are very common among populations (cf. Spence, 2011), synesthesia was disputed for many years (cf. Baron-Cohen et al., 1996; Baron-Cohen and Harrison, 1997, 1073) and estimations of how many synesthetes there are vary (cf. Ramachandran and Hubbard, 2001, 6, who propose 1 in 200 people). Because of the similarity between both phenomena, Martino and Marks (2001) suggested that cross-modal correspondences should even be considered weak forms of synesthesia. A subsequent criticism by Deroy and Spence (2013) revealed, for example, that, while synesthesia is an automatic phenomenon, there is a degree of control in cross-modal correspondences. Furthermore, the former is a conscious experience, whereas the latter not necessarily. Also, taking into account a size–pitch correspondence, it becomes evident that the correspondence here is relative, while synesthesia operates in absolute relationships (cf. Deroy and Spence, 2013, 653ff. for a complete list of arguments).

In the real world, certain sensory information co-occur together more frequently. For example, the pitch–elevation correspondence, also investigated in the current study, has been reported and replicated many times to date (e.g., Bernstein and Edelstein, 1971; Melara and O'Brien, 1987; Parise et al., 2014; Pratt, 1930) and is said to be very robust across populations (Evans and Treisman, 2010; Spence, 2011). There seems to be a degree of how prevalent the correspondence occurs that may also vary from modality to modality. For example, a correspondence between *high* pitch and *high* elevation can be seen as iconic – high is high and low is low; ergo, the meaning resembles the form. It may be less evident when we consider an example given in the previous paragraph with front vowels and voiceless stops suggesting a light beer. The correspondence between sound and taste seems less iconic at first glance, but it is still cross-modal. According to Spence, the multisensory integration is influenced by cross-modal correspondences (Spence, 2011, 982). He states that "[t]he stronger the coupling, the more likely it is that the original unimodal signals will be fused completely into a single integrated multisensory percept" (Spence, 2011, 984). According to the Bayesian view, the strength of the coupling is modulated by the statistics in the natural environment correlations (Ernst, 2006; Parise and Spence, 2009). Therefore, if two phenomena across two modalities co-occur together more often, they are more likely to be integrated into one. The reflection of the effect in the environment (coupled spatio-temporally) increases the anticipation that the two stimuli do correspond. This leads to the perceptualization of the effect (Spence, 2011, 985).

In the current experiment, I looked for a physiological explanation of the iconic correspondences between pitch and elevation, and pitch and size. Both correspondences are also cross-modal, and, as suggested by Spence (2011, 985), very likely perceptualized. Consequently, the effects in question may not be purely physiological anymore. In our data, we were able to identify that the majority of the participants were not prone to head movement (cf. Chapter 3.5.3), yet this did not have an influence on the fundamental frequency (cf. Figure 3.7). This finding supports the perceptualization of the correspondence between fundamental frequency and elevation and suggests that other factors than head movement are at play.

Some of the cross-modal correspondences, based on multisensory perception and integration, are iconic in nature, such as those between pitch and elevation and pitch and size. Still, some correspondences that are not iconic in our sense also get habituated because of the natural environment statistics. The goal of future research might be to disentangle which of the cross-modal correspondences are iconic in their nature and which of them are not. It is also possible that sensory phenomena should be integrated into our thinking about iconicity more tightly. Differently from cross-modal correspondences, iconicity itself has been investigated in both perception and production, which are tightly coupled with one another (e.g., Ohala, 1996). Especially in speech production, certain sensory responses connected to movement – for example, of the articulators (Lieberman, 1970) – might occur and become grounded in the sensory-motor systems (Leshinskaya and Caramazza, 2016). I propose, therefore, that properties of iconic relationships are sensory-based. In further investigations of the origins of iconic phenomena, apart from the sensory-motor representations, also theories concerning sensory integration phenomena should be considered.

Chapter 4

Ideophones

4.1 Introduction to Ideophones

Previous chapters provided insight into how iconicity manifests itself on the segmental and suprasegmental levels. However, iconicity is probably the most evident on a word level to any language user. Onomatopoeic words such as *boom*, *bang*, or *swish* are present in the pop culture and in daily life. Film titles, song lyrics, commercials, and even books contain many such examples. In the early discussion on the principles of linguistic signs, Saussure (2011) claims that, despite their depicting character, onomatopoeic words do not undermine the principle of arbitrariness. They are "limited in number, but also they are chosen somewhat arbitrarily, for they are only approximate and more or less conventional imitations of certain sounds (cf. English *bow-bow* and French *ouaoua*)" (Saussure, 2011, 69). The author overlooked, firstly, the role of the phonological system and phonotactics. Both phonology and phonotactics shape the form of the word by constraining, which sounds and syllable structures are allowed in the given language (cf. Akita, 2009, 37). Secondly, Saussure did not mention that there exist many more examples for sound-symbolic words other than those based on sound imitation.

An ideophone is defined as "[a] member of an open lexical class of marked words that depict sensory imagery" (Dingemanse, 2019, 16). In other words, ideophones portray any sensory experience, whether it is a sound, an image, or a feeling. For example, Japanese mimetics (a term used for Japanese sound-symbolic words, cf. Akita 2009, 9) *kôrokoro* and *gôrogoro* both depict an object rolling. The difference in voicing iconically depicts a difference in the quality of the object. While *kôrokoro* is used for a small, light object rolling, *gôrogoro* depicts a large, heavy object rolling (Akita, 2009, 38). Ideophones are marked, because they often stand out syntactically, prosodically, or even phonotactically. As an example, Akita and Dingemanse (2018, 3) note that, in Hausa, many ideophones, unlike other members of the lexicon, have consonants in the final position, e.g., tsít 'in a complete silence' or kàzàʀ-kázáʀ 'in an energetic manner'. Lastly, Dingemanse (2019) characterizes ideophones as belonging to an open lexical class. A whole lexical class of ideophones – also referred to as mimetics, or expressives – exists in languages such as Japanese (Japonic; Akita, 2009), Siwu (Niger-Congo; Dingemanse, 2011), or Awetí (Tupian; Reiter, 2011). Numerous other examples are given by Voeltz and Kilian-Hatz (2001). The members of the lexical class of ideophones go far beyond the scope of the auditory domain. In her book on Pastaza Quechua sound-symbolic words, Nuckolls describes categories of ideophones depicting contact and penetration (Nuckolls, 1996, 178ff.), deformation (Nuckolls, 1996, 232ff.), or suddenness and completiveness (Nuckolls, 1996, 250ff.). An example for an ideophone belonging to the latter category is *dzing*, that depicts "a sudden awarness or intuition" (Nuckolls, 1996, 250).

Now, onomatopoeias are strictly connected to the imitation of sound. Ideophones, however, are not restricted to sound imitation and can depict other dimensions, too. The German ideophone *ruckzuck* 'very fast,' for example, refers to a kind of movement. According to both Akita (2009) and Kilian-Hatz (1999, after Dingemanse, 2012, 657), there exist differences whether onomatopoeic ideophones should be regarded separately to other kinds of ideophones within a language, because onomatopoeic ideophones are said to have a "more peripheral syntactic realization" (Dingemanse, 2012, 657). Here, I maintain that onomatopoeic words are a subgroup of the larger class of ideophones, namely that they are sound ideophones.

It is a fact that languages vary greatly in the number of ideophones. In some languages, they represent a whole lexical class. In Japanese, there are over a thousand mimetics (Akita, 2009, 1), which are dispersed across syntactic classes – such as verbs, adverbs, nouns, and adjectives (McLean, 2020, 3) – and which are widely used in everyday conversations and writing (Nuckolls, 2004). They express subtle differences through iconic properties of sounds, e.g., voiced consonants depict heavy or large objects, while voiceless consonants light or small objects (Hamano, 1998, 172). In English however, the number of sound-symbolic words is limited, similarly to other Indo-European languages such as German (Akita and Dingemanse, 2018, 6; Dingemanse, 2019, 21; Laing, 2019, 16; Nuckolls, 2004, 132). There exists work on German onomatopoeia (Havlik, 1981), however, it focuses on sound-imitative words used in comics and covers a wide range of idiosyncratic instances, e.g., *marf* 'traffic noise' (Havlik, 1981, 102), or *plunkty* 'jangling dress made of metal plates' (Havlik, 1981, 121). Why then are Indo-European languages, in comparison with, e.g., Japanese, or Bantu languages, so ideophonically impoverished (Nuckolls, 2004)?

According to Nuckolls (2004), the use of ideophones simulates the described event. Ideophones are performative and transfer the interlocutors into the situation. The use of ideophones requires a connection between the human and non-human world – between one him/herself and nature. Nuckolls (2004, 133) notices that Quechua speakers, whose language is abundant in sound-symbolic words, change their use of ideophones when their experiential framework changes. The group that she studies sustains a relationship with its surroundings based on animacy. However, its members have been faced with an increased influence of the Ecuadorian government and nation. In turn, their view of the connection with nature deteriorates towards a Judeo-Christian view of nature. This view, on the other hand, projects nature as antagonistic to humans, and has widely influenced the cultures of the Indo-European languages leading to diminishing the number of sound-symbolic words (Nuckolls, 2004, 133).

Another factor briefly mentioned by Nuckolls (2004) is an increased literacy. The constraints of orthography and written language seem to have a deteriorating effect on ideophony. This assumption, however, contradicts the existence of a rich class of ideophones in Japanese, where ideophones are also used in written language (Akita, 2017). At this point, it is worth noting that the sound-symbolic words in German are rarely used in formal speech or written language. And even in other cultures, such as the Awetí community, speakers are reluctant to write down the ideophones. As revealed by Sabine Reiter (author of Reiter, 2011) in a private conversation, after being asked about why he omitted an ideophone during the transcription, one of the members of the community told her "this isn't important, we do not write that." Thus, even members of the Awetí community, which widely uses ideophones, resist using them in writing. The situation may be similar to other words frequently used in human interaction that rarely find entrance into the orthographical conventions (e.g., false starts, filled pauses like "uh", repairs, etc.).

Ideophones can be seen as iconic, yet, not necessarily across languages. The German ideophone *ruckzuck*, 'very fast,' contains sound-symbolic means that give insight into its meaning, such as syllable repetition with onset variation, or the combination of continuants in the onsets and plosives in the codas (cf. Kentner, 2017). Recent evidence shows that iconicity helps learn new words, and naïve participants can infer the meaning of sound-symbolic words better than arbitrary words (Lockwood et al., 2016a,b; Nygaard, 2008; Nygaard et al., 2009a). However, the exact meaning of the ideophone within a language remains opaque despite the iconicity. As shown by Iwasaki et al. (2007b), English native speakers can guess the intention of Japanese ideophones only partly. The authors conclude that, to some degree, the iconicity in ideophones is language-specific and conventionalized. Therefore, ideophones have cross-linguistic attributes, but for non-native speakers, the sensory experience of the ideophone and the convention within a given community is missing for a correct interpretation. To some extent, the culture shapes our perception of iconicity.

In this part of my dissertation, I will bring the topic of ideophones closer by inspecting the ideophone inventory in German. Diffloth (1972, 440) notes that "a wide geographic and historical distribution indicates that ideophones are characteristic of natural language in general, even though they are conspicuously undeveloped and poorly structured in the languages of Europe" (after Dingemanse, 2019, 2). The goal set for the following chapters is to gather and present a legitimate scope of the ideophone inventory in German.

Similar to what has been said for Japanese above, German ideophones are scattered across the lexicon. Instances of sound-symbolic words are present among interjections, e.g., *nanu* depicting surprisal or amazement, adverbs, e.g., *ruckzuck*, 'very fast', but also nouns, e.g., *Kuddelmuddel* depicting a chaotic state, and possibly other parts of speech. Therefore, as a first step, I will present a newly collected data set of German ideophones. Despite its broad range, a data set is a mere collection of instances and does not contain various critical information, such as the intended meaning. Thus, in the next step, I will organize the findings in the form of a dictionary proposal. So far, there exists a lexicon of German onomatopoeia found in comic books (Havlik, 1981), but no work exists that also shows examples of use and the variety of meaning. To the best of my knowledge, a comparable work exists only for Polish onomatopoeia by Bańko (2008). In Polish, like in German, ideophones are scattered across the lexicon. They are also less frequent than in Japanese; therefore, ideophone inventories of both languages may serve for comparison. The book by Bańko (2008) encompasses not only onomatopoeia but also other sound-imitative words, like *łap-cap*, 'grasping or catching something in a fast manner,' or *rach-ciach*, 'fast,' which depict a kind of movement. A collection of a lexicon of German ideophones and its further expansion to a full-fledged dictionary would fill the gap in German linguistic research. The design should profit language scientists from various specializations, e.g., phonetics, semantics, and morphosyntax. The completion of an entire dictionary lies outside the scope of this work. Therefore, the purpose of the current lexicon and subsequent dictionary proposal is to set a framework in which a comprehensive collection of ideophones, including various sensory modalities rather than just sound imitations, could be presented to a broader public. With such a collection at hand, comparative work on ideophone inventories across cultures and languages could be conducted in the future.

This qualitative work is followed by two quantitative analyses of the German ideophones. The first, in Chapter 4.4, aims to test the ideophone hierarchy proposed by Dingemanse (2012), after which there is a particular order in which ideophones can occur in a language. The second, in Chapter 4.5, involves testing a hypothesis that the use of ideophones in books varies for different age groups. Both experiments are based on the initially collected data set and demonstrate how the data set can be employed for further hypothesis testing. Finally, the chapter is capped with a general discussion.

4.2 Newly Collected Data Set of German Ideophones

Ideophones can be easily found in comic books. An example of their use is displayed in Figure 4.1. The ideophones – all words in Figure 4.1 except for *Hilfe!* 'help!' – stem from other parts of speech, mostly verbs. Starting with *knurr*, from the verb *knurren*, which means 'to growl,' then *brumm*, from the verb *brummen* meaning 'to buzz,' and finally *kratz*, which stems from the verb *kratzen*, 'to scratch.' The word *tatz* can be identified as stemming from the noun *Tatze* (f.), 'paw' or 'large, strong hand.' Lastly, *zack* can be interpreted twofold – as stemming from the verb *zacken* 'to cut in a form of zig-zag,' or from the noun *Zacke* 'peak, point, tooth.' However, *Zack* functions as

FIGURE 4.1: Ideophone use in German comic books (Bottaro et al., 1997, 97). The upper left square shows the multitude of sound-symbolic words – with the exception of *Hilfe!*, 'help!'. Some of them stem from verbs (e.g., *kratz* from *kratzen* 'to scratch') or nouns (e.g., *tatz* from *Tatze* (f.) 'paw' or 'large, strong hand').

an alone-standing noun in expressions such as *auf Zack sein*, where it refers to being prepared to react *quickly*. Before delving into the subject of German ideophones, a legitimate question to ask is whether those instances truly are ideophones if their root can be easily identified in other parts of speech?

To answer the above question, I will refer back to the definition of what an ideophone is – "[a] member of an open lexical class of marked words that depict sensory imagery" (Dingemanse, 2019, 16). An open lexical class entails that new members can be added to the class. Thus, new ideophones can be made up – whether collectively, idiosyncratically (as we will see below), or ad hoc (Maduka-Durunze, 2001). The argument of markedness fundamentally concerns spoken language. Nevertheless, markedness can be deduced from the typography in Figure 4.1. Prosody is reflected in the writing via punctuation (Chafe, 1988), or letter replication (Fuchs et al., 2019). The font itself can also be used to represent prosodic prominence by implementing boldface or variable size (Rosenberger, 1998). Therefore, when read, those typographically marked words would receive an appropriate iconic prominence marking, too. Finally, do these words depict sensory imagery? *Knurr* [knʊɐ̯] and *brumm* [bʁʊm] are mainly based on the perception of a certain sound. The latter is often used with children to name a sound made by a car or machinery. However, the former can be used to depict a sound, or maybe even more so a feeling, of an empty stomach, as in *Mein Magen knurrt* 'my stomach rumbles.' The sounds contained in *kratz* [kʁat͜s] imitate the sound of scratching, and those in *zack* [t͜sak] sound like something fast, thanks to a continuant in the onset affricate and a plosive in the coda. *Tatz* contains a plosive in the onset, therefore symbolizes something abrupt, with a coda similar to previously mentioned *kratz*, meaning 'to scratch' – something that a paw can do. That said, the words in Figure 4.1 do present examples of ideophones after the definition of Dingemanse (2019, 16).

4.2.1 The Origin of German Ideophones

However, the examples in Figure 4.1 call for a question about the origin of German ideophones. There exist a number of ideophones that seem to be systematically derived from verbs, like *kratz* from *kratzen* 'to scratch', *brumm* from *brummen* 'to buzz', and *knurr* from *knurren* 'to growl' (cf. Figure 4.1). This morphological form in German is a non-inflectional construction (NIC), which is created by stripping the verb to its stem (Bücking and Rau, 2013, 62). This type of construction can be used (1) for any German verb, and also (2) as a complex structure in combination with other words (Bücking and Rau, 2013, 59f.):

(1) *dich in den Arm nehm*
you.ACC in the.M.ACC.SG arm take.STEM

(2) *licht-wieder-an-knips*
light-again-on-turn.STEM

The origin of the German NIC is disputed, however it is certain that it was popularized by comic books and, subsequently, web speech (cf. Schlobinski 2001; Teuber 1998; after Bücking and Rau 2013, 61f.). The NIC can be found already in the original version of *Hänsel und Gretel* from 1812 (cf. Grimm and Grimm, 2013, KHM 15), which says[1]:

(3) *knuper, knuper, kneisch-en, wer knuper-t an mein-em*
nibble.STEM nibble.STEM gnaw-INF who nibble-3SG at my-N.DAT.SG
Häus-chen!
house-DIM.N
'nibble, nibble, gnaw, who's nibbling on my house!'

And also in *Max und Moritz*, a comic published by Wilhelm Busch in 1865. Here an example from the fourth story[2]:

(4) *Aber Moritz aus der Tasche zieh-t die*
but Moritz from the.F.DAT.SG pocket pull-3.SG the.F
Tinte-n-pulver-flasche, und geschwinde, stopf, stopf, stopf!
ink-INTF-powder-bottle and quickly stuff.STEM stuff.STEM stuff.STEM
Pulver in den Pfeife-n-kopf.
powder in the.M.ACC.SG pipe-INTF-head

[1] The form *knupern* refers to the modern German verb *knuspern* 'to nibble, to crunch.'

[2] The original drawings and text can be found here: `https://www.wilhelm-busch.de/werke/max-und-moritz/max-und-moritz-streich-4/blatt-3/` (accessed on June 10, 2021).

'But Moritz takes the ink bottle out of his pocket and quickly stuffs it into the pipe!'

The NIC were widely used by Erika Fuchs – a German translator of Disney comics from the early 1950s (Wormer, 2017). It may, thus, be the case that many German ideophones are derived from existing verbs through the NIC.

However, even though NIC can be based on any verb, those that are mainly used in comics seem to have a sound-imitating character. The question is whether those verbs originate from sound-imitative words, i.e., onomatopoeia. Only later would they be turned to NIC to be used as ideophones. The process of such origin would be as follows:

(5) original sound (inspires) → sound-imitative word (turns to) → verb (is stripped to) → non-inflectional construction (is used as) → ideophone

To answer the question, whether German NIC used as ideophones stem from verbs based on sound-imitative words, I examined a few instances from Figure 4.1 and similar. For the verb *brummen*, inspiring the ideophone *brumm* in Figure 4.1, the etymological dictionary of German (Pfeifer, 2021b) points out to the earliest form in Old High German *breman* "brüllen" 'to roar' from the 9th century and notes the similarity of this sound-imitating verb to other languages, establishing the Indoeuropean root **bhrem-* "brummen, summen, surren" (Pfeifer, 2021a), 'to roar, to hum, to buzz'. *Knurr* in Figure 4.1 stems from *knurren* meaning "drohende Kehllaute ausstoßen (vom Hund), kollernde Töne von sich geben (vom Magen), mürrisch, unfreundlich, gereizt und undeutlich sprechen" (Pfeifer, 2021c), 'uttering threatening guttural sounds (from dog), making rumbling sounds (from stomach), speaking sullenly, unfriendly, irritably, and slurred'. The author indicates that the verb *knurren*, first noted in the 16th century, has a sound-imitative origin, similar to verbs *knarren* and *knirschen*. The verb *kratzen*, connected to the ideophone *kratz* in Figure 4.1, means "mit einem rauhen, scharfen oder spitzen Gegenstand schaben, reiben, ritzen" (Pfeifer, 2021e), 'scraping, rubbing, scratching with a rough, sharp or pointed object'. It originates from Old High German *krazzōn* from the 9th century. In Middle High German, there existed both *kratzen* and *kretzen*, which leads Pfeifer (2021e) to believe that the original root was the Indoeuropean **gred-*. He, however, does not remark on the sound-imitating, nor on the sound-symbolic character. An interesting example outside the scope of Figure 4.1 is the verb *puffen*, which was first noted in the 16th century and described "schlagen, stoßen", 'to strike, to push', or "(durch stoßartig entweichende Luft) einen dumpfen Knall von sich geben" (Pfeifer, 2021f), 'emitting a dull bang (by air escaping impulsively)'. The verb originates from the Early New High German nouns *buf* (15th century), and *puff* (16th century) that Pfeifer (2021f) views as sound-imitating for "(...) dull noises and sound impressions, such as those produced by air escaping impulsively (...)".[3] Only later, in the 18th century, did the noun *Puff* change the meaning to 'brothel', however still implying the sound-imitating characteristic. Lastly, the verb *krachen* "laut und polternd schallen, tönen, lärmen" (Pfeifer, 2021d), 'to resound loud and rumbling, to sound, to make noise', dates back to Old High German *krahhen* around 800. The noun *Krach* "Lärm, Getöse, polternder Schlag" (Pfeifer, 2021d), 'noise, roar, rumbling beat', is a back-formation from the verb, with first occurences in the 10th century in Old High German as *krah* or *krac*. Pfeifer (2021d) mentions that *krachen* – alongside with *krächzen*, *krähen*, and *Kranich* – are various expansions of the Indoeuropean sound-imitating root **ger-* or **grā-* meaning 'to shout hoarsely'.

The above analysis of a few NIC-serving verbs, despite its limited scope, reveals that the etymological process suggested in (5) may, in fact, hold true for a substantial number of ideophones derived as NIC. As noted by Pfeifer (2021b), verbs that may be later re-interpreted as NIC often have a sound-symbolic character in their origin.

Any future research on the etymology of ideophones in German should, above all, take into account the *iconic treadmill hypothesis* by Flaksman (2017). Flaksman (2017) describes various stages of de-iconization of iconic words – a process that occurs due to morphophonological processes, such as sound change or morphological integration. Iconic words become conventionalized with time and, thus, lose their expressivity. Then again, the need for expressivity drives the creation of new iconic words (Flaksman, 2017, 33f.). Therefore, the process suggested in (5) might fulfill this need for expressivity by reanalyzing conventionalized words (i.e., verbs) as iconic words (i.e., NIC used as ideophones). In order to examine this hypothesis thoroughly, future studies should investigate a more extensive set of German ideophones and the role that non-inflectional constructions play for them.

4.2.2 Ideophones in Children's Books

Knowing the above examples from comic books, we searched for another ideophone-rich genre to build the data set. Children's literature revealed itself as an ideal source. On the one hand, the prime goal was to gather as many German ideophones as possible in a structured data set. On the other hand, we wanted to test the sensory scope of German ideophones (cf. Chapter 4.4) and the use of ideophones across ages (cf. Chapter 4.5). It has frequently been noted that sound symbolism, and most notably ideophones, facilitates word-learning in children (e.g., Imai et al., 2008, 2015; Kita et al., 2010; Laing, 2019), also across languages (Kantartzis et al., 2011; Yoshida, 2012). Caretakers use sound-symbolic words most frequently in early child development, and the number decreases with age (Perlman et al., 2017). This finding stresses the importance of sound symbolism at the onset of language acquisition. Laing (2019, 18f.) notes that onomatopoeias are ideal in early interactions due

[3]Translated by the author, original: "(...) dumpfe Geräusche und für Schalleindrücke, wie sie bei stoßartig entweichender Luft entstehen (...)" (Pfeifer, 2021f).

to their simplicity – meaning both the cross-linguistic similarity, as well as a phonological simplicity when a fine-phonological ability is still lacking. In an analysis of a language acquisition diary of Annalena – originally published by Elsen (1991) – Laing (2014) establishes that onomatopoeias largely contribute to the vocabulary of the child in the earliest stages. Through onomatopoeias, Annalena develops the phonetic abilities that she does not show elsewhere in the lexicon, e.g., she produces certain consonants only in onomatopeia, but not in proper words (Laing, 2014, 398 and 401). Laing says that onomatopoeias facilitate the phonetic and phonological development in that "their fuzzy phonological boundaries appear to facilitate production by providing an impressionistic template which incorporates lexical meaning while also allowing a wide margin of phonological error" (Laing, 2014, 403). Alongside these results are those by Perry et al. (2015) and the most recent ones by Sidhu et al. (2021a), showing that early acquired words are higher in iconicity than later acquired words.

Adult authors wrote the books used for our data set collection, thus resembling a caretaker–child relationship. In a caretaker–child situation, those books are read to children, which should facilitate language acquisition. Thus, one of the goals of collecting data from children's books is to investigate whether the use of ideophones in these books changes across ages, i.e., whether their use in child-directed and child-produced speech is mirrored in the written child-directed language.

Children's books are written for young audiences of various ages, and the age recommendation is an available data point. It all allows for both gathering an extensive data set and analyzing the use of ideophones across different age groups. One reason to assume that caretakers use ideophones more frequently towards younger children might be the depictive character of sound-symbolic words allowing for a broad sensory experience (Motamedi et al., 2020). Therefore, before examining the use of ideophones in books for different age groups, it was relevant to test which sensory domains native speakers of German connect with some of the ideophones. As suggested by Nuckolls (2019), the sensori-semantic categories of ideophone overlap; hence, an assignment to a single sensory domain may not be factual. We based the categories in those previously established in an implicational hierarchy by Dingemanse (2012) and tested how unanimously do native speakers of German assign the ideophones to the categories.

Next, I resume to present the data collection process and define the criteria applied to assess the ideophone character of a word. Assembling the data set allows for further investigations. An immediate step following the collection of this kind of a large data set is its structuring in the form of a dictionary proposal, an initial attempt of which is offered in Chapter 4.3.1. It is an example of a theoretical work, yet crucial given that a proper ideophone dictionary for German is lacking. My work in this thesis is strongly experimental; therefore, I apply the data set to answer compelling research questions experimentally. As noted in the paragraph above, those are tightly connected to one another. The first investigates the implicational hierarchy of ideophones while at the same time tackling the question of the dispersion of the ideophone meaning across sensory domains (cf. Chapter 4.4). The second question investigates the use of ideophones in books for different age groups (cf. Chapter 4.5).

4.2.3 Methods of the Data Collection

The data set was collected from February to August 2018 in the Helene-Nathan-Bibliothek in Berlin[4]. This library was chosen due to its rich selection of children's books. The criteria for the data collection are attributable to the key elements of what an ideophone is, as discussed above, and based on the theory by Dingemanse (2019, 16). Words with sound-symbolic ideophone-like features were not treated as ideophones if they showed morphosyntactic integration (Dingemanse, 2012, 656). In that sense, verbs that remind of ideophones, such as *krächzen* from *krächz* 'croak' or *piepen* from *piep* 'squek,' were not included in the data set (cf. also Chapter 4.2). Similarly, nouns that bear signs of declension were excluded from the data set. As shown by Dingemanse (2017a) and Dingemanse and Akita (2016), morphosyntactic integration and the expressive character of ideophones are inversely correlated. Hence, only nouns that were not morphosyntactically integrated, i.e., not declined as well as marked, were incorporated in the data set. Each instance of the ideophone was entered into the data set. Also, the original form of the ideophones was kept, e.g., *ahuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu*, or *huuuaaaaaaaaaah*, in order to allow for an analysis of prosodic amplification in the future (similar to, e.g., Fuchs et al., 2019, for adjective antonym pairs). However, a unified written form was noted beside the original one, e.g., *ahu*, or *huah* for the above, which allows for comparing the number of word forms (original forms as seen in the book) with the number of lemmas (unified forms as would be seen in a dictionary).

Next, we collected metadata on the book, in which a given ideophone was found, such as the title, the author, the age recommendation, the original language, and the publishing year. We hope that this metadata might allow for testing other effects in the future, for example, whether the language of origin has an influence on the variability and the number of ideophones in the book (original language), or whether the use of ideophones in children's books changes over time (publishing year).

The data set also entails the context of the ideophone and its meaning in context, if possible to establish. The interpretation of the meaning in context is subjective and might be a case for an in-depth investigation in the future. A photo was taken of the page on which the ideophone was found to ease future work with ideophones in context. In this respect, photos are a part of the data set. The unique name of each photo for a given page, on which the ideophone was found, is noted in the data set. Two other columns are the part of speech and category, yet, these are by

[4]The library is located in Karl-Marx-Str. 66, 12043 Berlin, Germany.

far the least developed parts of the data set so far. An idea for future research might be to comprehensively structure ideophones for a German ideophone dictionary by parts of speech, but also across the categories in which ideophones may fall, such as sound, movement, or texture ideophones (cf. Dingemanse 2012, 663; McLean 2020, 18; and Chapter 4.4).

Lastly, the data set comprises the information on the number of words on a given page and the number of pages in a given book. The former was extracted semi-automatically using picture-to-text recognition available in Microsoft OneNote (Microsoft Corporation, 2018). In the cases that the text was not recognized correctly, the number of words was counted manually. This situation occurred, e.g., when the picture quality was worse and when the text was written by hand or not horizontally from left to right. A word was treated as an orthographic unit between two blank spaces with a few exceptions. In some cases, ideophones were concatenated into one word, e.g., *oh je*, *ach herrje*, or *tra, ri, ra*. Those examples concern ideophones, which often co-occur and are officially written separately which, however, tend to be concatenated in informal writing – compare *oh je* with 19,959 vs. *ohje* with 15,624 occurrences, as given in the WebXL corpus of the DWDS (*DWDS – Digitales Wörterbuch der deutschen Sprache* 2021, accessed on February 5, 2021). All words were counted, even those written on the graphic elements and which were merely elements of the story; see Figure 4.2 for an example. In some cases, the pictures did not comprise the whole page, i.e., parts of the text were cut out. Then, an estimation of the word count, based on the visible parts and other available pages, was given. The number of pages was copied from the library database. The number of words on a page, along with the age recommendation data, allows for testing the hypothesis, whether there are more ideophones in books for younger audiences (cf. Chapter 4.5).

After completing the data collection, we established that the criteria we applied to assert the ideophone status might be problematic for German. We heavily relied on the universality of the already-cited definition – "[a] member of an open lexical class of marked words that depict sensory imagery" (Dingemanse, 2019, 16) – and based our criteria upon that. Most importantly, the word should be marked (syntactically, phonologically, or prosodically) and depict some kind of a sensory representation or experience. We also payed attention not to include morphosyntactically integrated cases (Dingemanse, 2017a; Dingemanse and Akita, 2016). Using these criteria, we now notice two sub-categories of ideophones in the data set. The first encompasses what we call *expressive* ideophones, with instances like *oh*, *ach*, *oje*, or *autsch*. The second sub-category includes what we call *depictive* ideophones, such as *dingdong*, *ruckzuck*, *schwups*, or *krach*. Now, the first category ideophones overlap with interjections according to Dingemanse (2021, 2): "Ideophones are typically depictions *of* events, while interjections are typically responses *to* them. Perhaps a slap in the face will help the reader to appreciate the difference. The *sound* of the slap is the main business of an ideophone to depict; your *outcry* in response to it is an interjection."

FIGURE 4.2: The figure illustrates an example of a page on which an ideophone occurs – *Au!* at the top of the page. Other than that, there are many words incorporated in the image. All of those words were counted into the data on words per page. The figure is a part of the whole page from the book *Pino Pfote - Ab die Post!* by Freeman (2016).

However true the example by Dingemanse (2021, 2) is, the dictionary of Polish onomatopoeia (Bańko, 2008) also includes items that would be classified as interjections; cf. the entry for *au* (other spellings, e.g., *aua*, *auć*, or *jau*), an equivalent of the German *autsch*:

> Names that imitate an animal howling, loud, prolonged crying, or loud yawning. Also names for the sound made involuntarily by someone who is suddenly in pain. It is difficult to separate them precisely, although it can be seen that some imitate a prolonged sound (e.g., *auu*), others a short sound (e.g., *auć*), and others one or the other depending on the word stress, cf. a**u̲** (howling) and **a̲**u (pain symptom)

(Bańko, 2008, 145).[5] It is worth noting that, in Polish, word stress in not semantically relevant, Bańko notes, however, the semantic difference between a**u̲** and **a̲**u, pointing to the phonological markedness of ideophones (e.g., Dingemanse, 2021, 1). Dingemanse (2021, 2) also notes that interjections, like the above pain outcry, are indexical, whereas ideophones are iconic in the Peircean sense (Peirce, 1955). Nevertheless, it is also possible to use these interjections, or what we call here *expressive* ideophones, in a *depictive*, i.e., iconic, manner. Imagine watching someone being slapped. The

[5]Translated by the author, original text: "Nazwy imitujące wycie zwierzęcia, głośny, przeciągły płacz lub głośne ziewanie. Także nazwy dźwięku, jaki wydaje mimowolnie ktoś, kogo nagle coś zabolało. Trudno je precezyjne rozdzielić, choć widać, że niektóre imitują przeciągły dźwięk (np. *auu*), inne krótki dźwięk (np. *auć*), a jeszcze inne jeden lub drugi zależnie od akcentu por. a**u̲** (wycie) i **a̲**u (objaw bólu)." (Bańko, 2008, 145)

sympathetic reaction to such an event would rather involve the vocalization *autsch* 'ow!' than the vocalization *klatsch* 'slap'. In this case, *autsch* 'ow!' is depictive of the pain experience. As shown by Singer et al. (2004), seeing someone else in pain only triggers an affective, not the sensory, response in the brain. Therefore, in this context, *autsch* 'ow!' can be interpreted as a demonstration of an internal state cause by the image of someone else's pain. Similarly, *aua* 'ow!' can be used expressively as a reaction to a pain-inducing stimulus, but also in a depictive sense, e.g., *Hast du ein Aua?* 'Do you have an ow?', where it is even nominalized.

Regardless of the above interpretation, the reader must be informed that our data set might be inaccurate in relation to Dingemanse (2021), who states that "[i]f we lump them [ideophones and interjections] together, we lose the ability to explain how and why they differ in terms of markedness, morphosyntax and mode of signification." Consideration on the ideophone-status of some interjections, such as *autsch* or *aua* 'ow!', would be a valuable addition for future research on German ideophones, or even ideophones in general. Here, however, I proceed with outlining the data and subsequent research questions as they were planned and conducted before the latest insights (notably Dingemanse, 2021) came to my attention.

4.2.4 Results of the Data Collection

The data was collected from 431 books, 270 of which were originally written in German and 161 translated from foreign languages. Among the translated books, 88 were originally English, 25 French, and 16 Dutch. There is a total of 3,631 data points. If only one occurence of an ideophone per page(!) is counted (i.e., disregarding ideophone repetitions on the same page) there are 3,006 data points. If only one occurence of an ideophone per book(!) is counted (i.e., disregarding ideophone repetitions within a book), there are 2,479 data points. From among all data points which include repetitions, there are 1,020 word forms (so where *aa* vs. *aaaa* are separate instances) and a total of 650 lemmas, i.e., unified word forms (so where *aa* vs. *aaaa* are uniformly written and count as one instance). In order to put this number into perspective, I provide a few examples. Flaksman (2017) found about 1,500 iconic roots in the English lexicon, 60% of which are de-iconized by now. Taylor (2007) lists 1,200 instances of imitative utterances in English comic books. Lastly, Havlik (1981), provides about a list of 2,000 onomatopoeias used in German comics (comparable to word forms due to variable writing, e.g., *auweh*, *au wei*, and *au weia* are treated separately).

The age recommendation was reliably divided into two groups: children up to 6 years of age and children from 7 to 9 years old. There are 342 books in the first age group (up to 6 years) and 89 in the second age group (7–9 years). As for the ideophones themselves, Table 4.1 summarizes the most frequently found ideophone lemmas. To identify the lemmas correctly, some prosodic phenomena in writing, like lengthening or reduplication, were minimized and unified.

TABLE 4.1: The table shows the most frequent ideophones found in the corpus. The frequency was calculated across all data points. It is based on lemmas, i.e., the lengthening and reduplication are minimized and unified. Also, some instances are concatenated (cf. Chapter 4.2.3).

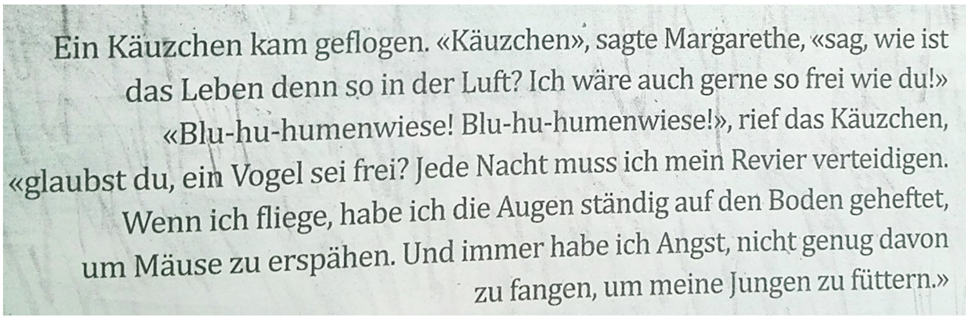
Ideophone	**Frequency**
na	228
oh	187
ach	145
hm	71
oje	65
tak	58
he	53
ah	51
platsch	50
ohnein	44
äh	43
pst	41
nagut	36
hui	34
puh	34
naja	33
bäh	32
aha	31
hopp	31
hurra	29
haha	28
nanu	28
au	26
miau	26
piep	24
aua	23
huhu	22
achso	21
hicks	21
mäh	21
auweia	20
hä	20
ups	20

At this point, I would like to mention a few noteworthy cases, which became apparent during the data examination. Firstly, as has been observed before, some ideophones often occur together with other words, like particles, e.g., *na gut*, *na ja* (or: *naja*), *na klar*, *oh ja* (or: *o ja*), *oh nein* (or: *o nein*), *oh nee* (or: *o nee*). The question that arises here, is whether those ideophones should be treated separately to the particles they co-occur with. Just like with the previously mentioned case of *oh je* and *ohje*, – whose frequency of occurrence in the WebXL corpus of the DWDS (*DWDS – Digitales Wörterbuch der deutschen Sprache* 2021) is comparable – it is similar for some of the words given here as examples. The most striking evidence for treating the ideophone–particle combination as one, is the case of *naja*. Written together, it occurs in the DWDS WebXL corpus 584,120 times, but written separately as *na ja* 134,420 times (*DWDS – Digitales Wörterbuch der deutschen Sprache* 2021, accessed on February 5, 2021). An exhaustive discussion on this notion is, however, outside the scope of this work. Future research might resolve this issue by examining both word forms and lemmas available in the corpus. Also, as mentioned in Chapter 4.2, a morphosyntactic analysis of the data set might also tackle many instances of verbs with a sound-symbolic character, such as: *kreischen* 'to squel', *poltern* 'to crash', *jammern* 'to moan', *seufzen* 'to sigh', *fiepen* 'to whimper', *blitzen* 'to flash', *sausen* 'to whistle', or *wippen* 'to bob'. An in-depth analysis of those instances, alongside with the NIC based on those verbs – i.e., *kreisch*, *polter*, *seufz*, and so on – could reveal whether such verbs are in fact based in sound imitation.

Furthermore, some cases occur where ideophones are embedded in other words, as can be seen in Figure 4.3. In all three examples given, the authors use different strategies to embed sound-symbolic elements in the word, but those cases also demonstrate different goals such an embedding might have. In Figure 4.3a, the ideophone *huhu* is embedded in the word *Blumenwiese*, 'a flowery meadow'. Here, *huhu* is the sound made by the animal saying the words, a tawny owl. The author uses the long stressed vowel in the initial syllable to embed the animal sound – [ˈbluːmənˌviːzə] becomes [ˌbluːhuˈhuːmənviːzə].[6] Figure 4.3b, shows a different case. Here, the sound-symbolic repetition is used to create a rhythmic pattern of the phrase, possibly also as an emphasis. *Flo-Fla-Flieger-Held* has three radically different vowels and, thanks to the orthographic separation with a dash and the minimal segmental contrast, all three initial syllables may be realized as prosodically prominent, such as in [ˈfloː ˈfla ˈfliːgəhɛlt]. The last example, in Figure 4.3c, shows how the profession of the character may be combined with an ideophone. In this case, it is not a simple *au*, an expression of pain, but a *Piraten-Au*, an expression of pain by a pirate.

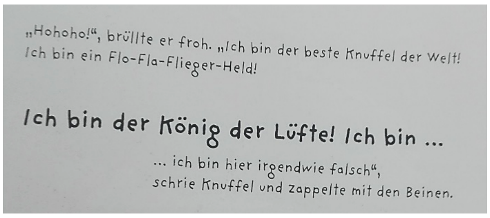
The cases mentioned above serve as a preface to the idiosyncrasy of ideophones in children's literature. The most frequent ideophones are given in Table 4.1; however, there are 307 ideophone lemmas (47% of a total of 650 lemmas) that only occur once in the whole corpus. This finding motivated a comparison of the ideophone

[6]The stress pattern here is only a proposal.



Ein Käuzchen kam geflogen. «Käuzchen», sagte Margarethe, «sag, wie ist das Leben denn so in der Luft? Ich wäre auch gerne so frei wie du!» «Blu-hu-humenwiese! Blu-hu-humenwiese!», rief das Käuzchen, «glaubst du, ein Vogel sei frei? Jede Nacht muss ich mein Revier verteidigen. Wenn ich fliege, habe ich die Augen ständig auf den Boden geheftet, um Mäuse zu erspähen. Und immer habe ich Angst, nicht genug davon zu fangen, um meine Jungen zu füttern.»

(A) "Blu-hu-humenwiese! Blu-hu-humenwiese!" (Schlatter, 2019).

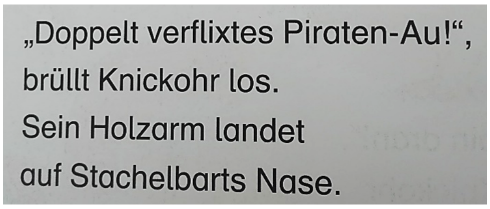


„Hohoho!", brüllte er froh. „Ich bin der beste Knuffel der Welt! Ich bin ein Flo-Fla-Flieger-Held!

Ich bin der König der Lüfte! Ich bin …

… ich bin hier irgendwie falsch", schrie Knuffel und zappelte mit den Beinen.

(B) "Flo-Fla-Flieger-Held!" (Witzigmann and Bjarke, 2010).



„Doppelt verflixtes Piraten-Au!", brüllt Knickohr los. Sein Holzarm landet auf Stachelbarts Nase.

(C) "Piraten-Au!" (Koenig, 2011).

FIGURE 4.3: Examples of ideophones embedded in other words.

distribution according to Zipf's law. Zipf's law is a power law, prevalent in languages of the world (Calude and Pagel, 2011; Piantadosi, 2014) but also in other phenomena, like city populations or company sizes (cf. Li, 2002, 16ff.). For natural languages, it states that the frequency of the word is inversely proportionate to its rank, i.e., the most frequent word has the rank 1, the second most frequent word the rank 2, and so on. Figure 4.4 illustrates how the collected data applies to Zipf's law. As can be seen, the observed distribution in blue varies slightly from the theoretical distribution in red. First of all, it can be seen that the proportion between the first two most frequent words is disturbed – the first most frequent word is less frequent than expected, and the second most frequent word is more frequent than expected. Mid-frequent words (with a log frequency rank of approx. 2.5–4) occur more frequently than expected. This all may be caused by the high proportion of words that occur only once in the corpus. Those instances of *hapax legomenon* – i.e., unique instances – are visible on the lower blue points of the plot that extend to vertical lines. The data set also does not quite meet the criteria of another power law, the Pareto principle, also known as the 80/20 rule. In our data set, 20% of the top-ranked words ($N = 130$) account for 73% of use (2651 out of the total of 3631 occurrences), while a number closer to 80% should be expected.

The unique instances of ideophones signify a high degree of idiosyncratic symbolic forms in children's books. Some of them are single items, like *krrks*, *tsurum*, or *pottepenk*. Some of them involve reduplications, like *ohwehohwehohweh*, reduplications with onset variation, like *Schnuffel-tuffel-muffel* (for more details on this kind of reduplication, see Kentner, 2017), or reduplications with ideophone combination, like *schwubbelschwabbelschwischwaschworen*. Ideophone combinations are not

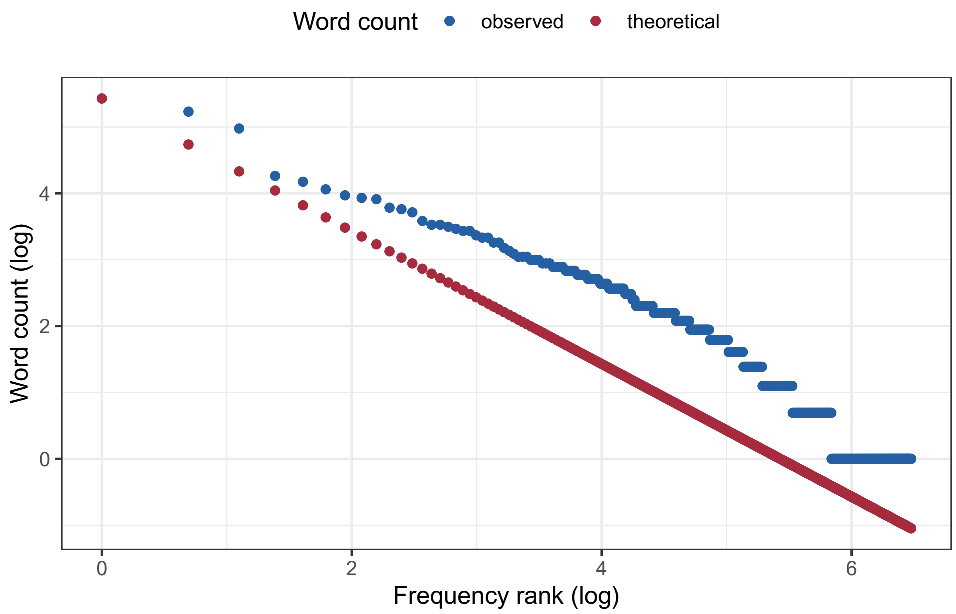


FIGURE 4.4: Zipf's law of the collected ideophones with word count on the y-axis and frequency rank on the x-axis. Both variables are presented in the logarithmic scale. The observed data in blue diverges from Zipf's theoretical distribution in red. The lower blue points show that numerous unique instances occur only once.

uncommon, too, and include among the idiosyncratic ideophones, e.g., *blatzpadatz*, or *bumplatschbong*.

An example of ideophone co-occurrence among the more common ideophones is given in Figure 4.5. The exact sequence there gives the impression of the event – the two characters crash with *bumms*, the umbrella hits the crocodile on the head with *knall*, and the crocodile falls on the floor with *rumms*. Also, not only the idiosyncratic ideophones undergo reduplication, but the more common ones, too. E.g., *igitt-igitt-igitt*, *blubb! blubb!*, *zackzackpengpeng*, the latter example being both a reduplication and an ideophone combination.

Having had a glimpse into the richness of the collected data, I will now propose how the data could be structured for further publication and research. Because the data set involves a collection of words with their context and possible meanings, I will outline a potential design of a lexicon of German ideophones.

4.3 Lexicon of German Ideophones

The data set collected as a part of this research project is a rich source for further analyses. As such, an effort should be made to make the data available to the interested public. The proposal made here is, due to the language of the whole publication, conceived in English. Nevertheless, a complete lexicon should be written

FIGURE 4.5: An example of ideophones occurring together from Kulot (2016). The sequence of the ideophones illustrates the event of being hit and falling down.

in German, as the primary target group would be German readers. Another argument for that is the nature of ideophones, which in their sound symbolism may be hard or even impossible to translate (cf. Chapter 2.5). Among the addressees of such a collection are primarily researchers and students of linguistics or German philology. However, also the general public might appreciate the publication of the first German ideophone lexicon. Like dictionaries in general, it might also help learners of German to better comprehend everyday language that they might not find elsewhere.

However, not only the collected data set should be the basis of the lexicon, but further sources, too. This expansion is essential, as children's books do not entail some ideophones commonly used in daily speech, e.g., *futsch* 'gone', a word commonly used even in newspapers (e.g., an article *Bezirksamt legt zwei Meter breiten Radweg an — 300 Parkplätze futsch* 'District office creates a two-meter-wide cycle path – 300 parking spaces gone' published by Hildburg Bruns in Berliner Zeitung on February 4, 2021[7]). Other common ideophones not present in the corpus include, e.g., *ballaballa* 'crazy', *flugs* 'fast', *Hallodri* 'an unreliable man', *Tohuwabohu* 'a chaos', *Rambazamba* 'a fuss', *ratzfatz* 'fast', or *Remmidemmi* 'a big hustle and bustle'.

On the other hand, not all instances present in the corpus may be useful for the lexicon. Hereby I mean especially the idiosyncratic ideophones discussed above and frequent cases of hapax legomenon – i.e., unique occurrences of words. Their exclusion from the lexicon does not diminish their expressiveness. Nevertheless, the lexicon should, first of all, include words that are generally used in everyday language, and it would not be the case with the ad-hoc or idiosyncratic ideophones. Although, an extensive introduction in the nature of ideophones should accompany the lexicon, including a description of their possible idiosyncrasy.

The introduction to the lexicon should also discuss the constraints of ideophones, mostly the sound-imitating ones, such as the phonological system and phonotactic rules. Those rules govern which sounds are given and permitted in a language and how they can be composed into syllables and words. Moreover, the phonological image of the word is represented graphemically depending on the grapheme–phoneme correspondence of the individual language. In turn, the phonetic, real-world sound of a dog's bark, becomes filtered into the German *wau* [va͜ʊ], the English *woof* [wuːf], the Polish *hau* [xaw], and the Japanese ワン *wan* [ɰᵝã̠ɴ].

Lastly, the data collected so far would certainly profit from further annotations. First and foremost, an annotation of the syllable structure of the ideophone. Such information would allow for an in-depth analysis of the ideophone phonotactics. As found by Nuckolls et al. (2016), ideophones in Pastaza Quechua employ mechanisms that are barely found in other parts of the lexicon. Among the sounds unusual for Pastaza Quechua that are, however, frequent in ideophones are for example palatalized obstruents /kʲ, pʲ, tʲ/, the affricate /ts/, or the vowel /o/. Some new sounds also developed in ideophones only, such as palatalized voiced obstruents /bʲ, gʲ/,

[7] The article can be found here; last accessed on March 9, 2021.

or expressively aspirated /p^h, t^h, k^h/. Moreover, ideophones in Pastaza Quechua use stress patterns and syllable structure not present anywhere else in the lexicon. Phonological markedness of ideophones has been noted elsewhere for, e.g., Dagaare (Bodomo, 2006), Yir-Yoront (Alpher, 1994), and Kisi (Childs, 1988). It is still an open question, how the phonology and phonotactics of ideophones look like in languages where ideophone inventories seem less-pronounced at first sight. Further analyses on that matter would allow establishing the characteristics of sounds that depict a certain sensory experience, such as abruptness of movement, or a structure of a visual image. Similar to sound symbolism, sounds with particular acoustic traits may trigger the denotation of the sensory experience in ideophones.

4.3.1 German Ideophone Dictionary: Design Proposal

The proposal created based on the data set and external research on German ideophones can serve as a basis of a first to date dictionary of German ideophones. As previously noted, it is crucial to include those ideophones, which native speakers commonly use. Therefore, the unique instances, i.e., hapax legomenon, from the collected data set should not be incorporated in a dictionary unless they find their way to the common language. The data set gives an insight primarily into the possibilities of an open lexical class that ideophones form (Dingemanse, 2019, 15). But because they are present across different word classes in German, their primary word class, together with the definition and an example of use, are crucial to disentangle the meaning for native speakers and language learners.

For German, a similar collection was previously published by Havlik (1981) for onomatopoeia found in comic books, and for Polish, an onomatopoeia dictionary, however expanding across other sensory modalities, was written by Bańko (2008). For languages which are known for their ideophone inventory, such collections, alongside exhaustive descriptions, are more common; a few examples include Nuckolls (1996) for Pastaza Quechua, Dingemanse (2011) for Siwu, or Reiter (2011) for Awetí.

A creation of a complete dictionary is outside the scope of this thesis. However, I would like to present a proposed structure for a compact dictionary of German ideophones. The following example consists of 30 entries of ideophones, which are also used in the subsequent analysis. The number of ideophones chosen for this example is arbitrary and should only serve as a diverse example containing ideophones with variable spelling, different meanings, and of diverse word classes. The items were chosen on the basis of the collected data set, however, taking into account their presence in everyday language. This way, idiosyncratic cases were ommitted. Every entry presented below comprises of six blocks. The first block is the ideophone itself (with alternative spelling, if necessary), followed by the IPA transcription in the second block. Further, the part of speech is given, alternatively, different possible parts of speech known in the use. This is necessary to distinguish, because, as noted before, German ideophones are scattered across the lexicon and do not belong to one part of speech only. The next block consists of the type of the ideophone (as revealed in the following Chapter 4.4, based on Dingemanse, 2012). This is followed by the meaning of the ideophone and, finally, by an example of its use. All examples stem from the corpus.

ach /ax/
interjection • inner feelings and cognitive states
1. positive exclamation: excitement, surprise
2. negative exclamation: indignation, despair
Example:
1. Ach, wie ist das Leben schön
Eng. Ach, how nice life is
2. Ach du meine Güte!
Eng. Ach my goodness

aha /aˈha/
interjection • inner feelings and cognitive states
exclamation of understanding, insight
Example:
"Aha", sagte Großmutter. "Du warst also die Babysitterin (...)"
Eng. "Aha", said grandmother. "So you were the babysitter (...)"

aua /ˈa͜ua/
interjection • other sensory perceptions
exclamation of pain
Example:
Aua, das tut weh!
Eng. Aua, it hurts!

auweia /a͜uˈva͜ɪa/
interjection • inner feelings and cognitive states
exclamation of worry, despair
Example:
Auweia, das sieht nicht gut aus!
Eng. Auweia, it's not looking good!

bums /bʊms/
interjection • sound
sound of an impact
Example:
und bums! liegt sie auf dem Boden
Eng. and bums! she's laying on the floor

dingdong | ding-dong /dɪŋdɔŋ/
interjection • sound

sound of a bell
Example:
Ding-dong, klingelte der Wolf am Gartentor
Eng. Ding-dong, the wolf rang at the garden gate

hä /hɛ/
interjection • cognitive state
exclamation of confusion, despair
Example:
Hä? Wo willst du denn hin?
Eng. Hä? Where are you going?

hm | hmm /hm̩/
interjection • inner feelings and cognitive states
1. exclamation of thinking (with a falling tone)
2. exclamation of delight (with a high or raising-falling tone)
Example:
1. Hmm, ich sehe das Problem.
Eng. Hmm, I see the problem.
2. Hmm, lecker!
Eng. Hmm, tasty!

hoppla /ˈhɔpla/
interjection • inner feelings and cognitive states
exclamation of surprise
Example:
Hoppla! Mama tritt auf die Bremse.
Eng. Hoppla! Mama hits the brakes.

hurra /ˈhʊʁa/
interjection • inner feelings and cognitive states
exclamation of excitement
Example:
Hurra, gleich ist Ida da!
Eng. Hurra, Ida will be here soon!

husch /hʊʃ/
interjection • movement
1. exclamation of chasing someone or something away
2. creeping, silent movement
Example:
1. "Husch, husch, hau bloß ab!"
Eng. "Husch, husch, just go away!"
2. wie er leise, husch, husch, husch, schleicht durch Wiese, Feld und Busch
Eng. how he quietly, husch, husch, husch, creeps through meadow, field, and bush

klackerdiklack /ˌklakɐdiːˈklak/
interjection • sound
clicking sound of a machine
Example:
Klackerdiklack (above hands typing on a machine)

Krimskrams | Krimsenkrams, ***m*** /ˈkʁɪmsˌkʁams/ | /ˈkʁɪmsənˌkʁams/
noun • visual
things, stuff
Example:
Dann noch Krimsenkrams für die Haare
Eng. Then also Krimsenkrams for the hair

Kuddelmuddel, ***m*** **|** ***n*** /ˈkʊdl̩ˌmʊdl̩/
noun • inner feelings and cognitive states
mess, confusion
Example:
Das große Jahreszeiten-Kuddelmuddel
Eng. The big season-Kuddelmuddel

Mischmasch, ***m*** /ˈmɪʃmaʃ/
noun • visual
a mix of everything, hotchpotch
Example:
Mischmaschzeit.
Eng. Mischmasch-time.

naja | na ja /naˈjaː/
interjection • inner feelings and cognitive states
exclamation of conceding, admitting
Example:
"Na jaaa", meinte Hagen schließlich. "Eigentlich haben die Jungs recht."
Eng. "Na jaaa", Hagen said finally. "Actually the boys are right."

nanu /naˈnuː/
interjection • inner feelings and cognitive states
1. exclamation of surprise, amazement
2. exclamation of confusion
Example:
1. Nanu! Plötzlich kommt das weiße Pony angetrabt!
Eng. Nanu! Suddenly the white pony comes trotting!
2. Anton scheint mich kaum zu bemerken. Nanu?
Eng. Anton hardly seems to notice me. Nanu?

oh /oː/
interjection • inner feelings and cognitive states

general accentuation of the context, e.g.,
1. exclamation of surprise (both positive and negative)
2. exclamation of despair
3. exclamation of joy
Example:
1. positive: *Oh! Ernest als Baby!*, *Eng.* Oh, Ernest as a baby!; negative: *Oh, Mist!*, *Eng.* Oh, crap!
2. Oh, ich weine täglich
Eng. Oh, I cry every day
3. Oh, wie toll, wie toll
Eng. Oh, how great, how great

oje | ohje /oˈjeː/
interjection • inner feelings and cognitive states
exclamation of worry, despair
Example:
Oje, das passt nie und nimmer alles ins Auto
Eng. Oje, all of this will never fit into the car

papperlapapp | Papperlapapp, ***m*** /ˌpapɐlaˈpap/
interjection | *Noun* • inner feelings and cognitive states
condescending on speech that is deemed worthless; nonsense, rubbish
Example:
Papperlapapp! Ich bin schon mit der mächtigen Hexe Priska durch die Gegend geflogen
Eng. Papperlapapp! I've already flown around with the mighty witch Priska

pardauz /paʁˈda͜ut͜s/
interjection • sound
sound of an impact
Example:
Das Päckchen fällt heraus. Pardauz!
Eng. The package falls out. Pardauz!

pfui /p͜fʊ͜ɪ/
interjection • other sensory perception
expression of disgust
Example:
Pfui, ist wohl gar nicht lecker.
Eng. Pfui, it's not tasty at all.

platsch /plat͜ʃ/
interjection • sound
sound of an impact with water
Example:
Und dann – platsch! – landet er im Teich.

Eng. And then – platsch! – he ends up in the pond.

potzblitz | potz Blitz *(obsolete)* /ˌpɔt͡sˈblɪt͡s/
interjection • inner feelings and cognitive states
exclamation of surprise; cursing
Example:
Potz Blitz, ich hab doch tatsächlich verschlafen!
Eng. Potz Blitz, I actually overslept!

pst | psst /ps̩t/
interjection • sound
silencing expression
Example:
Psst, sei still!
Eng. Psst, be quiet!

ratsch /ʁat͡ʃ/
interjection • sound
tearing sound
Example:
Plötzlich macht es RATSCH, und das Trampolin ist kaputt.
Eng. Suddenly it goes RATSCH, and the trampoline is broken.

ruckzuck | ruck zuck /ˌrʊkˈt͡sʊk/
adverb • movement
fast
Example:
Wenn ihr uns leuchtet, finden wir mein Kissen ruckzuck!
Eng. If you light us up, we'll find my pillow ruckzuck!

schwups | schwupps /ʃvʊps/
interjection • movement
expression of a fast movement
Example:
Schwups! Schon steckt es im Beutel.
Eng. Schwups! It's already in the bag.

ups /ʊps/
interjection • inner feelings and cognitive states
expression of (an unpleasant) surprise, a failure
Example:
Ups. Jetzt habe ich das Geheimnis verraten
Eng. Ups. Now I've revealed the secret

Zickzack | Zick-Zack, *m* /ˈt͡sɪkˌt͡sak/
noun • visual
a recurring pattern from one side to another

Example:
Im Zickzack und so wild, dass sie dabei fast eine alte Dame umgehüpft hätte (...)
Eng. In Zickzack and so wild that she almost knocked over an old lady (...)

4.3.2 Repository

The underlying data of the corpus is available in the OSF repository under the following address: `https://osf.io/6udxz/`. It includes the CSV file with the main corpus, the R script for the descriptive analyses (partly described in Chapter 4.5), the table with the list of books, and the table with the frequency of the lemmas. As mentioned earlier, the pictures of the ideophones in children's books are an essential part of the corpus. They are, however, copyrighted and will, therefore, not be made available in the repository. They can be made accessible upon request and quoted only accompanied by a proper reference.

4.4 Classification of German Ideophones

So far, we have been talking about the data set of German ideophones that we managed to collect through children's books. In the following, I would like to present the first experimental analysis conducted using the collected data. We have (a lot of) German ideophones at hand, and now we want to see *how* they are.

As Kilian-Hatz (1999, after Dingemanse, 2012, 662) notices, there are differences between languages when it comes to the semantic breadth of ideophones. While sound-imitating ideophones seem to be widespread, other domains are less frequent. Thus far, there exists little work on the semantic typology of ideophones. One notable example is the implicational hierarchy proposed by Dingemanse (2012). It serves primarily as a guideline to structure the semantic categories of ideophones in a given language, but it also allows to make predictions on how the ideophone inventory of a language may develop.

In 2012, Dingemanse proposed that the ideophone inventory within a language develops along an implicational hierarchy. The hierarchy follows the pattern (Dingemanse, 2012, 663):
"SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES".
It implies that if there are ideophones in a language, they will first belong to the sound category – these could be, for example, onomatopoeia. Therefore, if there are ideophones for movement in a language, some sound ideophones have to exist there, too. Further, the existence of visual ideophones in a language implies that there are as well some ideophones for sound and movement. The hierarchy rules out the possibility of finding a language with ideophones for other sensory perceptions if there exist no sound, movement, and visual patterns ideophones in this language.

Dingemanse (2012, 662ff.) does not specify the number of ideophones that must belong to the category, i.e., that the primary categories must be the richest ones, but simply states the implicational order of the ideophone inventory.

In a recent stimulus-based elicitation study, McLean (2020) found that two further categories are prevalent in Japanese ideophones – FORM and TEXTURE. This adaptation seems to go beyond the scope of Japonic languages. An investigation on ideophone inventories of 22 languages shows that the revised hierarchy holds true for a multitude of language families. It also suggests that the two novel categories should be included after SOUND and MOVEMENT in the original hierarchy by Dingemanse (2012). The revised hierarchy looks as follows:
"SOUND < MOVEMENT < FORM < TEXTURE < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER FEELINGS AND COGNITIVE STATES"

However, it has to be kept in mind that a classification of ideophones or other sound-symbolic words into a single sensory domain is questionable (Akita, 2009; Nuckolls, 2019). Perception is generally multisensory (cf. Calvert et al., 2004; Ghazanfar and Schroeder, 2006) and a percept of seemingly one domain may implicate a multitude of other information (Nuckolls, 2019, 171). Similarly, when thinking of semantic features to decompose the meaning of a lexical item, those features are based on various sensory domains. We cannot say that coffee that we drink in the morning is based on taste only. It is also its profound smell, color, and even the tactile feeling of the cup and the fluid itself wandering through the mouth that create its sense. All of the sensory cues are integrated into the concept (cf. Spence, 2018), and the integration occurs even beyond the neocortex (Ghazanfar and Schroeder, 2006, 282). As studies show, even sensory, non-ideophone, words, like 'barbecued' or 'fishy', are highly multisensory (Winter, 2019, 141; also see Lynott and Connell, 2009). In Winter (2019, 143), 'harsh' is classified across all sensory domains, most strongly connected to the auditory and the visual domain. Therefore, a classification of ideophones as belonging to a single domain may not be possible. Only seven from among 500 ideophones tested by Nuckolls (2019, 193) are, in fact, monosensory.

Findings such as those reported by Nuckolls (2019) challenge the categorization to a single sensory domain. It has also been previously noted by Rácová (2014, 22) that many of the Bengali ideophones are polysemous, as they denote various semantic concepts, e.g., *gargar* may represent a feeling, a sound, or something visual (cf. Rácová, 2014, 22). The implicational hierarchy does not mention that an ideophone may belong to multiple categories. Nevertheless, as Dingemanse (2012, 663) states explaining the order of the hierarchy, "[M]OVEMENT frequently comes packaged together with sound in sensory input (...)". Hence, an ideophone related to movement may also express the sound that this movement involves. The categories put forth intertwine. As multisensory as perception is, a clear-cut categorization of words as expressive as ideophones is, thus, questionable.

The purpose of the current investigation is to classify some of the German ideophones into sensory domains. Based on native speakers' judgments, the study empirically evaluates the implicational hierarchy proposed by Dingemanse (2012) and its revision by McLean (2020). Further, it aims to test whether German ideophones can be multisensory in nature, as shown by Nuckolls (2019). Finally, we perform exploratory analyses, which group the ideophones together.

4.4.1 Method

Experimental Design

An online experiment was designed to gather native speakers' judgments on the sensory domains of German ideophones. Participants were asked to indicate the domain they associate the ideophone with. There was a total of seven possible answers, which were based on Dingemanse's hierarchy (2012) and its revision by McLean (2020). They included: *ein Geräusch* 'a sound', *eine Art von Bewegung* 'a type of movement', *etwas Visuelles* 'something visual', *ein Gefühl oder Zustand* 'a feeling or state', *eine Oberflächenstruktur* 'a surface texture', *eine Gestalt* 'a form', and *eine andere Wahrnehmung. Welche?* 'a different perception. Which one?'. The latter possibility was an open question and participants could type their feeling into a box. For each item, only one option could be chosen.

Each stimulus was embedded in a carrier sentence *Was verbinden Sie mit ...?* 'What do you associate with ...?'. There were a total of 70 ideophones. Those included 56 ideophones from the collected data set (cf. Chapter 4.2), some of which were among the most frequent ones, and some with only a few instances of occurrence. Fourteen remaining ideophones were not found in our data set, however, they are commonly used in everyday language and were selected by a native speaker. The choice of the ideophones was motivated by the judgment of the same native speaker to represent a whole semantic and sensory breadth of German ideophones. The number of ideophones was motivated by the total length of the experiment. Alongside the ideophones, 15 control items were implemented into the experiment. These included words that most likely belong to a single sensory domain, such as *Dreieck* 'triangle', *rot* 'red', or *joggen* 'to jog'. Thus, the task comprised 85 items in total. Both the items, i.e., questions, and the possible answers within a question were randomized. All ideophones and control items were presented in lowercase, regardless of their lexical class.

Table 4.2 shows the complete list of items, both ideophones and control items, and their translations. The list, as well as other files regarding the project, are available in an OSF repository: `https://osf.io/syw8k/`.

An online questionnaire was generated using SoSci Survey (Leiner, 2019), and the participants were recruited via a commercial platform Clickworker. The whole procedure took about 10 minutes. The participants were monetarily compensated for their time.

TABLE 4.2: The list of ideophones and control items used in the experiment, with translations, where appropriate, sorted alphabetically. Regardless of the lexical class, all items were written in lowercase.

Ideophones		**Control items**
ach	*nanu* 'well'	*bunt* 'colorful'
aha	*oh*	*Dreieck* 'triangle'
aua 'ouch'	*oje* 'oh dear'	*fruchtig* 'fruity'
auweia 'ouch'	*pah* 'huh'	*joggen* 'to jog'
bäh 'ugh'	*papperlapapp* 'rubbish'	*kariert* 'checked'
ballaballa 'whacko'	*pardauz* 'whoops'	*krank* 'sick'
brumm 'buzz'	*pfui* 'ugh'	*kriechen* 'to crawl'
bums 'bang'	*platsch* 'splash'	*laut* 'loud'
dallidalli 'chop-chop'	*plemplem* 'cuckoo'	*leise* 'quiet'
dingdong	*plopp* 'plop'	*Liebe* 'love'
flapp 'flap'	*plums* 'plop'	*rau* 'rough'
flugs 'in a flash'	*potzblitz* 'goodness gracious'	*rot* 'red'
flutsch 'slip'	*pst* 'psst'	*salzig* 'salty'
hä 'eh'	*puff*	*schleimig* 'slimy'
hallodri 'playboy'	*puh* 'phew'	*stinkt* 'stinks'
hauruck 'heave-ho'	*quack*	
hicks 'hic'	*rabatz* 'fuss'	
hmm	*rambazamba* 'fuss'	
holterdiepolter 'helter-skelter'	*ratsch* 'rip'	
hoppla 'whoops'	*ratz fatz* 'lickety-split'	
huhu 'hey'	*remmidemmi* '(to make a) racket'	
hui 'whoosh'	*rickeracke* 'scritch-scratch'	
hurra 'hurray'	*ruck zuck* 'chop-chop'	
husch 'shoo'	*rülps* 'burp'	
igitt 'yuck'	*schickimicki* 'posh'	
klackerdiklack 'click-clack'	*schnipp schnapp* 'snap-snap'	
klatsch 'clap'	*schwups* 'whoosh'	
knurr 'growl'	*singsang* 'ditty'	
krächz 'croak'	*tatütata* 'nee-naw'	
kracks 'crack'	*tohuwabohu* 'chaos'	
krimskrams 'junk'	*ups* 'oops'	
kuckuck 'cuckoo'	*wirrwarr* 'chaos'	
kuddelmuddel 'chaos'	*wumms* 'oomph'	
mischmasch 'hotchpotch'	*wusch* 'swoop'	
naja 'well'	*zickzack* 'zigzag'	

A total of 135 participants took part in the online questionnaire. It was crucial that all participants are German native speakers, as ideophones exploit language-specific behavior (Akita, 2015; Iwasaki et al., 2007a,b). After excluding non-native speakers and incomplete cases, there were 121 valid cases left. Among them, there were 60 female and 61 male participants. The age of those 121 participants ranged from 18 to 69 years (mean = 38.41, SD = 12.05).

Data Analysis

All analyses were carried out in `R` (R Core Team, 2019). We used `tidyverse` package for data processing (Wickham, 2017), and `ggplot2` for data visualization (Wickham, 2016).

All items were controlled for the frequency of response across participants. In case of an open answer (*eine andere Wahrnehmung. Welche?* 'a different perception. Which one?'), the typed answers were inspected and grouped into possible sub-categories. Subsequently, ideophones were categorized according to the top-picked category. The percent value of the total number of votes was calculated to evaluate whether the choice of a category was straightforward. If the difference between the first two categories was less than 5%, we saw both categories as applicable for the given ideophone.

We also performed an exploratory k-means clustering (Likas et al., 2003) using the package `cluster` (Maechler et al., 2021), and `factoextra` (Kassambara and Mundt, 2020) for visualization. Clustering is a mathematical procedure of sorting things together according to different parameters given for those things. It is looking for similarities and dissimilarities across the data and grouping the data accordingly. As Romesburg (1984, 2) puts it, "if we gathered a set of pebbles from a stream shore, noted their attributes of size, shape, and color, and sorted similar pebbles into the same piles, we would be physically performing a cluster analysis". The number of piles would then represent the number of clusters.

The clustering was performed for all ideophones based on the percentage of participants who chose a given category ($N = 7$). Hence, we used percentages of all categories for every ideophone, similarly to the attributes size, shape, and color used for pebbles in the example above. Using this approach, we wanted to see which ideophones will be grouped based on votes for all categories. The data was scaled prior to clustering. First, we computed k-means for $k = 2 : 5$, i.e., 2–5 clusters, in order to visually determine the optimal number of clusters. The estimation of the optimal number of clusters was done using the elbow, silhouettes, and gap statistics method (Milligan and Cooper, 1985; Tibshirani et al., 2001). While the latter method returned that one cluster is optimal, the two first methods indicated that four are optimal; therefore, the results will refer to clustering where $k = 4$. Finally, a χ^2-test was performed to examine how the clusters correlate with the first-choice categories.

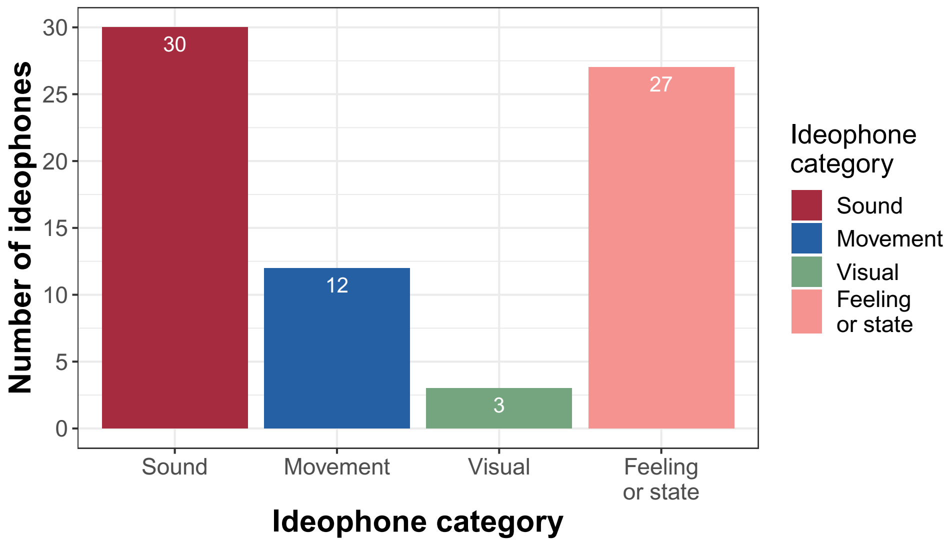


FIGURE 4.6: Ideophone categories, which were identified for the subset of the corpus ideophones. The categories are presented in the order proposed in the implicational hierarchy (Dingemanse, 2012).

4.4.2 Results

Implicational Hierarchy (Dingemanse, 2012; McLean, 2020)

The results of the analysis are shown in Figure 4.6. The categories in the plots are given in the order of the implicational hierarchy. The results for the first three categories – SOUND, MOVEMENT, and VISUAL PATTERNS – are in line with the hierarchy (Dingemanse, 2012). These results, however, question the revised hierarchy by McLean (2020), who suggested that SOUND and MOVEMENT should be followed by FORM and TEXTURE categories. These two categories did not receive any resonance in our experiment. We found that 30 ideophones from the data set were associated with SOUND, 12 with MOVEMENT, and three with VISUAL PATTERNS. No ideophone was identified for the category that follows VISUAL PATTERNS, i.e., OTHER SENSORY PERCEPTIONS. However, we found that 27 ideophones were classified among INNER FEELINGS AND COGNITIVE STATES. This finding challenges the implicational hierarchy, which forsees that, in a given language, ideophones relating to OTHER SENSORY PERCEPTIONS are a prerequsite for those expressing INNER FEELINGS AND COGNITIVE STATES (Dingemanse, 2012, 662ff.).

Multisensory Perception of Ideophones (Nuckolls, 2019)

Two of 70 ideophones could not be assigned to one category only, as the difference between the two top-picked categories amounted to less than 5%. This was the case for ideophones *holterdiepolter* and *schnipp schnapp*. *Holterdiepolter* 'hurried uneven noisy movement' was primarily assigned to the category MOVEMENT with 40.50% of the total votes; however, it also received 39.67% of the votes in the category SOUND.

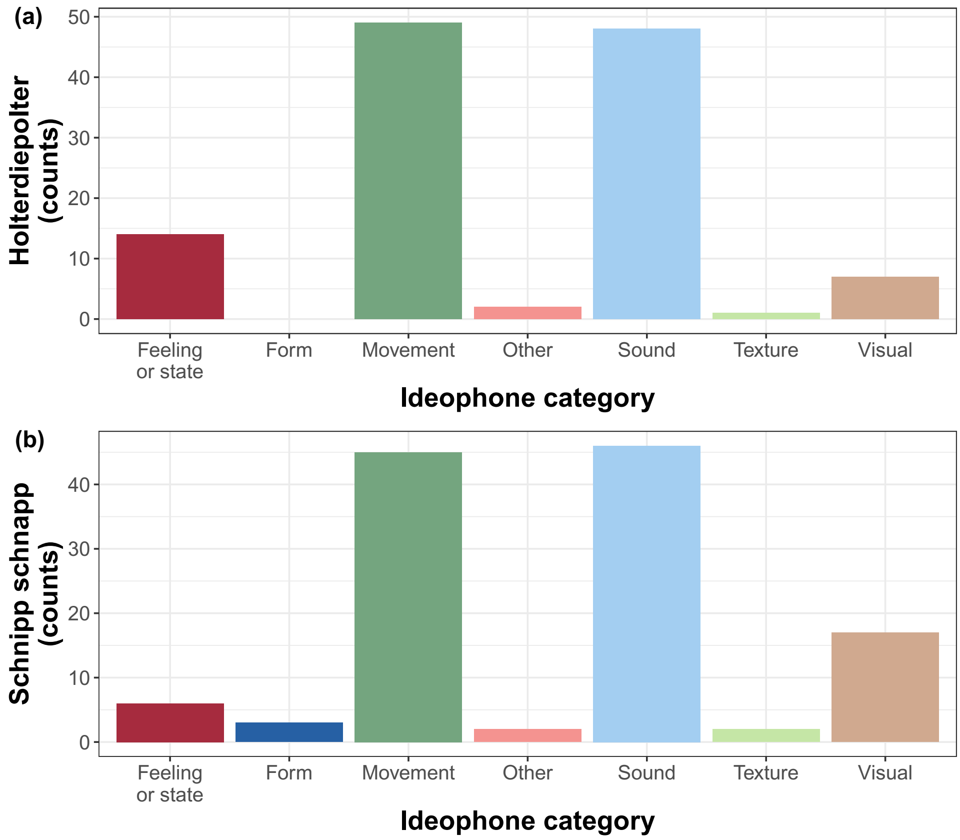


FIGURE 4.7: Category counts for (a) *holtediepolter* and (b) *schnipp schnapp*. Two categories MOVEMENT and SOUND are dominating, Importantly, they differ from one another only minimally.

For *schnipp schnapp* 'the sound of cutting', while 38.02% of the participants chose the category SOUND, 37.19% favored the category MOVEMENT.

Because the participants could only choose one category, we disabled the possibility of a within-participant multisensory rating. The multisensory rating is, however, visible in a between-participants manner. Figure 4.7 illustrates the vote counts across all categories for the two multisensory ideophones mentioned in the above paragraph – (a) *holterdiepolter* and (b) *schnipp schnapp*. It can be seen that the categories MOVEMENT and SOUND are almost equal. Hence, those ideophones show the tendency to be associated with both categories.

Exploratory Analyses

Even though we initially accepted the threshold of 5% between the first two categories, we also examined further inconsistencies in the category assignment. The primary goal was to gain a better picture of the sensory breadth of ideophones. Nevertheless, the following results are exploratory.

As mentioned above, two ideophones were clearly identified in two different categories. Extending the threshold of 5% difference between the first and the other category assignment, we saw a disagreement in 12 further ideophones. Within 10% range from the first-choice category, there are three further instances. Most notably, *wusch* 'faster than expected', which 47.93% of the participants assigned to the category MOVEMENT and 42.15% to the category SOUND (difference = 5.78%). Interestingly, for *rambazamba* 'hustle and bustle', INNER FEELINGS AND COGNITIVE STATES was the first-choice category with 32.23%, but was closely followed by two other categories: SOUND, with 25.63%, and MOVEMENT, with 23.97% (difference to the previous category = 6.60% and = 1.66%, respectively). Lastly, *pardauz* (cf. Chapter 4.3.1) was primarily assigned to SOUND, with 33.06% of the votes, and followed by INNER FEELINGS AND COGNITIVE STATES with 23.97% (difference = 9.09%). Looking forward into the following extension steps, we also found that *pardauz* was assigned to MOVEMENT with 14.88% (difference to previous category = 9.09%).

Four additional cases of divergence, alongside those mentioned in the previous paragraph, were found when we extended the threshold to 15% difference from the first-choice category. *Remmidemmi* 'hustle and bustle' was assigned to *sound* by 40.50% of the participants, however 27.27% categorized it as INNER FEELINGS AND COGNITIVE STATES (difference = 13.23%). The relative difference between the first and the second choice for the remaining three ideophones is equal and amounts to 14.05%. For *plums* 'dull clapping sound of falling' 51.24% of the participants decided for SOUND and 37.19% for MOVEMENT. *Kuddelmuddel* (cf. Chapter 4.3.1) was assigned to INNER FEELINGS AND COGNITIVE STATES by 41.32% of the participants and to VISUAL PATTERNS by 27.27%. And *zickzack* was primarily seen as MOVEMENT by 39.67% of the participants, followed by FORM with 25.62%.

The last threshold we adapted for the exploratory analysis is 20% range from the first-choice category. Here, we found five further divergent cases, in addition to the previously mentioned ones. With the smallest difference in this group, *hoppla* (cf. Chapter 4.3.1) was assigned to INNER FEELINGS AND COGNITIVE STATES with 47.11% of the votes, and to MOVEMENT with 30.58% (difference = 16.53%). For *Mischmasch* (cf. Chapter 4.3.1) three categories were identified within the range, as it yielded 37.19% of the votes for VISUAL PATTERNS, then 20.66% for TEXTURE, and 19.00% for INNER FEELINGS AND COGNITIVE STATES (difference to the previous category = 16.53% and = 1.66%, respectively). Similarly, three categories were revealed for *hui* 'exclamation of surprise; fast movement', primarily INNER FEELINGS AND COGNITIVE STATES with 42.98%, followed by SOUND with 25.62%, and MOVEMENT with 23.14% (difference to the previous category = 17.36% and = 2.48%, respectively). For *rickeracke* 'sound of crushing', 44.63% of the participants chose the category SOUND and 25.62% the category MOVEMENT (difference = 19.01%). Lastly, *wirrwarr* was identified as INNER FEELINGS AND COGNITIVE STATES by 45.45% of the participants, and as VISUAL PATTERNS by 25.62% (difference = 19.83%).

In light of those results, we performed a cluster analysis on all ideophones based on the percentage of votes for all seven categories. Four clusters were determined

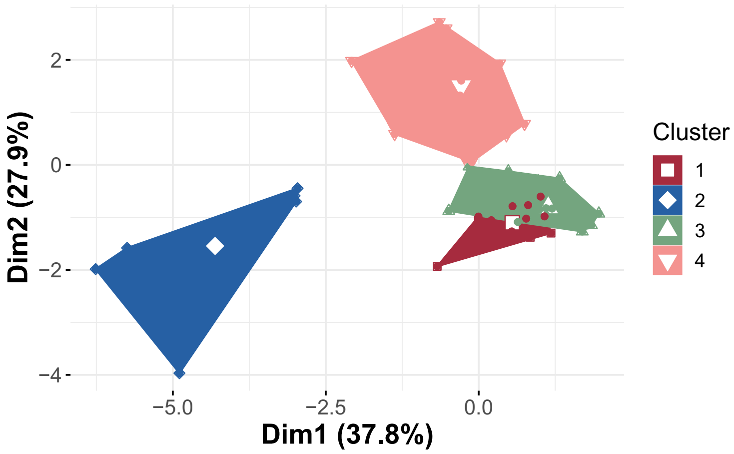


FIGURE 4.8: K-means clustering for 70 German ideophones based on participants' judgments for all ideophone categories. The optimal number of clusters was determined to be four, and these are depicted in different colors. The axes y and x show two dimensions created by the algorithm based on seven variables, i.e., ideophone categories. Together, those dimensions explain 65.7% of variance in the data.

to be optimal. Interestingly, ideophones were assigned to four from the seven categories (cf. Figure 4.6). The results of the clustering can be seen in Figure 4.8.

The first cluster contains $N = 12$ ideophones and has a within-cluster sum of squares (WCSS) of 18.11. The WCSS shows the variability of the observations within a cluster – the smaller the value, the more compact the cluster. The second cluster contains $N = 6$ ideophones and has a WCSS of 52.63, which can be seen by the spread of the blue cluster in Figure 4.8. The third cluster includes $N = 24$ items and has a WCSS = 23.39, and the fourth cluster comprises $N = 28$ ideophones and has a WCSS of 61.47. The list of ideophones by clusters is given in Table 4.3.

As Figure 4.8 reveals, there is an overlap between clusters 1 and 3. The overlapping clusters are the ones with the smallest variation, as shown by the WCSS. The overlap is a problem for interpretation and calls the appropriateness of the algorithm into question. However, for the sake of the exploratory analyses, we calculated whether the clusters correlate with the first-choice categories, as picked by the participants.

The results of a chi-squared test show that there is a significant relationship between cluster and the first-choice category, $\chi^2(9, N = 70) = 142.31, p < .001$. The corrected effect size, Cramer's V, was strong, $V = 0.81$. As Figure 4.9 shows, only one cluster (3) contains ideophones, which were initially assigned to the same category SOUND. However, four further SOUND ideophones are scattered across clusters 1 and 4, with two SOUND ideophones in each cluster. Apart from those, cluster 1 contains only MOVEMENT ideophones, and cluster 4 only INNER FEELINGS AND COGNITIVE STATES ideophones. Cluster 2 shows the most variance when compared

TABLE 4.3: The list of ideophones by clusters, sorted alphabetically. The numbers and symbols of the clusters correspond to the clusters from Figure 4.8. Within the cluster, some ideophones are colored to signify the first-choice category, corresponding to the color scheme in Figure 4.9. If the ideophone within the cluster is not colored, it corresponds to the main category of a given cluster from Figure 4.9.

Cluster 1 ■	Cluster 2 ◆	Cluster 3 ▲	Cluster 4 ▼
dallidalli	*krimskrams*	*brumm*	*ach*
flugs	*kuddelmuddel*	*bums*	*aha*
flutsch	*mischmasch*	*dingdong*	*aua*
hauruck	*schickimicki*	*flapp*	*auweia*
holterdiepolter	*wirrwarr*	*hicks*	*bäh*
husch	*zickzack*	*huhu*	*ballaballa*
plums		*klackerdiklack*	*hä*
ratz fatz		*klatsch*	*hallodri*
ruck zuck		*knurr*	*hmm*
schnipp schnapp		*krächz*	*hoppla*
schwups		*kracks*	*hui*
wusch		*kuckuck*	*hurra*
		platsch	*igitt*
		plopp	*naja*
		pst	*nanu*
		puff	*oh*
		quack	*oje*
		rabatz	*pah*
		ratsch	*papperlapapp*
		remmidemmi	*pardauz*
		rülps	*pfui*
		singsang	*plem plem*
		tatütata	*potzblitz*
		wumms	*puh*
			rambazamba
			rickeracke
			tohuwabohu
			ups

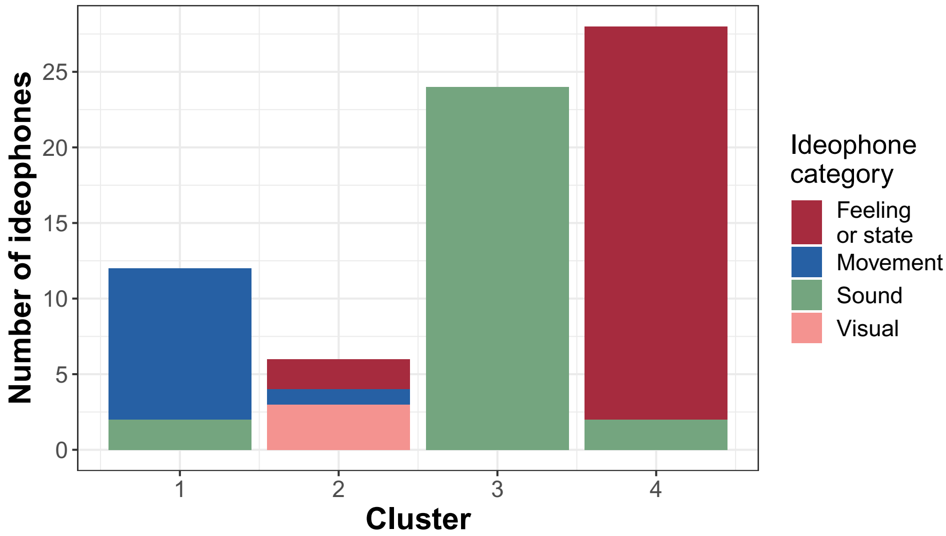


FIGURE 4.9: A stacked barplot of first-choice ideophone categories by clusters. Each cluster is represented as one bar on the x-axis. The y-axis shows the number of ideophones. Within a cluster, different colors represent different ideophone categories.

to the ideophone category preference by the participants. It is the smallest cluster with only six items, three of which belong to the category VISUAL, two to INNER FEELINGS AND COGNITIVE STATES, and a single item to MOVEMENT. Overall, we see a tendency for clustering to follow the pattern of the participants' category preference; there are, however, cases that seem to be contested.

4.4.3 Discussion

The purpose of the current study was to categorize German ideophones into domains put forth by the ideophone implicational hierarchy (Dingemanse, 2012), as revised by McLean (2020). The results reveal that, in the current data set consisting of 70 German ideophones, the hierarchy holds for the first three categories: SOUND, MOVEMENT, and VISUAL PATTERNS. And, while several ideophones were classified as the fifth category INNER FEELINGS AND COGNITIVE STATES, there were no ideophones in the fourth category OTHER SENSORY PERCEPTIONS. No evidence was found for the revised hierarchy adduced by McLean (2020).

Apart from that, we found evidence corroborating Nuckolls's (2019) assumption that ideophones do not necessarily belong to a single sensory domain. In our data set, two ideophones were assigned to two different categories, at a threshold less than 5% between those categories. However, further exploratory analyses revealed that by increasing the threshold, the number of divergent ideophones increased, some of which circulated not only among two but among three sensory domains.

Additionally, we performed an exploratory k-means cluster analysis that grouped the ideophones into four clusters based on the percentage of assignments across all categories. We found a strong correlation between the selected cluster and the participants' first-choice category. Nevertheless, we treat those results with precaution, primarily because of the overlap found between two clusters but also because of the exploratory nature of this analysis.

The surprising result for the INNER FEELINGS AND COGNITIVE STATES category led us to inspect the character of the ideophones which were assigned to this category. We found that 74% of those ideophones were interjections according to Duden online (Dudenredaktion, n.d.). In the category VISUAL, all ideophones are nouns, and in the category MOVEMENT the best-represented part of speech with 50% are adverbs. Even though there is a substantial number of interjections in the category SOUND, i.e., 56%, it is worth noting that 26% of the ideophones in that group were not included in the Duden online at all. Therefore we remark striking domination of interjections, especially among the INNER FEELINGS AND COGNITIVE STATES ideophones.

Ideophones are often found as a subclass of one or more parts of speech (Kilian-Hatz, 2006). As noted by Ameka (2006, 743), interjections "conventionally constitute utterances by themselves and express a speaker's current mental state or reaction toward an element in the linguistic or extralinguistic context". Among them, there is a subgroup of iconic words, such as ideophones (cf. Ameka, 2006, 743). What characterizes both ideophones and interjections is the syntactic independence (Dingemanse, 2017b) and a bizarre phonological and morphological behavior (cf. Ameka, 2006; Kilian-Hatz, 2006). Therefore, we may observe that the two notions overlap – not only in their structural aspects but also in their expressive character.

Dingemanse (2017b, 15) argues that ideophones enrich our view on language by noticing the variety of modes of representation, and interjections help us reach "beyond sentence level, and remind us that language is social-interactive at core". However, ideophones might be register-dependent, thus, serve as a similar social-interactive instrument, too (cf. D'Onofrio and Eckert, 2020). As Akita (2017, 326) shows, there is a difference in the morphosyntactic integration of ideophones between registers (i.e., normal and colloquial). In colloquial Japanese, ideophones are morphosyntactically integrated to a greater extent in comparison with plain speech. To the best of my knowledge, there exist no studies on the ideophone use in different registers of German. However, ideophones are rarely, if at all, used in formal contexts. Recently, primarily young companies use them in advertisements to approach young customers (e.g., *Durstexpress*, cf. Figure 4.10). Ideophones are used in less formal registers with friends and family, creating an atmosphere of closeness. The alternation between Standard German and a dialectal or more colloquial speech resembles code-switching (Blom and Gumperz, 2000; Woolard, 2005), which in monolinguals, who switch in the dialect use, has also been labeled as style-shifting (Ervin-Tripp, 2002, 47). As put by Irvine (2002, 21), "'style' crucially concerns distinctiveness (...)." In German, ideophones may as well serve as a marker that signals meaning beyond the sentence level and carries not only its iconic, sensory meaning

FIGURE 4.10: *Durstexpress* advertisement using ideophones. The text translates to 'You're like: Rämäläām! We're like: Ding dong!'. *Ding dong* imitates the sound of a bell (cf. Chapter 4.3.1), *rämäläām* is a no word that should most likely imitate the notion of partying with music.

but, as a function of style, also a kind of a social meaning.

Consequently, there may be a need to characterize closer what ideophones and interjections are in language systems such as German. As noted before, many of the ideophones tested here are defined as interjections. Nevertheless, just like not every quadrangle is a square, not every interjection is an ideophone (Ameka, 2006, 743). In German, also not every ideophone is an interjection (cf. Chapter 4.3.1). Ideophones in German, thus, seem to be a category overarching the parts of speech. The point is not to deny it or postulate for a separate word class of ideophones. It is to look deeper into what makes a certain member of a given word class an ideophone with special attention to the interjections.

As Dingemanse and Akita (2016) show that the more grammatically integrated a segment is, the less expressive it is. Then, the depictive character of ideophones becomes descriptive. And, as the process of deideophonisation occurs with time (Dingemanse, 2017a; Dingemanse and Akita, 2016), ideophones become more and more integrated and less and less expressive. Undoubtedly, we may see traces of such a process in German. As mentioned in Chapter 4.2.4, there exist verbs with sound-symbolic character, e.g., *poltern* 'to crash', or *fiepen* 'to whimper'. Among the ideophones we find *holterdiepolter* 'hurried uneven noisy movement', *fiep* 'sound of a bird'. An interesting question would be what is the origin of those words, as debated in Chapter 4.2.1. In turn, it would be possible to see whether we are dealing with deideophonisation or ideophonisation – whether the ideophone is becoming more integrated or whether the verb is losing its status and becoming more expressive. The iconic treadmill hypothesis by Flaksman (2017) suggests that the coinage of new iconic words in a language is motivated by the deiconization of old words. Through sound change and sense development, words lose their iconicity over time (Flaksman, 2017, 17f.) and this process stimulates the addition of new iconic words to the lexicon. It seems that the level of iconicity remains the same, and the shifts are best visible in the diachronic perspective.

The exploratory analysis with an extended threshold reveals a far greater sensory breadth and speaks for the strong coupling of sensory cues within an ideophone concept. Even though the task disabled a within-participant multicategorization, our data shows that in a between-participant manner. And when "[m]ost of our everyday experiences (...) are multisensory" (Spence, 2012, 37), why would not the "marked words that depict sensory imagery" (Dingemanse, 2012, 655) be multisensory, too? Importantly, previous studies show that words more linked to sensory experience are more iconic in their form (Perlman et al., 2018, for signed and spokes languages; and Winter et al., 2017, for English).

The pattern revealed by the explorative clustering analysis reveals some similarities to the categorization as shown by the global participants' judgments. Participants assigned the ideophones to four categories; similarly, four clusters were optimal for this data set. Three out of four clusters have a dominating category – MOVEMENT, SOUND, or INNER FEELINGS AND COGNITIVE STATES – and there is much variance in one of the clusters. This may again signal the multisensory character of some ideophones. Combining these careful assumptions with the results of the hierarchy testing, where two ideophones were assigned to two different categories by an almost equal number of participants, we can put forth a hypothesis that there exist monosensory and multisensory ideophones in German.

The results of the multisensory nature of ideophones (cf. Chapter 4.4.2) point to the intertwining of the two first categories: SOUND and MOVEMENT. As noted before, "[M]OVEMENT frequently comes packaged together with sound in sensory input (...)" (Dingemanse, 2012, 663). The reason why we found those two categories frequently coinciding in ideophones may be the existence of sound emission verbs in German (Goldberg, 1995; Levin and Rappaport Hovav, 1991, 1992). This, in turn, is dependent on a typological distinction between verb-framed and satellite-framed languages (Talmy, 1985; Talmy, 1991). In a verb-framed language, like Spanish or French, the verb focuses on a path of motion, i.e., *enter, exit, ascend, descend, pass*, etc. (Talmy, 1991, 489). In a satellite-framed language, the focus of a verb lies on the manner of motion. Both English and German serve as an example. "The bottle floated into the cave" (Talmy, 1991, 488) is a perfectly understandable sentence and can be translated to German as *Die Flasche trieb in die Höhle hinein.* In Spanish, the manner of *floating* cannot be incorporated into the verb, i.e., *La botella entró flotando a* *la cueva* 'The bottle entered (MOVED-in) floating to the cave' (Talmy, 1991, 488). Similarly, verbs expressing sounds, i.e., sound emission verbs, may be used to express movement relating to the sound. A sentence *The train screeched into the station* (Goldberg, 1995, 62) can be easily translated into German: *Der Zug quietschte in den Bahnhof hinein.* The link between movement and sound within a verb in satellite-framed languages may have implications on the nature of ideophones in those languages, such that they are more prone to be multisensory than in verb-framed languages. Studies reveal that while in English, verbs receive high iconicity ratings, in Spanish, iconicity in verbs is perceived as low (Perlman et al., 2018; Perry et al., 2015). Therefore, future research investigating the multisensory nature of ideophones, especially regarding SOUND and MOVEMENT ideophones, should take into consideration whether there exist sound emission verbs in the given language.

Lastly, the results of the ideophone categorization by native speakers show that clustering may be a powerful tool to group the highly multisensory character of ideophones. Even though hierarchical clusters are preferred for smaller data sets, and their results are more robust than k-means, I decided to use k-means for the exploratory analyses because of their simplicity. I refrain from drawing any conclusions from these results. I nevertheless hope that the findings reported here may serve as a basis for further research. I want to stress that future work should focus on using more robust methods, such as hierarchical clustering, with a possibly more extensive data set to uncover the complexity and entanglement of ideophone meaning.

In sum, then, I showed that a subset of the ideophones from the collected data set partly aligns with the implicational hierarchy by Dingemanse (2012). Further, I demonstrated that German ideophones have multisensory potential (Nuckolls, 2019). All in all, I believe that, in order to investigate ideophones in German, we first need to establish more concise guidelines on how to deal with words classified as interjections. Simply put, we need an answer to the question, when does an interjection have an ideophone character and when it does not. Armed with that, we will be able to better understand the interplay of sensory experiences hidden between the lines.

4.5 Age Hypothesis for the Use of Ideophones

Now I would like to move to the second experimental analysis based on the collected data. At the beginning of the last chapter, I used the adverb *how* to introduce my intentions in the investigation of ideophones. And I believe that the problem tackled in this chapter also refers to *how* the ideophones are, more specifically, *how* they are within books for different groups of children. But, also, though somehow philosophical, this chapter may help answer the question *why* ideophones are.

Iconicity is "a powerful vehicle for bridging between language and human sensori-motor experience" (Perniss and Vigliocco, 2014, 1) and it can serve as a base for establishing novel vocal symbols (Perlman et al., 2015b; Perlman and Lupyan, 2018). It has been shown that iconicity aids word learning in adults (Jones and Vigliocco, 2017; Nygaard et al., 2009a). Lockwood et al. (2016a) conducted two experiments to see how iconicity can boost word learning. In the first experiment, participants learned Japanese ideophones with their real meanings or with opposite meanings. After the main task, the participants were told that half of the words they learned had incorrect meanings. Subsequently, they were asked to perform a forced-choice task on the same set of ideophones in which they had to choose the meaning they felt was correct. The results revealed that participants learned better in the matching category and that they were also able to guess the real meaning above chance. In the second experiment, participants similarly learned Japanese adjectives – either with matching or opposite meanings. Here, no difference was shown between the matching and opposite conditions. Overall, Lockwood et al. (2016a) demonstrated that Dutch native speakers learn Japanese ideophones better than Japanese adjectives and that they can guess the ideophone meaning well above chance.

However, not only adult populations are prone to the effect iconicity has on word learning. Already preverbal infants are sensitive to sound-symbolic cross-modal correspondences (Asano et al., 2015). Studies with three-year-olds reveal that sound symbolism facilitates verb learning in both Japanese- (Imai et al., 2008) and English-speaking children (Kantartzis et al., 2019, 2011). Based on these and similar results, Imai and Kita (2014) put forth a "sound symbolism bootstrapping hypothesis". According to the hypothesis, sound symbolism, and iconicity in general, helps children learn the language and distinguish the reference of the word. Iconicity is able to do so, because it shaped the protolanguage at its origin (Imai and Kita, 2014, 9) – it is an innate need, "a general property of language" (Perniss et al., 2010, 1).

We see that children's perception is sensitive to iconicity – but the language of their caretakers is also full of it. This "feedback loop" yet again signifies the strength of iconicity in language development. Yoshida (2012) demonstrated that children of both Japanese- and English-speaking parents benefit from sound symbolism during verb learning, even though the two parent groups used different strategies to convey sound symbolism. While Japanese speakers relied on conventionalized sound-symbolic forms, English speakers invented new idiosyncratic ones. It shows that, regardless of the available iconic means in a language, caretakers still make use of iconicity (Yoshida, 2012, 13). In addition, a recent study by Perry et al. (2021) shows that iconicity does aid children to retain the novel word and helps them overcome the Quine's problem (as suggested by Imai and Kita 2014, 1, and Perniss and Vigliocco 2014; also see Quine 1960, 33).

An interesting question is what happens in the course of the development. How does the use of iconic words change over time? Perlman et al. (2017) conducted a longitudinal study with 35 children and their caretakers to examine this question. Both in child-produced and child-directed speech, iconicity decreased over time. In children's speech, the effect was caused by an escalated use of arbitrary words; the children maintained, however, the frequency of iconic words. The effect was the opposite in the caretakers' speech behavior. While they sustained the frequency of arbitrary words, they significantly decreased the use of iconic words. The change in parent's behavior was driven by the cognitive ability of children, which suggests that parents adapt their language to children's skills (Perlman et al., 2017, 917).

The prominent role of iconicity in language acquisition is endorsed by a growing body of research. Studies show that children acquire sound symbolic and iconic words earlier (Laing, 2019; Perry et al., 2015). In addition, sound-symbolic words are more frequent in children's speech, as well as in the child-directed speech of caretakers (Perry et al., 2018). Perry et al. (2018) also corroborate the findings by Perlman et al. (2017) that the effect of iconicity is most pronounced in younger children. Similarly, Massaro and Perlman (2017) report that iconicity is more widespread in early language acquisition, and they pinpoint the general vocabulary increase as a factor that decreases the proportion of iconicity in children's productive and receptive vocabulary. Recent evidence suggests that the effect goes beyond infancy and that children further profit from iconicity at later ages (Sidhu et al., 2021a, 22, for ages 7–10 years).

So far, the studies have focused on spoken language. However, as our data set shows, children's books are a unique source of iconic words (cf. Chapter 4.2). We were interested whether the sound symbolism bootstrapping hypothesis (Imai and Kita, 2014), and its role in different stages of the development (e.g., Perlman et al., 2017; Perry et al., 2018), is present in written language. The main question is if we will observe a decrease in the use of iconic words over time, similarly to Perlman et al. (2017) and Perry et al. (2018). If the frequency of ideophones in books for younger children is higher than in books for older children, this may be yet another point of evidence for the bootstrapping effect that sound symbolism and iconicity, in general, have on language development.

4.5.1 Method and Analysis

The data set of German ideophones collected from children's books was used to examine whether the ideophone use in written language changes over time. The details on the data set can be found in Chapter 4.2.4. Most importantly, as described in Chapter 4.2.4, the books used to compile the data set can be divided into two groups accordingly to the age recommendation. The first group – up to 6 years of age – includes data from 342 books, and the second group – for children between 7–9 years of age – 89 books. While the total corpus includes 1,020 word forms (i.e., ideophones as they were written in the book) and 650 lemmas (i.e., ideophones in a unified spelling), the counts for two age groups vary. In the first group, there are 856 word forms and 572 lemmas; in the second group, there are 333 word forms, and 239 lemmas.

We started the analysis by inspecting the mean number of pages, as well as the mean number of words per page in each age group. This descriptive analysis was done to see the fundamental differences there may be between the two categories. Because the data was not normally distributed, we additionally computed the medians to compare the two age groups with a Wilcoxon signed-rank test.

In order to compare the frequency of ideophones in both groups, we calculated the proportion of ideophones as compared to the number of all words on a book page. Thus, the number of words occurring on each page that was available through corpus collection was calculated. This was done semi-automatically (cf. Chapter 4.2.3) – we used a picture-to-text function in Microsoft OneNote (Microsoft Corporation, 2018) for the initial estimation, however we checked the recognition by hand for every page. The picture-to-text function did not work when the text was handwritten or arranged in other than the left-to-right dimension. In those cases, the words were counted by hand. Subsequently, we calculated the proportion of the ideophones to all words.

Apart from frequency, we were also interested in whether the variability of ideophones differs between the books for different age groups. Two calculations were used to check that. The first one summarized the total number of different lemmas – and the total number of word forms for comparison – in both age groups. The second proofed the repetitions of the same lemma – and the same word form for comparison – within a single book for both age groups.

Neither the proportion data nor the data on ideophone variability was normally distributed. Therefore, to check the effect strength statistically, we computed one-sided Wilcoxon signed-rank tests to compare both age groups in all conditions. It should be noted that this is a type of non-parametric test, and it does not rely on the comparison of the means but rather on the comparison of the medians. The tests were unpaired because the data stemmed from different sets of books, and one-sided because we assumed that the use of ideophones decreases over time. All *p*-values obtained from the Wilcoxon signed-rank tests were Bonferroni adjusted.

All analyses mentioned above were carried out in `R` (R Core Team, 2019). `Tidyverse` was used for data processing (Wickham, 2017), and `ggplot2` for data visualiztion (Wickham, 2016).

4.5.2 Results

Figure 4.11(a) shows the number of pages in books for the two age groups under consideration: up to 6 years, and between 7–9 years. The median for the first group is = 32, and the mean = 33.2, with $SD = 8.31$. For the second group, the median number of pages is = 64, the mean = 89.8, with $SD = 60$. The descriptive analysis points to a difference in the number of pages between age groups. Books for younger children have significantly fewer pages, as per a one-sided Wilcoxon signed-rank test ($Z = -47.32, p < .001, r = 0.78$). As for Figure 4.11(b), which shows the number

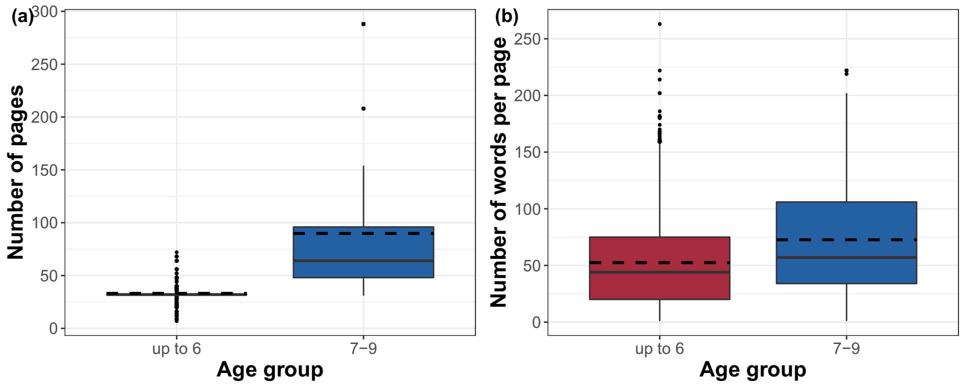


FIGURE 4.11: (a) The number of pages and (b) the number of words per page in books for the two age groups. The red bar on the left shows books for children up to 6 years old, and the blue bar on the right for children between 7 and 9 years old. The solid horizontal line on each bar signifies the median, and the dashed horizontal line shows the mean.

of words per page, the median is $= 44$, and the mean $= 52.5$, with $SD = 40.1$, for the first age group. In the second group, the median is $= 57$, and the mean number of words per page $= 72.6$, with $SD = 47.2$. The difference between groups is significant per a one-sided Wilcoxon signed-rank test, $Z = -12.79, p < .001, r = 0.21$. The effect size for the number of words per page is small, especially compared to a large effect found for the difference in the number of pages.

The proportion of ideophones across all words on a page differs significantly between the two age groups, as shown by a one-sided Wilcoxon signed-rank test, $Z = 11.61, p < .001, r = 0.26$. The median proportion of ideophones on a page for the first age group is $= 3.70$, with a mean $= 11.7$ and an $SD = 22.9$. As for the second group, the median proportion of ideophones per page is $= 2.36$, the mean is $= 3.42$, with $SD = 5.66$. The data is illustrated in Figure 4.12, with the proportion in the logarithmic scale for better interpretability. The small effect ($r = 0.26$) found in our data set shows that the proportion of ideophones across all words is greater in books for younger children.

Finally, we compared the variability of ideophones between groups. Across the board, there were 572 different ideophone lemmas in books for younger children and 239 in books for older children. As for word forms, there were 856 word forms in books for younger children and 333 in books for older children. We were specifically interested in the repetitions of the same ideophone within a book between the two age groups. First, we compared the lemma repetitions. In books for younger children, the median amounted to $= 1$, and the mean $= 4.30$, with $SD = 9.41$. In books for older children, the median was the same, thus $= 1$, but the mean number of repetitions of the same ideophone lemma in a book was $= 2.28$, with $SD = 2.03$. The result was not significant as per Wilcoxon signed-rank test, $Z = -0.17, p = 0.43, r = 0$. For comparison, we also looked at the repetitions of

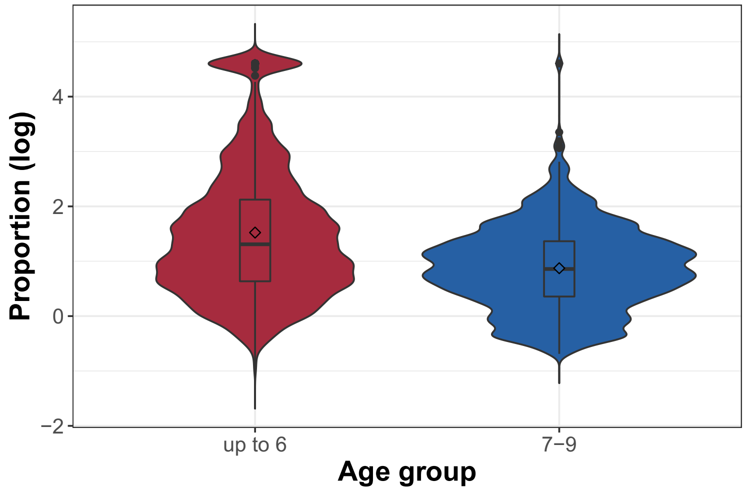


FIGURE 4.12: The proportion of ideophones per page by age group. The proportion on the y-axis is on a logarithmic scale for easier comparability. The violin plot shows the density of the data, i.e., the give log ratio in each of the groups (left/red for children up to 6 years, right/blue for 7–9 year-olds). The boxplot within the violin illustrates the median and interquartile ranges, and the diamond symbol shows the group mean proportion.

a word form within a book between the two age groups. Here, the medians for both groups were equal = 1 as well. However, other descriptive values differed. The mean in books for younger children was = 4.04, with $SD = 9.39$, and in books for older children, the mean was = 2.12, with $SD = 1.96$. Nevertheless, the Wilcoxon signed-rank test showed no effect, $Z = -0.56, p = 0.28, r = 0$. Overall, the difference in the variability is visible by looking at the total number of ideophone lemmas or word forms and the mean values of their repetition within a book. However, the statistical tests do not confirm this effect. It may thus be the case that ideophones are not generally more variable in books for younger children, but there are rather only a few ideophones that are repeated over and over again.

4.5.3 Discussion

The goal of the current study was to investigate the use of ideophones in books for children of two different age groups – up to 6 years and between 7–9 years of age. The results of a one-sided Wilcoxon signed-rank test show that the proportion of ideophones in books for younger children is significantly higher than in books for older children ($\tilde{x} = 3.70$ vs. $\tilde{x} = 2.36$, respectively). This finding corroborates the decline in the use of ideophones during language development reported by Perlman et al. (2017) and Perry et al. (2018). Furthermore, it extends this effect onto the domain of written language. Our result provides additional evidence for the sound symbolism bootstrapping hypothesis (Imai and Kita, 2014).

The first group in our study included books for children up to 6 years of age. This undoubtedly invites much within-group variation considering that Perlman et al. (2017) reported the decline throughout the time frame of two years, with an approximate mean age of children between 1;8–3;8 years, and Perry et al. (2018) over the course of roughly a year (1;2–2;6 years). Books for children, who are learning to speak, may differ drastically from those intended for children learning to read. The second group included books recommended for children between 7–9 years, thus the age span is two times smaller. Despite this within-group difference in the first group, we found an effect in the proportion of ideophones between the first and the second group. It can be the case that, when the age groups were divided more evenly, the effect would be more robust. In our case, the effect is small ($r = 0.26$), and the cause of this effect size may be the uneven age distribution between groups (6 years vs. 3 years, respectively). A more structured approach with evenly divided age groups remains a possibility for future research.

Globally, we also found significant differences in the format of the books recommended for children from different age groups. Books for younger children have significantly fewer pages and also fewer words per page. Combined with the increased proportion of ideophones per page, these results suggest that the ideophones are more pronounced in books for younger audiences. Compared with the books for older children, there are overall fewer words on a page in books for younger children, and a larger proportion of them are ideophones. Ideophones are syntactically, morphologically, and phonologically salient (cf. Ameka, 2006; Dingemanse, 2017b; Kilian-Hatz, 2006). It seems that the structural environment present in books for younger children is ideal for ideophones to show their full potential. With fewer "competing" words surrounding an ideophone, its salient character flourishes and may play a crucial role in transmitting the meaning (cf. Imai and Kita, 2014, 3).

We were also interested in comparing the ideophone repetitions within a book between both groups because repetitions may improve learning (cf. Cuddy and Jacoby, 1982; Endress et al., 2009). We found no difference in ideophone repetitions between the two groups. We did find that, cumulatively, there is more variability across ideophones in books for younger children than for older children, i.e., there are more different ideophones. However, this result should not be taken for granted as the number of books was not equal for both groups, and the sum of the different ideophones is cumulative across all books.

Altogether, our results contribute to the existing body of research on the role of iconicity in language acquisition (Laing, 2019; Massaro and Perlman, 2017; Motamedi et al., 2020; Perlman et al., 2017; Perry et al., 2015, 2018; Sidhu et al., 2021a). We showed that ideophones constitute a more significant portion of the linguistic input present in books for younger children, as compared to books for older children. Children's books are written by adult authors and are directed towards children. Whether knowingly or intuitively, the authors follow the pattern similar to the caretakers' speech behavior (Perlman et al., 2017) – they use fewer ideophones with increasing age of the audience.

4.6 Chapter Conclusion

This chapter focused on ideophones – also called expressives, or mimetics –, "marked words that depict sensory imagery" and that belong to an open lexical class (Dingemanse, 2019, 16). Through several studies and on the example of German, I showed that (1) ideophones are widely present in Western languages (cf. Chapter 4.2), (2) they can express and refer to a multisensory experience (cf. Chapter 4.4), and (3) they are a bootstrapping mechanism for language acquisition used in a written domain (cf. Chapter 4.5).

The process of collecting the ideophone corpus through children's literature enabled us to grasp the scope of ideophones in German. In total, we found 1,022 word forms, therein 657 ideophone lemmas (i.e., with unified writing). While some ideophones are used repeatedly, and undergo an only slight modification of their form, some are idiosyncratic and created ad hoc. This very first attempt to assemble a systematic collection of ideophones in the form of a dictionary (cf. Chapter 4.3.1) aims to lead the way to fill the present research gap. Currently, there exist no guidelines for languages such as German to ascertain what makes a member of a given word class an ideophone (cf. Chapter 4.4.3). We rely on works by Dingemanse (2012, 2019) and Kilian-Hatz (1999), or Nuckolls (1996), to name a few, and those undoubtedly paved the way into the world of ideophones. However, it seems that for languages, that are thought to be "idiophonically impoverished" (Nuckolls, 2004, 132), a more thorough differentiation is needed. Among the mentioned examples are cases of ideophones and verbs alike, i.e., *holterdipolter* and *poltern*, or interjections that widely behave like ideophones, i.e., *na*. Regardless, we observe that ideophones are present and important even in languages, in which they are scattered among the lexicon.

In Chapter 4.4, we showed that ideophones in German to some extent follow the implicational hierarchy by Dingemanse (2012), however not the proposal made by McLean (2020). No ideophones were classified as neither FORM nor TEXTURE (McLean, 2020). Also, in our data, there was a gap of the category OTHER SENSORY PERCEPTIONS posited by Dingemanse (2012). Moreover, we showed that ideophones exhibit a multisensory potential. Two ideophones could not be assigned to one category only. Further exploratory analyses exposed this multisensory nature even deeper. Ideophones are performative in that they connect the linguistic form to the meaning in the real world (cf. Nuckolls, 2004, 133). They can, through their iconic form, transmit a sensory experience. Seemingly an imitation of a sound may bear much more – as in our example of *schnipp schnapp*, which is not only a sound of cutting but also an act, a movement of cutting. The effect reported here was obtained on a between-participants basis; however, a within-participant experiment would be a valuable point of reference in the future.

Based on the initially collected data set, in Chapter 4.5 we provided evidence that younger children are exposed to more ideophones in books than older children. This result is interpreted in line with the sound symbolism bootstrapping hypothesis by Imai and Kita (2014). Numerous studies to date reported a possible beneficial effect of iconicity, sound symbolism, and/or ideophone use on language acquisition (e.g., Imai et al., 2008; Kantartzis et al., 2019, 2011; Massaro and Perlman, 2017; Perlman et al., 2017; Perry et al., 2021, 2015; Yoshida, 2012). Our study is the first that showed an enhanced use of ideophones towards younger children in the orthographic domain. Thus, children's books – written by adults and often, at least until a certain age, read by adults to their children – mirror the mechanisms found in child-directed speech (Perlman et al., 2017; Perry et al., 2018).

Nuckolls (2004, 133) notes that westernization and globalization deprive languages of ideophones, literacy being one of the many material conditions enhancing this effect. Literacy drastically influences human cognition (Huettig and Mishra, 2014). It is a form of "socialization into mainstream ways of using language in speech and print, mainstream ways of taking meaning, of making sense of experience" (Gee, 1986, 742). Are then ideophones made for written language? At this point, I recall what Sabine Reiter reminisced about the data collection on Awetí (Reiter, 2011) – a native speaker with whom she worked, said to her about an ideophone "this isn't important, we do not write that." (cf. Chapter 4.1). This story echoes what many works suggest, namely that ideophones lose their expressive character with increased integration into the language (Akita, 2017; Dingemanse, 2017a; Dingemanse and Akita, 2016).

In books for younger children, literary constraints, which may have a deleterious effect on ideophony (Nuckolls, 2004, 133), seem less potent than in books for older children. After all, children, to whom books are read, rely on oral communication. For German, children's literature presented itself as an optimal source of ideophone data. It includes instances of ideophones far beyond onomatopoeia. It may be the case that such a methodology lends itself to the initial data collection in other languages, in which ideophones seem to be "conspicuously underdeveloped and poorly structured" (Diffloth 1972, after Dingemanse 2019, 14).

Acquiring a language is a complex and challenging process. But with ideophones it goes a little bit more *ruckzuck*, 'fast', and it is not all such a *Wirrwarr*, 'chaos'. Ideophones provide a broad sensory experience, and maybe by doing so, they aid children – and us – to truly *feel* what we hear from others and what we want to say ourselves.

Chapter 5

General Discussion

5.1 Summary of the Findings

This thesis has explored the iconic potential of expression on various linguistic levels. In Chapter 1, I discussed the role of iconicity in the evolution of language. I concluded that not only iconic gestures but also iconic vocalizations are powerful enough to ground symbols, therefore, pinpointing the importance of multimodality. The studies in Chapters 2 to 4 of this dissertation sought to answer various questions regarding iconicity as an umbrella term, summarized in the following paragraphs.

Chapter 2 investigated the sound symbolism of German Pokémon names. The analysis showed little overlap with previous findings on other languages – only in the negative correlation between the Pokémon weight and the number of labial consonants. Nevertheless, a couple of other reliable mappings were found – a positive correlation between Pokémon defense and the number of back vowels, and a negative correlation between special defense and the number of high vowels. Based on these results, I conclude that German Pokémon names are sound-symbolic, however, in their own way. I suggest that phonology and phonotactics may govern the specific sound-symbolic expression in the language to a larger extent than it has been anticipated. Alternatively, with Pokémon names, the translation route may constrain sound symbolism.

In Chapter 3, I tested two novel hypotheses, both of which investigated the possible physiological roots of iconic prosody in pitch–elevation and pitch–size cross-modal correspondences. The first hypothesis proposed that fundamental frequency (F0) is influenced by head position. The second hypothesis suggested that the degree of jaw opening is affected by the size of an object being named. The experiment results showed no evidence for the second hypothesis and inconclusive evidence for the first hypothesis. I suggest that there is something more than physiology that drives the effect. That physiology might have been responsible at first but, throughout thousands of years, the correspondence might have been transferred into cognition. Therefore, the role of physiology might have shifted towards the role of sensory perception.

Finally, Chapter 4 was devoted to German ideophones. After collecting a data set of German ideophones from children's books and proposing a German ideophone dictionary design, I conducted two experiments with the data. The first experiment tested how German native speakers categorize ideophones. It showed that a previously proposed implicational hierarchy holds to some extent (Dingemanse, 2012), however, that its revised version finds no confirmation in German data (McLean, 2020). Furthermore, the results imply that ideophones in German are multisensory. The second experiment explored the use of ideophones in books for children of different age groups. It displayed that there are more ideophones in books for younger children, echoing the sound symbolism bootstrapping hypothesis by Imai and Kita (2014) on the example of ideophones in a written medium in German.

All of the results presented in this thesis explored the notion of iconicity, and they show that iconicity is prevalent and "pluripotential" (Abralin, 2021) due to its vast expression of different linguistic levels. Through a set of experiments, I showed evidence of iconicity on the phonetic/phonological level (Chapter 2), prosodic level (Chapter 3), and word level (Chapter 4). I also introduced how iconicity occurs in pre-linguistic signals and how it can serve to establish linguistic symbols (Chapter 1). Iconicity truly is "a general property of language" (Perniss et al., 2010) in all its facets.

5.2 Implications of the Findings

The general implication of the findings presented in this thesis is the abundance of iconicity. Iconicity, as a powerful cross-cultural tool that humans can use to ground meaning (e.g., Ćwiek et al., 2021; Nölle et al., 2020; Perlman et al., 2015b), is widespread not only in signed, but also in spoken language. And while some aspects of iconicity can be regarded as cross-linguistic (cf. Chapters 1.4, 3, and 4.4), others should be regarded within a language (cf. Chapters 2, 4.2, and 4.5). Below, I discuss the implications of each experimental chapter separately.

The experiment on sound symbolism of German Pokémon names in Chapter 2 revealed that the majority of previously established patterns did not hold for the German data (cf. Shih et al., 2019). The determined relationships, however, implied that similar traits could be encoded with different means. The expression of sound symbolism may, thus, be dependent on the phonology and phonotactics of a given system and the functional load that given characteristics carry in this language. Essentially, these results point towards a sound-symbolic functional load hypothesis, where a specific iconically prominent feature, such as size or strength, is more likely to be expressed by sound symbolism cross-linguistically. However, the expression itself underlies the phonological system – ergo, the phonemic load, as seen within a language. As demonstrated by Diffloth (1994, 112f.), there are exact oppositions to the most robust sound-symbolic relationships found thus far – /i/ = small, /a/ = big is the opposite in the case of Bahnar. This, however, as Diffloth (1994, 113) suggests, only shows how similar acoustic, articulatory, but also proprioceptive sensations (i.e., the form of the tongue vs. the amount of space left in the mouth) may produce opposite outcomes. Therefore, to seek robust sound-symbolic expressions, we should first look from a within-language perspective to later extrapolate onto cross-linguistic correspondences.

The findings in Chapter 3 signal no reliable physiological basis of iconic prosody. There was no evidence for a correspondence between the jaw opening and the size of the object, and the evidence for the influence of head position on fundamental frequency was inconclusive, given the discrepancy between the posterior probability and the 95% credible interval. There may be two possible implications from the latter. Either, there simply is no physiological ground for the pitch–elevation correspondence; but then, why would head position show an inconclusive effect and not a lack of effect? Alternatively, and what seems plausible, the effect caused by head position alone might be too subtle, for which the reasons are twofold. Firstly, in the task presented here, participants did not reach radical lowered and elevated head positions. Thus, the elicited range of movement might not have been enough to exhibit a reliable effect, as shown by the data, the priors, and the model. And, secondly, the effect might have become perceptualized throughout hundreds of thousands of years due to frequent co-occurrence (Parise et al., 2014; Spence, 2011), even though it was primordially driven by physiology.

Finally, Chapter 4 yields multiple implications. First and foremost, it shows that Indo-European languages, too, can be ideophone-rich (cf. Diffloth, 1972; Nuckolls, 1999, 2004). This is testified alone by the amount of data we were able to collect in children's books. The first experimental analysis, carried out with the collected dataset to test the implicational hierarchy by Dingemanse (2012) and McLean (2020), has two main indications. On the one hand, German ideophones exhibit multisensory characteristics as their categorization was not uniform across the participants. On the other hand, the distinction between ideophones and interjections in German proved to be difficult, which might have had an effect on the results of the first experiment. Assuming that an interjection is a reaction to an event and an ideophone a depiction of an event (cf. Dingemanse, 2021), the question emerges how to treat such cases like the German *aua* 'ow!', which can be used both as a reaction and as a depiction. Therefore, the concept of an ideophone, contrasted with the concept of interjection, may need further specification for languages such as German.

As for the second experiment in Chapter 4, the results suggest that ideophones do play a significant role in language acquisition, in line with Imai and Kita (2014). Generally, books for older children are considerably more lengthy, both in the number of pages and the number of words per page. However, the proportion of sound-symbolic words per page, i.e., ideophones, is greater in books for younger children. The effect was found despite the age span difference within the groups. The first group included books for children up to 6 years, and the second group only for children from 7–9 years. This study enriches current results on the role of iconicity in language acquisition with the dimension of written language. It signifies that children's books' authors adjust the use of iconic means depending on the age of addressees.

5.3 Future Directions

Given the scope of this work, there are multiple directions to pursue in the future. I will draft these separately for each of the areas tackled in the thesis.

First, an implication of Chapter 1, based on previous research by other authors, was that both multimodality and iconicity played a crucial role at the dawn of language. Unfortunately, there exist no time machines yet – to deliver unquestionable evidence – and I do not intend to propose inventing one as a future research goal. Instead, I suggest that future research should encourage projects to emphasize the expression of iconicity across modalities, i.e., how the visual and the oral–aural channels influence this expression between one another. A possible research question would be whether the modalities have an additive effect in reaching a communicative goal without an established communication system. Furthermore, in which cases may communication fail despite both channels? Is there a limit to the semantic breadth of iconicity in one or the other modality, or even when both are employed? Experimental methods which would help answer these questions include large-scale production and perception experiments comparable to Ćwiek et al. (2021), however integrating multimodal signals (similarly to Macuch Silva et al., 2020). Another possible route are iterated learning experiments (e.g., Edmiston et al., 2018; Erben Johansson et al., 2021; Kirby et al., 2014; Kirby and Hurford, 2002), which I have not discussed in this thesis, but which have been widely used in the context of symbol grounding, ergo, the evolution of communicative systems. Future studies with such methodology should nevertheless incorporate the role of multimodality for grounding linguistic symbols.

As mentioned in Chapter 2, a subset of German Pokémon names were analyzed here. Therefore, the next step would be to annotate a complete set of Pokémon names, not only the first generation. However, given the results, any future analysis should also include features that were not considered here. As I mentioned, a sound-symbolic functional load hypothesis crystallized in the course of the analysis, which proposes that there exist language-specific linguistic features that carry the sound-symbolic expression of certain attributes. Therefore, another possible avenue would be to compare the expression of robust correspondences (such as size–sound) across languages. It may be the case that where one language uses vowels, another language employs syllable structure.

The experiments in Chapter 3 yielded no direct answer whether iconicity was rooted in the body. One of the possible grounds for an inconclusive effect between head position and fundamental frequency might have been the insufficient difference between the highest and lowest object placement, which did not enforce a radical-enough head position. Hence, one possibility would be to introduce more radical vertical positions in the follow-up study to explore whether the effect is more pronounced in such conditions. However, another way would be to constrain the head movement and explore the effect in such a scenario. If the effect would still be present with constrained movements, we would find evidence for a robust perceptualization of this iconic relationship. Finally, I proposed in the chapter discussion that not only physiology, but also sensory perception should be considered by future studies investigating the roots of iconicity. Therefore, future research should be encouraged to seek for sensory bases of various cross-modal and iconic phenomena.

Undoubtedly, Chapter 4 encourages numerous future projects. First and foremost, as already suggested in the chapter itself, a publication of a German ideophone dictionary would fill a current research gap. For that, however, a deeper understanding of the ideophone category in German is needed. The present project exposed an overlap between ideophones and interjections that should be investigated and discussed more closely in the future. In this context, I also proposed that the use of ideophones in German might be register-dependent and that it might carry social meaning besides the sensory meaning. Thus far, no such experimental analyses exist for German. Research to come could also inspect the multisensory nature of ideophones, especially considering typological differences between languages, such as the previously mentioned satellite- and verb-framed languages. German, as a satellite-framed language, promotes the association of sound and movement; however, to the best of my knowledge, no experiments exist, which examine a multisensory nature of ideophones in a verb-framed language. The last experimental project in this thesis, regarding the use of ideophones in children's books, would greatly benefit from a more even age group distinction. By dividing the books into, e.g., two-year stretches, we could see whether the decline is gradual or whether there is a specific moment in the development that discourages the use of ideophones.

Finally, the ultimate goal of future research that I see is simply to explore the breadth of iconic expression. This thesis brings light to parts of iconicity, which, in fact, spans across all linguistic subfields and even further. As researchers examining iconicity, we should seek to exceed the boundaries that previous theories might have forced upon us.

5.4 Conclusion

The studies in this dissertation sought to expose the widespread nature of iconicity. Rather than exploring a single research question, this thesis tackled a few detailed issues, all of which concern linguistic iconicity. As an umbrella term for the subjects discussed here, iconicity was shown to be an incredibly prevalent force, anchored somewhere between language and cognition (cf. Abralin, 2021; Perlman, 2017; Perniss et al., 2010).

As humans, having this complex communication system that we call language, we may take pride in its form. "Speech perforce is largely arbitrary" (Hockett, 1978, 275), "[t]he word 'salt' is not salty nor granular; 'dog' is not 'canine' (...)" (Hockett, 1960, 90). It is easy to be proud of the seeming superiority because we, as a species, can discuss abstraction. However, it is also easy to forget how we got here. And, while there is little doubt that iconicity is crucial for symbol grounding at the heart of language (e.g., Everett, 2017; Peirce, 1955), we might be tempted to think that it is not useful anymore; that it contradicts arbitrariness, which we encounter constantly.

Nevertheless, mounting evidence, including the work in this dissertation, shows that iconicity is still doing well nowadays. It may be less evident in our vocabulary because not all words are equally iconic (Winter et al., 2017) and this tendency can differ across languages (Perlman et al., 2018). However, we do find ideophones even in languages, which are thought to be ideophone-poor (Diffloth, 1972; Nuckolls, 1999, 2004), such as the case of German (cf. Chapter 4). Moreover, as I have shown, ideophones in German have multisensory tendencies (Nuckolls, 2019) and can aid children in acquiring the language (Imai and Kita, 2014).

Crucially, to resonate with the goals for writing this thesis, I have also shown that iconicity wears different gowns. We can find form–meaning correspondences in the written language, and on the sound- and the prosodic level as well. Thus, iconic expressions come in all shapes and colors; they can be language-specific (e.g., Shih et al., 2019, or Chapter 2 of this thesis), but also cross-linguistic (e.g., Ćwiek et al., 2021). However, most importantly, we still find them in the language today. Iconicity is needed; therefore, it perseveres in us and our communication despite the changes that our communication undergoes (Flaksman, 2017).

The results of this thesis indicate that iconicity is an important property of language not only from the diachronic but also from the synchronic perspective. Iconicity may seem obscured; however, it employs a multitude of communicative channels to come to light. It is extensively used and perceived – without a convention but instead based on an inherent mechanism that we share. Iconicity mirrors parts of the sensory perception and benefits our development. Rather than being an outside discipline, iconicity constitutes the core of communication.

References

Aaker, Jennifer Lynn (1997). "Dimensions of Brand Personality". *SSRN Electronic Journal* 34.3, 347–356. DOI: 10.2139/ssrn.945432.

Abralin (2021). *Bodo Winter "Iconicity, not arbitrariness, is a design feature of language"*. https://youtu.be/R1ETw21oCGE.

Acredolo, Linda and Susan Goodwyn (1996). *Baby Signs: How to Talk with Your Baby Before Your Baby Can Talk*. Chicago, Illinois: Contemporary Books.

Akita, Kimi (2009). "A Grammar of Sound-Symbolic Words in Japanese: Theoretical Approaches to Iconic and Lexical Properties of Mimetics". PhD Thesis. Kobe, Japan: Kobe University.

Akita, Kimi (2015). "Sound symbolism". *Handbook of Pragmatics*. Ed. by Jan-Ola Östman and Jef Verschueren. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/hop.19.sou1.

Akita, Kimi (2017). "The linguistic integration of Japanese ideophones and its typological implications". *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62.2, 314–334. DOI: 10.1017/cnj.2017.6.

Akita, Kimi and Mark Dingemanse (2018). "Ideophones (Mimetics, Expressives)". *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 17.

Alderete, John and Alexei Kochetov (2017). "Integrating sound symbolism with core grammar: The case of expressive palatalization". *Language* 93.4, 731–766. DOI: 10.1353/lan.2017.0056.

Alpher, Barry (1994). "Yir-Yoront ideophones". *Sound Symbolism*. Ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala. Cambridge University Press, 161–177.

Ameka, Felix K. (2006). "Interjections". *Encyclopedia of Language and Linguistics*. 2nd Edition. Oxford: Elsevier, 743–746.

Anderson, Rindy C., Casey A. Klofstad, William J. Mayew, and Mohan Venkatachalam (2014). "Vocal Fry May Undermine the Success of Young Women in the Labor Market". *PLOS ONE* 9.5. Ed. by Joel Snyder, e97506. DOI: 10.1371/journal.pone.0097506.

Anderson, Rindy C., William A. Searcy, Melissa Hughes, and Stephen Nowicki (2012). "The receiver-dependent cost of soft song: a signal of aggressive intent in songbirds". *Animal Behaviour* 83.6, 1443–1448. DOI: 10.1016/j.anbehav.2012.03.016.

Anokhin, Peter K. (1974). *Biology and Neurophysiology of the Conditioned Reflex and Its Role in Adaptive Behavior: International Series of Monographs in Cerebrovisceral* *and Behavioral Physiology and Conditioned Reflexes, Volume 3*. Oxford, New York: Pergamon Press.

Apicella, C.L, D.R Feinberg, and F.W Marlowe (2007). "Voice pitch predicts reproductive success in male hunter-gatherers". *Biology Letters* 3.6, 682–684. DOI: 10.1098/rsbl.2007.0410.

Arbib, Michael A., Katja Liebal, and Simone Pika (2008). "Primate Vocalization, Gesture, and the Evolution of Human Language". *Current Anthropology* 49.6, 1053–1076. DOI: 10.1086/593015.

Argo, Jennifer J., Monica Popa, and Malcolm C. Smith (2010). "The Sound of Brands". *Journal of Marketing* 74.4, 97–109.

Arnott, Gareth and Robert W. Elwood (2010). "Signal residuals and hermit crab displays: flaunt it if you have it!" *Animal Behaviour* 79.1, 137–143. DOI: 10.1016/j.anbehav.2009.10.011.

Arrangoiz, Rodrigo, Fernando Cordera, David Caba, Manuel Muñoz, Eduardo Moreno, and Enrique Luque de León (2018). "Comprehensive Review of Thyroid Embryology, Anatomy, Histology, and Physiology for Surgeons". *International Journal of Otolaryngology and Head & Neck Surgery* 7.4, 160–188. DOI: 10.4236/ijohns.2018.74019.

Asano, Michiko, Mutsumi Imai, Sotaro Kita, Keiichi Kitajo, Hiroyuki Okada, and Guillaume Thierry (2015). "Sound symbolism scaffolds language development in preverbal infants". *Cortex* 63, 196–205.

Ashley, Richard (2004). "Musical Pitch Space across Modalities: Spatial and Other Mappings through Language and Culture". *Proceedings of the 8th International Conference on Music Perception and Cognition*. Ed. by S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster. Vol. 72. Adelaide, Australia: Causal Productions, 64–71.

Athaide, Gerard A. and Richard R. Klink (2012). "Creating Global Brand Names: The Use of Sound Symbolism". *Journal of Global Marketing* 25.4, 202–212. DOI: 10.1080/08911762.2012.744123.

Bańko, Miros\law (2008). *Współczesny polski onomatopeikon. Ikoniczność w języku*. Warszawa: PWN.

Barnes, Susan Kubic (2010). "Sign Language With Babies: What Difference Does It Make?" *Dimensions of Early Childhood* 38.1, 21–30.

Baron-Cohen, Simon, Lucy Burt, Fiona Smith-Laittan, John Harrison, and Patrick Bolton (1996). "Synaesthesia: Prevalence and Familiality". *Perception* 25.9, 1073–1079. DOI: 10.1068/p251073.

Baron-Cohen, Simon and John E. Harrison, eds. (1997). *Synaesthesia: Classic and contemporary readings*. Synaesthesia: Classic and contemporary readings. Malden: Blackwell Publishing.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of Memory and Language* 68.3. DOI: 10.1016/j.jml.2012.11.001.

Barre, Arnaud and Stéphane Armand (2014). "Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data". *Computer Methods and Programs in Biomedicine* 114.1, 80–87. DOI: 10.1016/j.cmpb.2014.01.012.

Barth, F. G. (1997). "Vibratory communication in spiders: Adaptation and compromise at many levels". *Orientation and Communication in Arthropods*. Ed. by Miriam Lehrer. EXS. Basel: Birkhäuser, 247–272. DOI: 10.1007/978-3-0348-8878-3_9.

Barthlott, W., J. Szarzynski, P. Vlek, W. Lobin, and N. Korotkova (2009). "A torch in the rain forest: thermogenesis of the Titan arum (Amorphophallus titanum)". *Plant Biology* 11.4, 499–505. DOI: 10.1111/j.1438-8677.2008.00147.x.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen (2015). "Parsimonious Mixed Models". *arXiv:1506.04967*.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2014). "Fitting Linear Mixed-Effects Models using lme4". *arXiv:1406.5823*.

Bavelas, Janet, Jennifer Gerwing, Chantelle Sutton, and Danielle Prevost (2008). "Gesturing on the telephone: Independent effects of dialogue and visibility". *Journal of Memory and Language* 58.2, 495–520. DOI: 10.1016/j.jml.2007.02.004.

Bee, Mark A., Stephen A. Perrill, and Patrick C. Owen (2000). "Male green frogs lower the pitch of acoustic signals in defense of territories: a possible dishonest signal of size?" *Behavioral Ecology* 11.2, 169–177. DOI: 10.1093/beheco/11.2.169.

Bentley, Madison and Edith J. Varon (1933). "An accessory study of "phonetic symbolism"". *The American Journal of Psychology* 45.1, 76–86. DOI: 10.2307/1414187.

Bergen, Benjamin K. (2004). "The Psychological Reality of Phonaesthemes". *Language* 80.2, 290–311. DOI: 10.1353/lan.2004.0056.

Bergman, Thore J. (2013). "Speech-like vocalized lip-smacking in geladas". *Current Biology* 23.7, R268–R269. DOI: 10.1016/j.cub.2013.02.038.

Berlin, Brent (1992). *Ethnobiological Classification*. Princeton University Press. DOI: 10.1515/9781400862597.

Berlin, Brent (2006). "The First Congress of Ethnozoological Nomenclature". *Journal of the Royal Anthropological Institute* 12.1, 23–44. DOI: 10.1111/j.1467-9655.2006.00271.x.

Bernstein, Ira H. and Barry A. Edelstein (1971). "Effects of some variations in auditory input upon visual choice reaction time". *Journal of Experimental Psychology* 87.2, 241–247. DOI: 10.1037/h0030524.

Biguer, B., M. Jeannerod, and C. Prablanc (1982). "The coordination of eye, head, and arm movements during reaching at a single visual target". *Experimental Brain Research* 46.2, 301–304. DOI: 10.1007/BF00237188.

Blasi, Damián E., Søren Wichmann, Harald Hammarström, Peter F. Stadler, and Morten H. Christiansen (2016). "Sound–meaning association biases evidenced across thousands of languages". *Proceedings of the National Academy of Sciences* 113.39, 10818–10823. DOI: 10.1073/pnas.1605782113.

Blauert, Jens (1969). "Sound Localization in the Median Plane". *Acustica* 22, 205–213.

Blom, Jan-Petter and John J. Gumperz (2000). “Social meaning in linguistic structure: code-switching in Norway”. *The Bilingualism Reader*. Ed. by Li Wei. 1st ed. New York, NY, USA: Routledge, 111–136.

Bodomo, Adams (2006). “The Structure of Ideophones in African and Asian Languages: The Case of Dagaare and Cantonese”. *Selected Proceedings of the 35th Annual Conference on African Linguistics: African Languages and Linguistics in Broad Perspectives*. Ed. by John Mugane. Cambridge, MA: Harvard University, 202–213.

Boersma, Paul and David Weenik (2018). *Praat: doing phonetics by computer [Computer program]*. http://www.praat.org/.

Bolinger, Dwight (1978). “Intonation across Languages”. *Phonology*. Ed. by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik. Vol. 2. Universals of Human Language. Stanford: Stanford University Press, 471–524.

Bolinger, Dwight (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford, California: Stanford University Press.

Bosker, Hans Rutger and David Peeters (2021). “Beat gestures influence which speech sounds you hear”. *Proceedings of the Royal Society B: Biological Sciences* 288.1943, 20202419. DOI: 10.1098/rspb.2020.2419.

Bottaro, Luciano, Carlo Chendi, and Giovan Battista Carpi (1997). “Die Erbschaft”. *Lustiges Taschenbuch - Die Piraten des gelben Meeres*. Ed. by Marco Andric. Vol. 9. Lustiges Taschenbuch. Egmont Ehapa Media, 77–156.

Bowling, D. L., M. Garcia, J. C. Dunn, R. Ruprecht, A. Stewart, K.-H. Frommolt, and W. T. Fitch (2017). “Body size and vocalization in primates and carnivores”. *Scientific Reports* 7.1, 41070. DOI: 10.1038/srep41070.

Brackbill, Yvonne and Kenneth B. Little (1957). “Factors determining the guessing of meanings of foreign words”. *The Journal of Abnormal and Social Psychology* 54.3, 312–318. DOI: 10.1037/h0042411.

Bremner, Andrew J., Karina Linnell, Serge Caparos, Charles Spence, Jan W. de Fockert, and Jules Davidoff (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. DOI: 10.1163/22134808-000s0089.

Brown, Roger W., Abraham H. Black, and Arnold E. Horowitz (1955). “Phonetic symbolism in natural languages”. *The Journal of Abnormal and Social Psychology* 50.3, 388–393. DOI: 10.1037/h0046820.

Brumm, Henrik, Rouven Schmidt, and Lars Schrader (2009). “Noise-dependent vocal plasticity in domestic fowl”. *Animal Behaviour* 78.3, 741–746. DOI: 10.1016/j.anbehav.2009.07.004.

Bücking, Sebastian and Jennifer Rau (2013). “German non-inflectional constructions as separate performatives”. *Beyond Expressives: Explorations in Use-Conditional Meaning*, 59–94. DOI: 10.1163/9789004183988_003.

Bulbagarden (2004). *Bulbapedia, the community-driven Pokémon encyclopedia*.

Bürkner, Paul-Christian (2017). "brms: An R Package for Bayesian Multilevel Models Using Stan". *Journal of Statistical Software* 80.1, 1–28. DOI: 10.18637/jss.v080.i01.

Bürkner, Paul-Christian (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms". *The R Journal* 10.1, 395. DOI: 10.32614/RJ-2018-017.

Busch, Wilhelm (1865). *Max und Moritz*. München, Germany: Braun und Schneider.

Bussmann, Hadumod (2006). *The Routledge Dictionary of Language and Linguistics*. 2nd. New York: Taylor & Francis.

Butterworth, Brian and Geoffrey Beattie (1978). "Gesture and Silence as Indicators of Planning in Speech". *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*. Ed. by Robin N. Campbell and Philip T. Smith. NATO Conference Series. Boston, MA: Springer US, 347–360. DOI: 10.1007/978-1-4684-2532-1_19.

Byrne, Richard W. and Nadia Corp (2004). "Neocortex size predicts deception rate in primates". *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271.1549, 1693–1699. DOI: 10.1098/rspb.2004.2780.

Caldwell, Christine A. and Kenny Smith (2012). "Cultural Evolution and Perpetuation of Arbitrary Communicative Conventions in Experimental Microsocieties". *PLOS ONE* 7.8, e43807. DOI: 10.1371/journal.pone.0043807.

Calude, Andreea S. and Mark Pagel (2011). "How do we use language? Shared patterns in the frequency of word use across 17 world languages". *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1567, 1101–1107. DOI: 10.1098/rstb.2010.0315.

Calvert, Gemma, Charles Spence, and Barry E. Stein, eds. (2004). *The handbook of multisensory processes*. Cambridge, Mass: MIT Press.

Casasanto, Daniel, Webb Phillips, and Lera Boroditsky (2003). "Do we think about music in terms of space? Metaphoric representation of musical pitch". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 25. Boston, MA: Cognitive Science Society, 2.

Cassidy, Kimberly Wright, Michael H Kelly, and Lee'at J Sharoni (1999). "Inferring Gender From Name Phonology". *Journal of Experimental Psychology: General* 128.3, 362–381.

Chafe, Wallace (1988). "Punctuation and the Prosody of Written Language". *Written Communication* 5.4, 395–426. DOI: 10.1177/0741088388005004001.

Chauvet, Jean-Marie, Eliette Brunel Deschamps, and Christian Hillaire (1996). *Dawn of art: the Chauvet Cave: the oldest known paintings in the world*. New York: H.N. Abrams.

Chen, Trevor H. and Dominic W. Massaro (2008). "Seeing pitch: Visual information for lexical tones of Mandarin-Chinese". *The Journal of the Acoustical Society of America* 123.4, 2356–2366. DOI: 10.1121/1.2839004.

Cheney, Dorothy L. and Robert M. Seyfarth (1980). "Vocal recognition in free-ranging vervet monkeys". *Animal Behaviour* 28.2, 362–367. DOI: 10.1016/S0003-3472(80)80044-3.

Childs, George (1988). "The Phonology of Kisi Ideophones". *Journal of African Languages and Linguistics*.

Clark, Nathaniel, Marcus Perlman, and Marlene Johansson Falck (2014). "Iconic Pitch Expresses Vertical Space". *Language and the creative mind*. Ed. by Mike Borkent, Barbara Dancygier, and Jennifer Hinnell. Stanford: SCLI Publications, 393–410.

Condillac, Etienne (2001). *Essay on the origin of human knowledge*. Trans. by Hans Aasleff. Cambridge, UK: Cambridge University Press.

Corballis, Michael C. (2002). *From hand to mouth: the origins of language*. Princeton, New Jersey: Princeton University Press.

Crockford, Catherine, Roman M. Wittig, Roger Mundry, and Klaus Zuberbühler (2012). "Wild Chimpanzees Inform Ignorant Group Members of Danger". *Current Biology* 22.2, 142–146. DOI: 10.1016/j.cub.2011.11.053.

Crystal, David (1995). "Phonaesthetically speaking". *English Today* 11.02, 8. DOI: 10.1017/s026607840000818x.

Cuddy, Lauren J. and Larry L. Jacoby (1982). "When forgetting helps memory: an analysis of repetition effects". *Journal of Verbal Learning and Verbal Behavior* 21.4, 451–467. DOI: 10.1016/S0022-5371(82)90727-7.

Cunnington, Glenn M. and Lenore Fahrig (2010). "Plasticity in the vocalizations of anurans in response to traffic noise". *Acta Oecologica* 36.5, 463–470. DOI: 10.1016/j.actao.2010.06.002.

Cuskley, Christine and Simon Kirby (2013). "Synesthesia, cross-modality, and language evolution". *Oxford Handbook of Synesthesia*. Ed. by Julia Simner and Edward Hubbard. Oxford Handbooks Online. New York, NY, USA: Oxford University Press, 869–907. DOI: 10.1093/oxfordhb/9780199603329.013.0043.

Cutler, Anne, James McQueen, and Ken Robinson (1990). "Elizabeth and John: sound patterns of men's and women's names". *Journal of Linguistics* 26.02, 471. DOI: 10.1017/s0022226700014754.

Ćwiek, Aleksandra and Susanne Fuchs (2019). "Iconic prosody is rooted in sensorimotor properties: fundamental frequency and the vertical space". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 41. Montreal, Canada: Cognitive Science Society, 1572–1578.

Ćwiek, Aleksandra and Susanne Fuchs (2021). "Hand-Mouth Coordination in a Pointing Task Requiring Manual Precision". *Proceedings of the 12th ISSP*. New Haven, CT: Haskins Laboratories.

Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, Sofia Koutalidis, Manfred Krifka, Pärtel Lippus, Gary Lupyan, Grace E. Oh, Jing Paul, Caterina Petrone, Rachid Ridouane, Sabine Reiter, Nathalie Schümchen, Ádám Szalontai, Özlem Ünal-Logacev, Jochen Zeller, Marcus Perlman, and Bodo Winter (2022). "The bouba/kiki effect is robust across cultures and writing systems". *Philosophical Transactions of the Royal Society B: Biological Sciences* 377.1841, 20200390. DOI: 10.1098/rstb.2020.0390.

Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, Sofia Koutalidis, Manfred Krifka, Pärtel Lippus, Gary Lupyan, Grace E. Oh, Jing Paul, Caterina Petrone, Rachid Ridouane, Sabine Reiter, Nathalie Schümchen, Ádám Szalontai, Özlem Ünal-Logacev, Jochen Zeller, Bodo Winter, and Marcus Perlman (2021). "Novel vocalizations are understood across cultures". *Scientific Reports* 11.1, 10108. DOI: 10.1038/s41598-021-89445-4.

Ćwiek, Aleksandra and Petra Wagner (2018). "The Acoustic Realization of Prosodic Prominence in Polish: Word-level Stress and Phrase-level Accent". *9th International Conference on Speech Prosody 2018*. ISCA, 922–926. DOI: 10.21437/SpeechProsody.2018-186.

D'Anselmo, Anita, Giulia Prete, Przemysław Zdybek, Luca Tommasi, and Alfredo Brancucci (2019). "Guessing meaning from word sounds of unfamiliar languages: a cross-cultural sound symbolism study". *Frontiers in Psychology* 10. DOI: 10.3389/fpsyg.2019.00593.

D'Onofrio, Annette (2014). "Phonetic detail and dimensionality in sound-shape correspondences: refining the *bouba-kiki* paradigm". *Language and Speech* 57.3, 367–393. DOI: 10.1177/0023830913507694.

D'Onofrio, Annette and Penelope Eckert (2020). "Affect and iconicity in phonological variation". *Language in Society* 50, 29–51. DOI: 10.1017/S0047404520000871.

Darwin, Charles (2009a). *On the Origin of Species (first published 1859)*. Ed. by William Bynum. London: Penguin Books.

Darwin, Charles (2009b). *The Descent of Man and Selection in Relation to Sex*. This edition first published 1871. New York: Cambridge University Press. DOI: 10.1017/cbo9780511703829.

Davis, Charles P., Hannah M. Morrow, and Gary Lupyan (2019). "What Does a Horgous Look Like? Nonsense Words Elicit Meaningful Drawings". *Cognitive Science* 43.10, e12791. DOI: 10.1111/cogs.12791.

De Carolis, Léa, Egidio Marsico, Vincent Arnaud, and Christophe Coupé (2018). "Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm". *PLOS ONE* 13.12, e0208874. DOI: 10.1371/journal.pone.0208874.

Deely, John (2006). "Semiotics: History". *Encyclopedia of Language and Linguistics*. 2nd Edition. Oxford: Elsevier, 216–229.

Dehaene, Stanislas, Serge Bossini, and Pascal Giraux (1993). "The mental representation of parity and number magnitude." *Journal of Experimental Psychology: General* 122.3, 371–396. DOI: 10.1037//0096-3445.122.3.371.

Deroy, Ophelia and Charles Spence (2013). "Why we are not all synesthetes (not even weakly so)". *Psychonomic Bulletin & Review* 20.4, 643–664. DOI: 10.3758/s13423-013-0387-2.

Dickens, Charles (1838). *Oliver Twist; or, The Parish Boy's Progress*. London: Richard Bentley, New Burlingtion Street.

Diffloth, Gérard (1972). "Notes on expressive meaning". *Chicago Linguistic Society 8*. Chicago, Illinois, 440–447.

Diffloth, Gérard (1994). "i: big, a: small". *Sound Symbolism*. Ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala. Cambridge: Cambridge University Press, 107–114.

Dingemanse, Mark (2011). "The Meaning and Use of Ideophones in Siwu". PhD Thesis. Nijmegen: Radboud University Nijmegen.

Dingemanse, Mark (2012). "Advances in the Cross-Linguistic Study of Ideophones". *Language and Linguistics Compass* 6.10, 654–672. DOI: 10.1002/lnc3.361.

Dingemanse, Mark (2013). "Ideophones and gesture in everyday speech". *Gesture* 13.2, 143–165. DOI: 10.1075/gest.13.2.02din.

Dingemanse, Mark (2015). "Ideophones and reduplication: Depiction, description, and the interpretation of repeated talk in discourse". *Studies in Language* 39.4, 946–970. DOI: 10.1075/sl.39.4.05din.

Dingemanse, Mark (2017a). "Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu". *STUF – Language Typology and Universals* 70.2, 363–384. DOI: 10.1515/stuf-2017-0018.

Dingemanse, Mark (2017b). "On the margins of language: Ideophones, interjections and dependencies in linguistic theory". *Dependencies in language*. Ed. by N. J. Enfield. Zenodo, 195–202. DOI: 10.5281/ZENODO.573781.

Dingemanse, Mark (2019). "Chapter 1. 'Ideophone' as a comparative concept". *Iconicity in Language and Literature*. Ed. by Kimi Akita and Prashant Pardeshi. Vol. 16. Amsterdam: John Benjamins Publishing Company, 13–33. DOI: 10.1075/ill.16.02din.

Dingemanse, Mark (2021). *Ideophones*. Tech. rep. PsyArXiv. DOI: 10.31234/osf.io/u96zt.

Dingemanse, Mark and Kimi Akita (2016). "An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese". *Journal of Linguistics* 53.03, 501–532. DOI: 10.1017/s002222671600030x.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen, and Padraic Monaghan (2015). "Arbitrariness, iconicity, and systematicity in language". *Trends in Cognitive Sciences* 19.10, 603–615. DOI: 10.1016/j.tics.2015.07.013.

Dittmann, Allen T. and Lynn G. Llewellyn (1969). "Body movement and speech rhythm in social conversation". *Journal of Personality and Social Psychology* 11.2, 98–106. DOI: 10.1037/h0027035.

Dolscheid, Sarah and Daniel Casasanto (2015). "Spatial Congruity Effects Reveal Metaphorical Thinking, not Polarity Correspondence". *Frontiers in Psychology* 6. DOI: 10.3389/fpsyg.2015.01836.

Dolscheid, Sarah, Sabine Hunnius, Daniel Casasanto, and Asifa Majid (2014a). "Prelinguistic Infants Are Sensitive to Space-Pitch Associations Found Across Cultures". *Psychological Science* 25.6, 1256–1261. DOI: 10.1177/0956797614528521.

Dolscheid, Sarah, Shakila Shayan, Asifa Majid, and Daniel Casasanto (2013). "The thickness of musical pitch". *Psychological Science* 24.5, 613–621. DOI: 10.1177/0956797612457374.

Dolscheid, Sarah, Roel M. Willems, Peter Hagoort, and Daniel Casasanto (2014b). "The relation of space and musical pitch in the brain". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 36. Quebec City, Canada: Cognitive Science Society, 421–426.

Dudenredaktion (n.d.). *Duden online*.

Dunbar, Robin (1998). *Grooming, Gossip, and the Evolution of Language*. Harvard University Press.

DWDS – Digitales Wörterbuch der deutschen Sprache (2021).

Edmiston, Pierce, Marcus Perlman, and Gary Lupyan (2018). "Repeated imitation makes human vocalizations more word-like". *Proceedings of the Royal Society B: Biological Sciences* 285.1874, 20172709. DOI: 10.1098/rspb.2017.2709.

Elsen, Hilke (1991). "Erstspracherwerb. Der Erwerb des deutschen Lautsystems". PhD Thesis. München, Deutschland: Ludwig-Maximilians-Universität München.

Elsen, Hilke (2005). "Das Kunstwort". *Muttersprache: Vierteljahresschrift für deutsche Sprache* 2, 142–149. DOI: 10.5282/ubm/epub.14674.

Elsen, Hilke (2006). "Pseudomorpheme – Fiktive Namen im Übergangsbereich von Phonologie und Morphologie". *Muttersprache: Vierteljahresschrift für deutsche Sprache* 116.3, 242. DOI: 10.5282/ubm/epub.14675.

Elsen, Hilke (2014). "Lautsymbolik – ein vernachlässigter Forschungsgegenstand der Sprachwissenschaft". *Glottotheory* 5.2, 185–218. DOI: 10.1515/glot-2014-0016.

Elsen, Hilke (2015). "Der Faktor Lautsymbolik". *JournaLIPP* 4, 27–42.

Endress, Ansgar D., Marina Nespor, and Jacques Mehler (2009). "Perceptual and memory constraints on language acquisition". *Trends in Cognitive Sciences* 13.8, 348–353. DOI: 10.1016/j.tics.2009.05.005.

Enquist, Magnus (1985). "Communication during aggressive interactions with particular reference to variation in choice of behaviour". *Animal Behaviour* 33.4, 1152–1161. DOI: 10.1016/S0003-3472(85)80175-5.

Erben Johansson, Niklas, Jon W Carr, and Simon Kirby (2021). "Cultural evolution leads to vocal iconicity in an experimental iterated learning task". *Journal of Language Evolution* 6.1, 1–25. DOI: 10.1093/jole/lzab001.

Erickson, Donna, Thomas Baer, and Katherine S. Harris (1982). "The role of the strap muscles in pitch lowering". *Haskins Laboratories: Status Report on Speech Research* 70, 275–284.

Ernst, Marc O. (2006). "A Bayesian view on multimodal cue integration". *Human body perception from the inside out*. Ed. by Günther Knoblich, M. Grosjean, I. Thornton, and M. Shiffrar. Oxford, United Kingdom: Oxford University Press, 105–131.

Ervin-Tripp, Susan (2002). "Variety, style-shifting, and ideology". *Style and Sociolinguistic Variation*. Ed. by Penelope Eckert and John R. Rickford. 1st ed. Cambridge, UK: Cambridge University Press, 44–56. DOI: 10.1017/CBO9780511613258.003.

Evans, Karla K. and Anne Treisman (2010). "Natural cross-modal mappings between visual and auditory features". *Journal of Vision* 10.1, 6–6. DOI: 10.1167/10.1.6.

Everett, Daniel (2017). *How Language Began: The Story of Humanity's Greatest Invention*. New York: Liveright.

Fant, Gunnar (1970). *Acoustic Theory of Speech Production*. Walter de Gruyter.

Fay, Nicolas, Michael Arbib, and Simon Garrod (2013). "How to Bootstrap a Human Communication System". *Cognitive Science* 37.7, 1356–1367. DOI: 10.1111/cogs.12048.

Fay, Nicolas and Stephanie Lim (2010). "From hand to mouth: an experimental simulation of langauge origin". *The Evolution of Language*. WORLD SCIENTIFIC, 401–402. DOI: 10.1142/9789814295222_0065.

Fay, Nicolas, Casey J. Lister, T. Mark Ellison, and Susan Goldin-Meadow (2014). "Creating a communication system from scratch: gesture beats vocalization hands down". *Frontiers in Psychology* 5. DOI: 10.3389/fpsyg.2014.00354.

Feinberg, D. R., B. C. Jones, L. M. DeBruine, J. J. M. O'Connor, C. C. Tigue, and D. J. Borak (2011). "Integrating fundamental and formant frequencies in women's preferences for men's voices". *Behavioral Ecology* 22.6, 1320–1325. DOI: 10.1093/beheco/arr134.

Feinberg, D. R., B. C. Jones, A. C. Little, D. M. Burt, and D. I. Perrett (2005). "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices". *Animal Behaviour* 69.3, 561–568. DOI: 10.1016/j.anbehav.2004.06.012.

Fernandes, Lincoln (2006). "Translation of Names in Children's Fantasy Literature: Bringing the Young Reader into Play". *New Voices in Translation Studies* 2, 44–57.

Fetscherin, Marc, Adamantios Diamantopoulos, Allan Chan, and Rachael Abbott (2015). "How are brand names of Chinese companies perceived by Americans?" *Journal of Product & Brand Management* 24.2, 110–123. DOI: 10.1108/JPBM-02-2014-0501.

Firth, John R. (1935). "The use and distribution of certain English sounds". *English Studies* 17.1-6, 8–18. DOI: 10.1080/00138383508596629.

Fischer, Julia and Tabitha Price (2017). "Meaning, intention, and inference in primate vocal communication". *Neuroscience & Biobehavioral Reviews*. An overview of nonhuman primates' communication and social abilities through behavioral and neuroscientific approaches 82, 22–31. DOI: `10.1016/j.neubiorev.2016.10.014`.

Fitch, W. Tecumseh (1997). "Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques". *The Journal of the Acoustical Society of America* 102.2, 1213–1222. DOI: `10.1121/1.421048`.

Fitch, W. Tecumseh (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. DOI: `10.1017/cbo9780511817779`.

Fitch, W. Tecumseh (2017). "Empirical approaches to the study of language evolution". *Psychonomic Bulletin & Review* 24.1, 3–33. DOI: `10.3758/s13423-017-1236-5`.

Fitch, W. Tecumseh and Jay Giedd (1999). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging". *The Journal of the Acoustical Society of America* 106.3, 1511–1522. DOI: `10.1121/1.427148`.

Fitch, William Tecumseh Sherman (1994). "Vocal Tract Length Perception and the Evolution of Language". PhD Thesis. Brown University.

Flaksman, Maria (2017). "Iconic treadmill hypothesis: the reasons behind continuous onomatopoeic coinage". *Dimensions of Iconicity*. Ed. by Angelika Zirker, Matthias Bauer, Olga Fischer, and Christina Ljungberg. Vol. 15. Iconicity in Language and Literature. Amsterdam: John Benjamins Publishing Company, 15–38.

Flaksman, Maria (2020). "Pathways of de-iconization". *Iconicity in Language and Literature 17*. Ed. by Pamela Perniss, Olga Fischer, and Christina Ljungberg. Amsterdam: John Benjamins Publishing Company, 76–103.

Forstmeier, Wolfgang, Claudia Burger, Katja Temnow, and Sébastien Derégnaucourt (2009). "The Genetic Basis of Zebra Finch Vocalizations". *Evolution* 63.8, 2114–2130. DOI: `https://doi.org/10.1111/j.1558-5646.2009.00688.x`.

Fort, Mathilde, Imme Lammertink, Sharon Peperkamp, Adriana Guevara-Rukoz, Paula Fikkert, and Sho Tsuji (2018). "SymBouki: a meta-analysis on the emergence of sound symbolism in early language acquisition". *Developmental Science* 21.5, e12659. DOI: `10.1111/desc.12659`.

Freeman, Tor (2016). *Pino Pfote - Ab die Post!* 1. Bamberg: Magellan.

Frishberg, Nancy (1975). "Arbitrariness and Iconicity: Historical Change in American Sign Language". *Language* 51.3, 696–719. DOI: `10.2307/412894`.

Fröhlich, Marlen, Christine Sievers, Simon W. Townsend, Thibaud Gruber, and Carel P. van Schaik (2019). "Multimodal communication and language origins: integrating gestures and vocalizations". *Biological Reviews* 94.5, 1809–1829. DOI: `10.1111/brv.12535`.

Fuchs, Susanne, Egor Savin, Stephanie Solt, Cornelia Ebert, and Manfred Krifka (2019). "Antonym adjective pairs and prosodic iconicity: evidence from letter replications in an English blogger corpus". *Linguistics Vanguard* 5.1. DOI: `10.1515/lingvan-2018-0017`.

Gallace, Alberto and Charles Spence (2006). "Multisensory synesthetic interactions in the speeded classification of visual size". *Perception & Psychophysics* 68.7, 1191–1203. DOI: 10.3758/BF03193720.

Gardner, R. Allen and Beatrice T. Gardner (1969). "Teaching Sign Language to a Chimpanzee". *Science* 165.3894, 664–672.

Garrod, Simon, Nicolas Fay, John Lee, Jon Oberlander, and Tracy MacLeod (2007). "Foundations of Representation: Where Might Graphical Symbol Systems Come From?" *Cognitive Science* 31.6, 961–987. DOI: 10.1080/03640210701703659.

Gasser, Michael (2004). "The Origins of Arbitrariness in Language". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 26. Chicago, Illinois: Cognitive Science Society.

Gebels, Gustav (1969). "An investigation of phonetic symbolism in different cultures". *Journal of Verbal Learning and Verbal Behavior* 8.2, 310–312. DOI: 10.1016/S0022-5371(69)80083-6.

Gee, James Paul (1986). "Review of Orality and Literacy: From The Savage Mind to Ways with Words". *TESOL Quarterly* 20.4, 719–746. DOI: 10.2307/3586522.

Gentilucci, Maurizio and Riccardo Dalla Volta (2007). "The motor system and the relationships between speech and gesture". *Gesture* 7.2, 159–177. DOI: 10.1075/gest.7.2.03gen.

Gerwing, Jennifer and Meredith Allison (2011). "The flexible semantic integration of gestures and words: Comparing face-to-face and telephone dialogues". *Gesture* 11.3, 308–329. DOI: 10.1075/gest.11.3.03ger.

Ghazanfar, Asif A. and Charles E. Schroeder (2006). "Is neocortex essentially multisensory?" *Trends in Cognitive Sciences* 10.6, 278–285. DOI: 10.1016/j.tics.2006.04.008.

Ghazanfar, Asif A., Daniel Y. Takahashi, Neil Mathur, and W. Tecumseh Fitch (2012). "Cineradiography of Monkey Lip-Smacking Reveals Putative Precursors of Speech Dynamics". *Current Biology* 22.13, 1176–1182. DOI: 10.1016/j.cub.2012.04.055.

Gick, Bryan, Ian Wilson, and Donald Derrick (2013). *Articulatory Phonetics*. Wiley-Blackwell.

Gingras, B., M. Boeckle, C. T. Herbst, and W. T. Fitch (2013). "Call acoustics reflect body size across four clades of anurans". *Journal of Zoology* 289.2, 143–150. DOI: 10.1111/j.1469-7998.2012.00973.x.

Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, Illinois: University of Chicago Press.

Grice, H. P. (1957). "Meaning". *The Philosophical Review* 66.3, 377. DOI: 10.2307/2182440.

Grimm, Jacob and Wilhelm Grimm (2013). *Kinder- und Hausmärchen*. Berliner Ausgabe. North Charleston, USA: Holzinger.

Guasparri, Andrea (2019). "Polysemy Revisited". *Altorientalische Forschungen* 46.1, 61–87. DOI: 10.1515/aofo-2019-0006.

Gussenhoven, Carlos and Aoju Chen (2000). "Universal and Language-Specific Effects in the Perception of Question Intonation". *ICSLP 2000-Beijing*. Beijing.

Haiman, John (1994). "Ritualization and the Development of Language". *Perspectives on Grammaticalization*. Ed. by William Pagliuca. Amsterdam Studies in the Theory and History of Linguistic Science. Series 4, Current Issues in Linguistic Theory. Amsterdam, The Netherlands: John Benjamins Publishing Company, 9–28.

Hamano, Shoko (1998). *The Sound-Symbolic System of Japanese*. Masayoshi Shibatani. Studies in Japanese Linguistics 10. CSLI Publications.

Havlik, Ernst (1981). *Lexikon der Onomatopöien: die lautimitierenden Wörter im Comic*. Frankfurt am Main: Fricke.

Hays, Terence E. (1994). "Sound Symbolism, Onomatopoeia, and New Guinea Frog Names". *Journal of Linguistic Anthropology* 4.2, 153–174. DOI: `https://doi.org/10.1525/jlin.1994.4.2.153`.

Hermann, Eduard (1942). *Probleme der Frage*. Vol. 1. Nachrichten von der Akademie der Wissenschaften in Göttingen. Göttingen: Vandenhoeck & Ruprecht.

Hewes, Gordon W. (1973). "Primate Communication and the Gestural Origin of Language". *Current Anthropology* 14.1/2, 5–24.

Hidaka, Masamitsu (1997). *Pokémon: Indigo League*.

Hinton, Leanne, Birch Moonwomon, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner, and Hazel Corcoran (1987). "It's Not Just the Valley Girls: A Study of California English". *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, 117–128.

Hinton, Leanne, Johanna Nichols, and John J. Ohala (1994). *Sound symbolism*. Ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala. Cambridge: Cambridge University Press.

Hirst, Daniel and Albert Di Cristo (1998). *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.

Hockett, Charles F. (1960). "The origin of speech". *Scientific American* 203.3, 88–96. DOI: `10.1038/scientificamerican0960-88`.

Hockett, Charles F. (1978). "In search of Jove's brow". *American Speech* 53.4, 243–313. DOI: `10.2307/455140`.

Honda, Kiyoshi (1988). "Various laryngeal mechanisms in controlling the voice fundamental frequency". *The Journal of the Acoustical Society of America* 84.S1, S82–S82. DOI: `10.1121/1.2026504`.

Honda, Kiyoshi (1995). "Laryngeal and extra-laryngeal mechanisms of F0 control". *Producing Speech: Contemporary Issues for Katherine Stafford Harris*. Ed. by F. Bell-Berti and L. J. Raphael. New York: AIP Press, 215–232.

Honda, Kiyoshi (1996). "Biological Mechanisms for Tuning Voice Fundamental Frequency". *Koutou (THE LARYNX JAPAN)* 8.2, 109–115. DOI: `10.5426/larynx1989.8.2_109`.

Hoole, Philip, Lasse Bombien, Marianne Pouplier, Christine Mooshammer, and Barbara Kühnert (2012). *Consonant Clusters and Structural Complexity*. Walter de Gruyter.

Hoole, Philip and Christine Mooshammer (2002). "Articulatory analysis of the German vowel system". *Silbenschnitt und Tonakzente*. Ed. by Peter Auer, Peter Gilles, and Helmut Spiekermann. Berlin, Boston: DE GRUYTER. DOI: 10.1515/9783110916447.129.

Hopkins, William D. and E. Sue Savage-Rumbaugh (1991). "Vocal communication as a function of differential rearing experiences in Pan paniscus: A preliminary report". *International Journal of Primatology* 12.6, 559–583. DOI: 10.1007/BF02547670.

Hopkins, William D., Jared P. Taglialatela, and David A. Leavens (2007). "Chimpanzees differentially produce novel vocalizations to capture the attention of a human". *Animal Behaviour* 73.2, 281–286. DOI: 10.1016/j.anbehav.2006.08.004.

Hornibrook, Jeremy, Tika Ormond, and Margaret Maclagan (2018). "Creaky voice or extreme vocal fry in young women". *The New Zealand Medical Journal* 131.1486, 6.

How many words are there in English? | Merriam-Webster (2020).

Huettig, Falk and Ramesh K. Mishra (2014). "How Literacy Acquisition Affects the Illiterate Mind – A Critical Examination of Theories and Evidence". *Language and Linguistics Compass* 8.10, 401–427. DOI: https://doi.org/10.1111/lnc3.12092.

Hughes, Melissa (2000). "Deception with honest signals: signal residuals and signal function in snapping shrimp". *Behavioral Ecology* 11.6, 614–623. DOI: 10.1093/beheco/11.6.614.

Hwan Hong, Ki, Ming Ye, Young Mo Kim, Kevin F. Kevorkian, and Gerald S. Berke (1997). "The role of strap muscles in phonation—In vivo caninelaryngeal model". *Journal of Voice* 11.1, 23–32. DOI: 10.1016/S0892-1997(97)80020-3.

Imai, Mutsumi and Sotaro Kita (2014). "The sound symbolism bootstrapping hypothesis for language acquisition and language evolution". *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1651, 20130298. DOI: 10.1098/rstb.2013.0298.

Imai, Mutsumi, Sotaro Kita, Miho Nagumo, and Hiroyuki Okada (2008). "Sound symbolism facilitates early verb learning". *Cognition* 109.1, 54–65. DOI: 10.1016/j.cognition.2008.07.015.

Imai, Mutsumi, Michiko Miyazaki, H. Henny Yeung, Shohei Hidaka, Katerina Kantartzis, Hiroyuki Okada, and Sotaro Kita (2015). "Sound Symbolism Facilitates Word Learning in 14-Month-Olds". *PLOS ONE* 10.2. Ed. by Andrew Bremner, e0116494. DOI: 10.1371/journal.pone.0116494.

Irvine, Judith T. (2002). ""Style" as distinctiveness: the culture and ideology of linguistic differentiation". *Style and Sociolinguistic Variation*. Ed. by Penelope Eckert and John R. Rickford. Cambridge: Cambridge University Press, 21–43. DOI: 10.1017/CBO9780511613258.002.

Itagaki, Sachi, Hikari Onishi, Shizuko Hiryu, and Kohta I. Kobayasi (2020). "Sound symbolism on synthetic speech continuum in judgment of size of visual stimulus". *Acoustical Science and Technology* 41.1, 413–415. DOI: 10.1250/ast.41.413.

Iwasaki, Noriko, David P. Vinson, and Gabriella Vigliocco (2007a). "How does it hurt, kiri-kiri or siku-siku? Japanese mimetic words of pain perceived by Japanese speakers and English speakers". *Applying theory and research to learning Japanese as a foreign language*. Ed. by Masahiko Minami. Newcastle: Cambridge Scholars Publishing, 2–19.

Iwasaki, Noriko, David P. Vinson, and Gabriella Vigliocco (2007b). "What do English speakers know about gera-gera and yota-yota?: A cross-linguistic investigation of mimetic words for laughing and walking". *Japanese-Language Education around the Globe* 17, 53–78.

Jawad, Hisham A. (2010). "Sound symbolism, schemes & literary translation". *Babel* 56.1, 47–63. DOI: 10.1075/babel.56.1.04jaw.

Jeffreys, Sir Harold (1998). *The Theory of Probability*. Third Edition. Oxford Classic Texts in the Physical Sciences. Oxford, New York: Oxford University Press.

Johansson, Niklas and Jordan Zlatev (2013). "Motivations for sound symbolism in spatial deixis: a typological study of 101 languages". *Public Journal of Semiotics* 5.1, 3–20. DOI: 10.37693/pjos.2013.5.9668.

Johnson, Keith (2011). *Acoustic and Auditory Phonetics*. Wiley-Blackwell.

Jones, Matthew and Gabriella Vigliocco (2017). "Iconicity in Word Learning: What Can We Learn from Cross-Situational". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 39. London, UK: Cognitive Science Society, 2321–2326.

Joordens, Josephine C. A., Francesco d'Errico, Frank P. Wesselingh, Stephen Munro, John de Vos, Jakob Wallinga, Christina Ankjærgaard, Tony Reimann, Jan R. Wijbrans, Klaudia F. Kuiper, Herman J. Mücher, Hélène Coqueugniot, Vincent Prié, Ineke Joosten, Bertil van Os, Anne S. Schulp, Michel Panuel, Victoria van der Haas, Wim Lustenhouwer, John J. G. Reijmer, and Wil Roebroeks (2015). "Homo erectus at Trinil on Java used shells for tool production and engraving". *Nature* 518.7538, 228–231. DOI: 10.1038/nature13962.

Joyce, James (2012). *Finnegans Wake*. Oxford world's classics. Oxford: Oxford University Press.

Jurafsky, Daniel (1996). "Universal Tendencies in the Semantics of the Diminutive". *Language* 72.3, 533–578. DOI: 10.2307/416278.

Jürgens, Andreas, Suk-Ling Wee, Adam Shuttleworth, and Steven D. Johnson (2013). "Chemical mimicry of insect oviposition sites: a global analysis of convergence in angiosperms". *Ecology Letters* 16.9, 1157–1167. DOI: 10.1111/ele.12152.

Kantartzis, Katerina, Mutsumi Imai, Danielle Evans, and Sotaro Kita (2019). "Sound Symbolism Facilitates Long-Term Retention of the Semantic Representation of Novel Verbs in Three-Year-Olds". *Languages* 4.2, 21. DOI: 10.3390/languages4020021.

Kantartzis, Katerina, Mutsumi Imai, and Sotaro Kita (2011). "Japanese Sound-Symbolism Facilitates Word Learning in English-Speaking Children". *Cognitive Science* 35.3, 575–586. DOI: 10.1111/j.1551-6709.2010.01169.x.

Kassambara, Alboukadel and Fabian Mundt (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*.

Kawahara, Shigeto, Mahayana C. Godoy, and Gakuji Kumagai (2020). "Do Sibilants Fly? Evidence from a Sound Symbolic Pattern in Pokémon Names". *Open Linguistics* 6.1, 386–400. DOI: 10.1515/opli-2020-0027.

Kawahara, Shigeto and Gakuji Kumagai (2019). "Expressing evolution in Pokémon names: Experimental explorations". *Journal of Japanese Linguistics* 35.1, 3–38. DOI: 10.1515/jjl-2019-2002.

Kawahara, Shigeto and Gakuji Kumagai (2021). "What voiced obstruents symbolically represent in Japanese: evidence from the Pokémon universe". *Journal of Japanese Linguistics* 37.1, 3–24. DOI: 10.1515/jjl-2021-2031.

Kawahara, Shigeto and Jeff Moore (2021). "How to express evolution in English Pokémon names". *Linguistics*. DOI: 10.1515/ling-2021-0057.

Kawahara, Shigeto, Atsushi Noto, and Gakuji Kumagai (2018). "Sound Symbolic Patterns in Pokémon Name". *Phonetica* 75.2.

Kawahara, Shigeto, Atsushi Noto, and Gakuji Kumagai (n.d.). "Sound (symbolic) patterns in Pokémon names: Focusing on voiced obstruents and mora counts". *PLOS ONE* (), 21.

Kawahara, Shigeto, Michinori Suzuki, and Gakuji Kumagai (2019). "The sound symbolic patterns in Pokémon move names in Japanese", 18.

Kelter, Riko (2020). "Analysis of type I and II error rates of Bayesian and frequentist parametric and nonparametric two-sample hypothesis tests under preliminary assessment of normality". *Computational Statistics*. DOI: 10.1007/s00180-020-01034-7.

Kendon, Adam (1991). "Some Considerations for a Theory of Language Origins". *Man* 26.2, 199–221. DOI: 10.2307/2803829.

Kendon, Adam (2011). "Vocalisation, speech, gesture, and the language origins debate: An essay review on recent contributions". *Gesture* 11.3, 349–370. DOI: 10.1075/gest.11.3.05ken.

Kendon, Adam (2017). "Reflections on the "gesture-first" hypothesis of language origins". *Psychonomic Bulletin & Review* 24.1, 163–170. DOI: 10.3758/s13423-016-1117-3.

Kentner, Gerrit (2017). "On the emergence of reduplication in German morphophonology". *Zeitschrift für Sprachwissenschaft* 36.2, 233–277. DOI: 10.1515/zfs-2017-0010.

Kilian-Hatz, Christa (1999). "Ideophone: Eine typologische Untersuchung unter besonderer Berücksichtigung afrikanischer Sprachen". Habilitation. Köln: Universität zu Köln.

Kilian-Hatz, Christa (2006). "Ideophones". *Encyclopedia of Language and Linguistics*. 2nd Edition. Oxford: Elsevier, 508–512.

Kirby, Simon, Tom Griffiths, and Kenny Smith (2014). "Iterated learning and the evolution of language". *Current Opinion in Neurobiology*. SI: Communication and language 28, 108–114. DOI: 10.1016/j.conb.2014.07.014.

Kirby, Simon and James R. Hurford (2002). "The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model". *Simulating the Evolution of Language*. Ed. by Angelo Cangelosi and Domenico Parisi. London: Springer, 121–147. DOI: 10.1007/978-1-4471-0663-0_6.

Kisler, Thomas, Uwe Reichel, and Florian Schiel (2017). "Multilingual processing of speech via web services". *Computer Speech & Language* 45, 326–347. DOI: 10.1016/j.csl.2017.01.005.

Kita, Sotaro, Katerina Kantartzis, and Mutsumi Imai (2010). "Children learn sound symbolic words better: Evolutionary vestige of sound symbolic protolanguage". *The Evolution of Language*. Ed. by Andrew D. M. Smith, Marieke Schouwstra, Bart de Boer, and Kenny Smith. Utrecht, The Netherlands: WORLD SCIENTIFIC, 206–213. DOI: 10.1142/9789814295222_0027.

Klerk, Vivian de and Barbara Bosch (1997). "The sound patterns of English nicknames". *Language Sciences* 19.4, 289–301. DOI: 10.1016/s0388-0001(96)00070-8.

Klimova, S. V. (1986). "Glagoly 'Nejasnogo Proiskhohdenija V Sokraschennom Oxfordskom Slovare: Elementy Etimologicheskoi Fonosemantiki (Verbs of 'Uncertain Origin' in The Concise Oxford English Dictionary: Elements of Etymological Phonosemantics)". PhD Thesis. Leningrad: University of Leningrad.

Klink, Richard R (2003). "Creating Meaningful Brands: The Relationship Between Brand Name and Brand Mark". *Marketing Letters* 14.3, 143–157.

Klink, Richard R. (2000). "Creating Brand Names With Meaning: The Use of Sound Symbolism". *Marketing Letters* 11.1, 5–20.

Klink, Richard R. (2001). "Creating Meaningful New Brand Names: A Study of Semantics and Sound Symbolism". *Journal of Marketing Theory and Practice* 9.2, 27–34. DOI: 10.1080/10696679.2001.11501889.

Klink, Richard R. (2009). "Gender differences in new brand name response". *Marketing Letters* 20.3, 313–326. DOI: 10.1007/s11002-008-9066-x.

Knoeferle, Klemens, Jixing Li, Emanuela Maggioni, and Charles Spence (2017). "What drives sound symbolism? Different acoustic cues underlie sound-size and sound-shape mappings". *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-05965-y.

Koenig, Christina (2011). *Tatze und die Eispiraten. Eisbärengeschichten*. Würzburg: Arena Verlag GmbH.

Kohler, K. J. (1982). "F0 in the Production of Lenis and Fortis Plosives". *Phonetica* 39.4-5, 199–218. DOI: 10.1159/000261663.

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

Kruschke, John K. (2015). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. 2nd. London: Elsevier. DOI: 10.1016/B978-0-12-405888-0.09999-2.

Kulot, Daniela (2016). *Ein kleines Krokodil mit ziemlich viel Gefühl*. Stuttgart: Thienemann.

Kumagai, Gakuji and Shigeto Kawahara (2017). "How abstract is sound symbolism: Labiality and diaper names in Japanese". *Proceedings of the 31st Annual Meeting of the Phonetic Society of Japan*. Tokyo, Japan, 49–54.

Kwon, Nahyun (2015). "The natural motivation of sound symbolism". PhD Thesis. Australia: University of Queensland Library. DOI: 10.14264/uql.2015.1013.

Kwon, Nahyun (2016). "Empirically Observed Iconicity Levels of English Phonaesthemes". *Public Journal of Semiotics* 7.2, 73–93. DOI: 10.37693/pjos.2016.7.16470.

Laidre, Mark E. (2009). "How Often Do Animals Lie about Their Intentions? An Experimental Test." *The American Naturalist* 173.3, 337–346. DOI: 10.1086/596530.

Laing, Catherine (2019). "A role for onomatopoeia in early language: evidence from phonological development". *Language and Cognition* 11.2, 173–187. DOI: 10.1017/langcog.2018.23.

Laing, Catherine E. (2014). "A phonological analysis of onomatopoeia in early word production". *First Language* 34.5, 387–405. DOI: 10.1177/0142723714550110.

Lameira, Adriano R., Madeleine E. Hardus, Adrian M. Bartlett, Robert W. Shumaker, Serge A. Wich, and Steph B. J. Menken (2015). "Speech-Like Rhythm in a Voiced and Voiceless Orangutan Call". *PLOS ONE* 10.1, e116136. DOI: 10.1371/journal.pone.0116136.

Lameira, Adriano R., Madeleine E. Hardus, Alexander Mielke, Serge A. Wich, and Robert W. Shumaker (2016). "Vocal fold control beyond the species-specific repertoire in an orang-utan". *Scientific Reports* 6.1, 30315. DOI: 10.1038/srep30315.

Leavens, David A., Jamie L. Russell, and William D. Hopkins (2010). "Multimodal communication by captive chimpanzees (Pan troglodytes)". *Animal Cognition* 13.1, 33–40. DOI: 10.1007/s10071-009-0242-z.

Lee, Michael S. Y., James B. Jago, Diego C. García-Bellido, Gregory D. Edgecombe, James G. Gehling, and John R. Paterson (2011). "Modern optics in exceptionally preserved eyes of Early Cambrian arthropods from Australia". *Nature* 474.7353, 631–634. DOI: 10.1038/nature10097.

Lega, Carlotta, Zaira Cattaneo, Lotfi B. Merabet, Tomaso Vecchi, and Silvia Cucchi (2014). "The effect of musical expertise on the representation of space". *Frontiers in Human Neuroscience* 8. DOI: 10.3389/fnhum.2014.00250.

Leiner, D. J. (2019). *SoSci Survey (Version 3.1.06) [Computer software].*

Leiss, Elisabeth (1997). "Synkretismus und Natürlichkeit". 31.1-2, 133–160. DOI: 10.1515/flin.1997.31.1-2.133.

Lengagne, Thierry (2008). "Traffic noise affects communication behaviour in a breeding anuran, Hyla arborea". *Biological Conservation* 141.8, 2023–2031. DOI: 10.1016/j.biocon.2008.05.017.

Leshinskaya, Anna and Alfonso Caramazza (2016). "For a cognitive neuroscience of concepts: Moving beyond the grounding issue". *Psychonomic Bulletin & Review* 23.4, 991–1001. DOI: 10.3758/s13423-015-0870-z.

Levin, Beth and Malka Rappaport Hovav (1991). "Wiping the slate clean: A lexical semantic exploration". *Cognition* 41.1, 123–151. DOI: 10.1016/0010-0277(91)90034-2.

Levin, Beth and Malka Rappaport Hovav (1992). "The lexical semantics of verbs of motion: the perspective from unaccusativity". *Thematic Structure: Its Role in Grammar*. Ed. by Iggy M. Roca. Vol. 16. Linguistic Models. Berlin, New York: Foris, 247–269.

Li, Wentian (2002). "Zipf's Law Everywhere". *Glottometrics* 5, 14–21.

Li, Xueqin, Garman Harbottle, Juzhong Zhang, and Changsui Wang (2003). "The earliest writing? Sign use in the seventh millennium BC at Jiahu, Henan Province, China". *Antiquity* 77.295, 31–44. DOI: 10.1017/S0003598X00061329.

Lieberman, Philip (1970). "Towards a Unified Phonetic Theory". *Linguistic Inquiry* 1.3, 307–322.

Likas, Aristidis, Nikos Vlassis, and Jakob J. Verbeek (2003). "The global k-means clustering algorithm". *Pattern Recognition*. Biometrics 36.2, 451–461. DOI: 10.1016/S0031-3203(02)00060-2.

Lindblom, B. (1990). "Explaining Phonetic Variation: A Sketch of the H&H Theory". *Speech Production and Speech Modelling*. Ed. by William J. Hardcastle and Alain Marchal. NATO ASI Series. Dordrecht: Springer Netherlands, 403–439. DOI: 10.1007/978-94-009-2037-8_16.

Little, Hannah, Marcus Perlman, and Kerem Eryilmaz (2017). "Repeated interactions can lead to more iconic signals". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 39. London, UK: Cognitive Science Society, 760–765.

Liu, Yadong, Arian Shamei, Una Y. Chow, Rachel Soo, Gina Pineda Mora, Gillian de Boer, and Bryan Gick (2020). "F0-related head movement in blind versus sighted speakers". *The Journal of the Acoustical Society of America* 148.2, EL190–EL194. DOI: 10.1121/10.0001353.

Lockwood, Gwilym, Mark Dingemanse, and Peter Hagoort (2016a). "Sound-symbolism boosts novel word learning". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42.8, 1274–1281. DOI: 10.1037/xlm0000235.

Lockwood, Gwilym, Peter Hagoort, and Mark Dingemanse (2016b). "How Iconicity Helps People Learn New Words: Neural Correlates and Individual Differences in Sound-Symbolic Bootstrapping". *Collabra* 2.1, 1–15. DOI: 10.1525/collabra.42.

Lowrey, Tina M. and L. J. Shrum (2007). "Phonetic Symbolism and Brand Name Preference". *Journal of Consumer Research* 34.3, 406–414. DOI: 10.1086/518530.

Lupyan, Gary and Bodo Winter (2018). "Language is more abstract than you think, or, why aren't languages more iconic?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 373.1752, 20170137. DOI: 10.1098/rstb.2017.0137.

Lynott, Dermot and Louise Connell (2009). "Modality exclusivity norms for 423 object properties". *Behavior Research Methods* 41.2, 558–564. DOI: 10.3758/BRM.41.2.558.

MacNeilage, Peter F. and Barbara L. Davis (2000). "On the Origin of Internal Structure of Word Forms". *Science* 288.5465, 527–531. DOI: 10.1126/science.288.5465.527.

Macuch Silva, Vinicius, Judith Holler, Asli Ozyurek, and Seán G. Roberts (2020). "Multimodality and the origin of a novel communication system in face-to-face interaction". *Royal Society Open Science* 7.1, 182056. DOI: 10.1098/rsos.182056.

Maduka-Durunze, Omen N. (2001). "Phonosemantic hierarchies". *Ideophones*. Ed. by Erhard Friedrich Karl Voeltz and Christa Kilian-Hatz. Typological Studies in Language 44. Amsterdam: John Benjamins Publishing Company, 193–203.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik (2021). *cluster: Cluster Analysis Basics and Extensions*.

Magnus, Margaret (2001). "What's in a Word? Studies in Phonosemantics". PhD Thesis. Trondheim: NTNU.

Maltzman, Irving, Lloyd Morrisett Jr., and Lloyd O. Brooks (1956). "An investigation of phonetic symbolism". *The Journal of Abnormal and Social Psychology* 53.2, 249–251. DOI: 10.1037/h0048406.

Marks, Lawrence E. (1975). "On colored-hearing synesthesia: Cross-modal translations of sensory dimensions". *Psychological Bulletin* 82.3, 303–331. DOI: 10.1037/0033-2909.82.3.303.

Martino, Gail and Lawrence E. Marks (2001). "Synesthesia: Strong and Weak". *Current Directions in Psychological Science* 10.2, 61–65. DOI: 10.1111/1467-8721.00116.

Marttila, Annu (2010). "A cross-linguistic study of lexical iconicity and its manifestation in bird names". PhD Thesis.

Massaro, Dominic W. and Marcus Perlman (2017). "Quantifying Iconicity's Contribution during Language Acquisition: Implications for Vocabulary Learning". *Frontiers in Communication* 2. DOI: 10.3389/fcomm.2017.00004.

Masuda, Junichi, Shigeki Morimoto, Tetsuya Watanabe, and Tamada Sousuke (1996). *Pokémon Red and Blue*.

Matrosova, Vera A., Ilya A. Volodin, Elena V. Volodina, and Andrey F. Babitsky (2007). "Pups crying bass: vocal adaptation for avoidance of age-dependent predation risk in ground squirrels?" *Behavioral Ecology and Sociobiology* 62.2, 181–191. DOI: 10.1007/s00265-007-0452-9.

Maurer, Daphne, Thanujeni Pathman, and Catherine J. Mondloch (2006). "The shape of boubas: sound-shape correspondences in toddlers and adults". *Developmental Science* 9.3, 316–322. DOI: 10.1111/j.1467-7687.2006.00495.x.

McCormick, Kelly, Jee Young Kim, Sara List, and Lynne C. Nygaard (2015). "Sound to Meaning Mappings in the Bouba-Kiki Effect". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 37. Pasadena, California: Cognitive Science Society.

McElreath, Richard (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 1st ed. Chapman and Hall/CRC. DOI: 10.1201/9781315372495.

McLean, Bonnie (2020). "Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic". *Linguistic Typology* -1.ahead-of-print. DOI: 10.1515/lingty-2020-2063.

McNeill, David (2012). *How Language Began: Gesture and Speech in Human Evolution*. Cambridge University Press. DOI: 10.1017/cbo9781139108669.

Meguerditchian, Adrien, Hélène Cochet, and Jacques Vauclair (2011). "From gesture to language: Ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization". *Primate communication and human language*. Ed. by Anne Vilain, Jean-Luc Schwartz, Christian Abry, and Jacques Vauclair. Amsterdam, The Netherlands: John Benjamins Publishing, 91–119. DOI: 10.1075/ais.1.07meg.

Meir, Irit (2010). "Iconicity and metaphor: Constraints on metaphorical extension of iconic forms". *Language* 86.4, 865–896. DOI: 10.1353/lan.2010.0044.

Meir, Irit, Carol A. Padden, Mark Aronoff, and Wendy Sandler (2007). "Body as subject". *Journal of Linguistics* 43.03. DOI: 10.1017/S0022226707004768.

Melara, Robert D. and Thomas P. O'Brien (1987). "Interaction between synesthetically corresponding dimensions". *Journal of Experimental Psychology: General* 116.4, 323–336. DOI: 10.1037/0096-3445.116.4.323.

Merin, Arthur and Christine Bartels (1997). "Decision-Theoretic Semantics for Intonation". *Arbeitspapiere des SFB 340* 88, 1–17.

Microsoft Corporation (2018). *Microsoft OneNote*.

Milligan, Glenn W. and Martha C. Cooper (1985). "An examination of procedures for determining the number of clusters in a data set". *Psychometrika* 50.2, 159–179. DOI: 10.1007/BF02294245.

Miron, M. S. (1961). "A cross-linguistic investigation of phonetic symbolism". *The Journal of Abnormal and Social Psychology* 62.3, 623–630. DOI: 10.1037/h0045212.

Mol, Lisette, Emiel Krahmer, and Marc Swerts (2009). "Alignment in Iconic Gestures: Does it make sense?" *International Conference on Auditory-Visual Speech Processing (AVSP)*. Ed. by Barry-John Theobald and Richard Harvey. Norwich, UK: University of East Anglia, 3–8.

Mompean, Jose A., Amandine Fregier, and Javier Valenzuela (2020). "Iconicity and systematicity in phonaesthemes: A cross-linguistic study". *Cognitive Linguistics* -1.ahead-of-print. DOI: 10.1515/cog-2018-0079.

Monaghan, Padraic, Richard C. Shillcock, Morten H. Christiansen, and Simon Kirby (2014). "How arbitrary is language?" *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1651, 20130299. DOI: 10.1098/rstb.2013.0299.

Moorthy, Sridhar, Ruth Pogacar, Samin Khan, and Yang Xu (2018). "Is Nike female? Exploring the role of sound symbolism in predicting brand name gender". *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processi*, 1128–1132.

Mooshammer, Christine and Christian Geng (2008). "Acoustic and articulatory manifestations of vowel reduction in German". *Journal of the International Phonetic Association* 38.2, 117–136. DOI: 10.1017/S0025100308003435.

Morales, Michael, Peter Mundy, and Jennifer Rojas (1998). "Following the direction of gaze and language development in 6-month-olds". *Infant Behavior and Development* 21.2, 373–377. DOI: 10.1016/S0163-6383(98)90014-5.

Morency, Louis-Philippe, Candace Sidner, Christopher Lee, and Trevor Darrell (2005). "Contextual recognition of head gestures". *Proceedings of the 7th international conference on Multimodal interfaces - ICMI '05*. Torento, Italy: ACM Press, 18. DOI: 10.1145/1088463.1088470.

Morey, Richard D., Jeffrey N. Rouder, Tahira Jamil, Simon Urbanek, Karl Forner, and Alexander Ly (2018). "Package 'BayesFactor'". *CRAN*.

Morton, Eugene S. (1977). "On the Occurrence and Significance of Motivation-Structural Rules in Some Bird and Mammal Sounds". *The American Naturalist* 111.981, 855–869. DOI: 10.1086/283219.

Morton, Eugene S. (1994). "Sound symbolism and its role in non-human vertebrate communication". *Sound Symbolism*. Ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala. Cambridge: Cambridge University Press.

Moseley, Dana Lynn, Graham Earnest Derryberry, Jennifer Nicole Phillips, Julie Elizabeth Danner, Raymond Michael Danner, David Andrew Luther, and Elizabeth Perrault Derryberry (2018). "Acoustic adaptation to city noise through vocal learning by a songbird". *Proceedings of the Royal Society B: Biological Sciences* 285.1888, 20181356. DOI: 10.1098/rspb.2018.1356.

Motamedi, Yasamin, Margherita Murgiano, Pamela Perniss, Elizabeth Wonnacott, Chloe Marshall, Susan Goldin-Meadow, and Gabriella Vigliocco (2020). "Linking language to sensory experience: onomatopoeia in early language development". *Developmental Science*, e13066. DOI: https://doi.org/10.1111/desc.13066.

Mudd, S. A. (1963). "Spatial stereotypes of four dimensions of pure tone". *Journal of Experimental Psychology* 66.4, 347–352. DOI: 10.1037/h0040045.

Munhall, K. G., Jeffery A. Jones, Daniel E. Callan, Takaaki Kuratate, and Eric Vatikiotis-Bateson (2004). "Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception". *Psychological Science* 15.2, 133–137. DOI: 10.1111/j.0963-7214.2004.01502010.x.

Munn, Charles A. (1986). "Birds that 'cry wolf'". *Nature* 319.6049, 143–145. DOI: 10.1038/319143a0.

Negus, V. E. (1949). *The comparative anatomy and physiology of the larynx*. New York: Hafner.

Newman, Stanley S. (1933). "Further Experiments in Phonetic Symbolism". *The American Journal of Psychology* 45.1, 53–75. DOI: 10.2307/1414186.

Nichols, Johanna (1971). "Diminutive Consonant Symbolism in Western North America". *Language* 47.4, 826–848. DOI: 10.2307/412159.

Niemitz, Carsten (2010). "The evolution of the upright posture and gait—a review and a new synthesis". *Naturwissenschaften* 97.3, 241–263. DOI: 10.1007/s00114-009-0637-3.

Nölle, Jonas, Stefan Hartmann, and Peeter Tinits (2020). "Language evolution research in the year 2020: A survey of new directions". *Language Dynamics and Change* 10.1, 3–26. DOI: 10.1163/22105832-bja10005.

Nuckolls, Janis B. (1996). *Sounds Like Life: Sound-Symbolic Grammar, Performance, and Cognition in Pastaza Quechua*. Ed. by William Bright. Oxford Studies in Anthropological Linguistics. Oxford University Press.

Nuckolls, Janis B. (1999). "The case for sound symbolism". *Annual Review of Anthropology* 28, 225–252.

Nuckolls, Janis B. (2004). "To Be or Not To Be Ideophonically Impoverished". *SALSA XI: Proceedings of the Eleventh Annual Symposium about Language and Society*. Ed. by Wai Fong Chiang, Elaine Chun, Laura Mahalingappa, and Siri Mehus. Austin: University of Texas, 131–142.

Nuckolls, Janis B. (2019). "The sensori-semantic clustering of ideophonic meaning in Pastaza Quichua". *Ideophones, Mimetics and Expressives*. Ed. by Kimi Akita and Prashant Pardeshi. Vol. 16. Iconicity in Language and Literature. John Benjamins Publishing Company, 167–198.

Nuckolls, Janis B., Elizabeth Nielsen, Joseph A. Stanley, and Roseanna Hopper (2016). "The Systematic Stretching and Contracting of Ideophonic Phonology in Pastaza Quichua". *International Journal of American Linguistics* 82.1, 95–116. DOI: 10.1086/684425.

Nygaard, Lynne C. (2008). "Sound Symbolism in Word Learning". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 30. Washington D.C.: Cognitive Science Society, 1912–1917.

Nygaard, Lynne C., Allison E. Cook, and Laura L. Namy (2009a). "Sound to meaning correspondences facilitate word learning". *Cognition* 112.1, 181–186. DOI: 10.1016/j.cognition.2009.04.001.

Nygaard, Lynne C., Debora S. Herold, and Laura L. Namy (2009b). "The semantics of prosody: acoustic and perceptual evidence of prosodic correlates to word meaning". *Cognitive Science* 33.1, 127–146. DOI: 10.1111/j.1551-6709.2008.01007.x.

O'Connor, Jillian J. M., Paul J. Fraccaro, Katarzyna Pisanski, Cara C. Tigue, Timothy J. O'Donnell, and David R. Feinberg (2014). "Social dialect and men's voice pitch influence women's mate preferences". *Evolution and Human Behavior* 35.5, 368–375. DOI: 10.1016/j.evolhumbehav.2014.05.001.

O'Connor, Jillian J.M. and Pat Barclay (2017). "The influence of voice pitch on perceptions of trustworthiness across social contexts". *Evolution and Human Behavior* 38.4, 506–512. DOI: 10.1016/j.evolhumbehav.2017.03.001.

Ogden, C. K. and I. A. Richards (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. New York: Harcourt, Brace & World, Inc.

Ohala, John J. (1980). "The acoustic origin of the smile". *The Journal of the Acoustical Society of America* 68.S1, S33–S33. DOI: 10.1121/1.2004679.

Ohala, John J. (1982). "The voice of dominance". *The Journal of the Acoustical Society of America* 72.S1, S66–S66. DOI: 10.1121/1.2020007.

Ohala, John J. (1984). "An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice". *Phonetica* 41.1, 1–16. DOI: 10.1159/000261706.

Ohala, John J. (1994). "The frequency code underlies the sound-symbolic use of voice pitch". *Sound symbolism*. Ed. by Leanne Hinton, Johanna Nichols, and John J. Ohala. Cambridge University Press, 325–347. DOI: 10.1017/cbo9780511751806.022.

Ohala, John J. (1996). "Speech perception is hearing sounds, not tongues". *The Journal of the Acoustical Society of America* 99.3, 1718–1725. DOI: 10.1121/1.414696.

Ohala, John J. (1997). "Sound Symbolism". *Proceedings of 4th Seoul International Conference on Linguistics [SICOL]*, 98–103.

Öhman, S. E. G. (1966). "Coarticulation in VCV Utterances: Spectrographic Measurements". *The Journal of the Acoustical Society of America* 39.1, 151–168. DOI: 10.1121/1.1909864.

Osorio, D. and M. Vorobyev (2008). "A review of the evolution of animal colour vision and visual communication signals". *Vision Research*. Vision Research Reviews 48.20, 2042–2051. DOI: 10.1016/j.visres.2008.06.018.

Parise, Cesare V., Katharina Knorre, and Marc O. Ernst (2014). "Natural auditory scene statistics shapes human spatial hearing". *Proceedings of the National Academy of Sciences* 111.16, 6104–6108. DOI: 10.1073/pnas.1322705111.

Parise, Cesare V. and Charles Spence (2009). "'When Birds of a Feather Flock Together': Synesthetic Correspondences Modulate Audiovisual Integration in Non-Synesthetes". *PLOS ONE* 4.5, e5664. DOI: 10.1371/journal.pone.0005664.

Parris, Kirsten M., Meah Velik-Lord, and Joanne M. A. North (2009). "Frogs Call at a Higher Pitch in Traffic Noise". *Ecology and Society* 14.1.

Pathak, Abhishek and Gemma A. Calvert (2020). "Sounds sweet, sounds bitter: How the presence of certain sounds in a brand name can alter expectations about the product's taste". *Food Quality and Preference*.

Pathak, Abhishek, Gemma Anne Calvert, and Lewis K. S. Lim (2020). "Harsh voices, sound branding: How voiced consonants in a brand's name can alter its perceived attributes". *Psychology & Marketing* n/a.n/a, 1–11. DOI: 10.1002/mar.21346.

Peirce, Charles S. (1955). *Philosophical Writings of Peirce*. Ed. by Justus Buchler. New York: Dover Publications.

Pelz, Jeff, Mary Hayhoe, and Russ Loeber (2001). "The coordination of eye, head, and hand movements in a natural task". *Experimental Brain Research* 139.3, 266–277. DOI: 10.1007/s002210100745.

Perkins, William H. and Raymond D. Kent (1986). *Functional Anatomy Of Speech, Language, And Hearing. A Primer*. Boston: Allyn and Bacon.

Perlman, Marcus (2011). "Empirical Perspective on Vocal Iconicity as a Starting Point of Language". PhD Thesis. Santa Cruz: University of California, Santa Cruz.

Perlman, Marcus (2017). "Debunking two myths against vocal origins of language". *Interaction Studies* 18.3, 376–401. DOI: 10.1075/is.18.3.05per.

Perlman, Marcus and Nathaniel Clark (2015). "Learned vocal and breathing behavior in an enculturated gorilla". *Animal Cognition* 18.5, 1165–1179. DOI: 10.1007/s10071-015-0889-6.

Perlman, Marcus, Nathaniel Clark, and Marlene Johansson Falck (2015a). "Iconic prosody in story reading". *Cognitive Science* 39.6, 1348–1368. DOI: 10.1111/cogs.12190.

Perlman, Marcus, Rick Dale, and Gary Lupyan (2015b). "Iconicity can ground the creation of vocal symbols". *Royal Society Open Science* 2.8, 150152. DOI: 10.1098/rsos.150152.

Perlman, Marcus, Riccardo Fusaroli, Deborah Fein, and Letita Naigles (2017). "The use of iconic words in early child-parent interactions". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 39. London, UK: Cognitive Science Society, 913–918.

Perlman, Marcus, Hannah Little, Bill Thompson, and Robin L. Thompson (2018). "Iconicity in Signed and Spoken Vocabulary: A Comparison Between American Sign Language, British Sign Language, English, and Spanish". *Frontiers in Psychology* 9. DOI: 10.3389/fpsyg.2018.01433.

Perlman, Marcus and Gary Lupyan (2018). "People can create iconic vocalizations to communicate various meanings to naïve listeners". *Scientific Reports* 8.1. DOI: 10.1038/s41598-018-20961-6.

Perlman, Marcus, Joanne E. Tanner, and Barbara J. King (2012). "A mother gorilla's variable use of touch to guide her infant". *Gesture Studies*. John Benjamins Publishing Company, 55–72. DOI: 10.1075/gs.6.04per.

Perniss, Pamela, Asli Özyürek, and Gary Morgan (2015). "The Influence of the Visual Modality on Language Structure and Conventionalization: Insights From Sign Language and Gesture". *Topics in Cognitive Science* 7.1, 2–11. DOI: 10.1111/tops.12127.

Perniss, Pamela, Robin L. Thompson, and Gabriella Vigliocco (2010). "Iconicity as a general property of language: evidence from spoken and signed languages". *Frontiers in Psychology* 1, 227. DOI: 10.3389/fpsyg.2010.00227.

Perniss, Pamela and Gabriella Vigliocco (2014). "The bridge of iconicity: from a world of experience to the experience of language". *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1651, 1–14. DOI: `10.1098/rstb.2013.0300`.

Perry, Lynn K., Stephanie A. Custode, Regina M. Fasano, Brittney M. Gonzalez, and Jordyn D. Savy (2021). "What Is the Buzz About Iconicity? How Iconicity in Caregiver Speech Supports Children's Word Learning". *Cognitive Science* 45.4, e12976. DOI: `https://doi.org/10.1111/cogs.12976`.

Perry, Lynn K., Marcus Perlman, and Gary Lupyan (2015). "Iconicity in English and Spanish and its relation to lexical category and age of acquisition". *PLOS ONE* 10.9. Ed. by Johan J. Bolhuis, e0137147. DOI: `10.1371/journal.pone.0137147`.

Perry, Lynn K., Marcus Perlman, Bodo Winter, Dominic W. Massaro, and Gary Lupyan (2018). "Iconicity in the speech of children and adults". *Developmental Science* 21.3, e12572. DOI: `10.1111/desc.12572`.

Peyo (1958). "Johan et Pirlouit: La flûte à six trous". *Spirou*.

Pfeifer, Wolfgang (2021a). "brummen". *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/etymwb/brummen>, abgerufen am 10.06.2021*.

Pfeifer, Wolfgang (2021b). *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/d/wb-etymwb>, abgerufen am 10.06.2021*.

Pfeifer, Wolfgang (2021c). "knurren". *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/etymwb/knurren>, abgerufen am 11.06.2021*.

Pfeifer, Wolfgang (2021d). "krachen". *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/etymwb/krachen>, abgerufen am 11.06.2021*.

Pfeifer, Wolfgang (2021e). "kratzen". *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/etymwb/kratzen>, abgerufen am 11.06.2021*.

Pfeifer, Wolfgang (2021f). "puffen". *Etymologisches Wörterbuch des Deutschen (1993), digitalisierte und von Wolfgang Pfeifer überarbeitete Version im Digitalen Wörterbuch der deutschen Sprache, <https://www.dwds.de/wb/etymwb/puffen>, abgerufen am 11.06.2021*.

Piantadosi, Steven T. (2014). "Zipf's word frequency law in natural language: A critical review and future directions". *Psychonomic bulletin & review* 21.5, 1112–1130. DOI: `10.3758/s13423-014-0585-6`.

Pietrandrea, Paola (2002). "Iconicity and Arbitrariness in Italian Sign Language". *Sign Language Studies* 2.3, 296–321.

Pinker, Steven and Paul Bloom (1990). "Natural language and natural selection". *Behavioral and Brain Sciences* 13.04, 707–727. DOI: 10.1017/s0140525x00081061.

Pisanski, Katarzyna (2014). "Human vocal communication of body size". PhD Thesis. Hamilton, Ontario: McMaster University.

Pisanski, Katarzyna, David Feinberg, Anna Oleszkiewicz, and Agnieszka Sorokowska (2017a). "Voice cues are used in a similar way by blind and sighted adults when assessing women's body size". *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-10470-3.

Pisanski, Katarzyna, Paul J. Fraccaro, Cara C. Tigue, Jillian J. M. O'Connor, and David R. Feinberg (2014a). "Return to Oz: Voice pitch facilitates assessments of men's body size". *Journal of Experimental Psychology: Human Perception and Performance* 40.4, 1316–1331. DOI: 10.1037/a0036956.

Pisanski, Katarzyna, Paul J. Fraccaro, Cara C. Tigue, Jillian J.M. O'Connor, Susanne Röder, Paul W. Andrews, Bernhard Fink, Lisa M. DeBruine, Benedict C. Jones, and David R. Feinberg (2014b). "Vocal indicators of body size in men and women: a meta-analysis". *Animal Behaviour* 95, 89–99. DOI: 10.1016/j.anbehav.2014.06.011.

Pisanski, Katarzyna, Sari G. E. Isenstein, Kelyn J. Montano, Jillian J. M. O'Connor, and David R. Feinberg (2017b). "Low is large: spatial location and pitch interact in voice-based body size estimation". *Attention, Perception, & Psychophysics* 79.4, 1239–1251. DOI: 10.3758/s13414-016-1273-6.

Pisanski, Katarzyna and Drew Rendall (2011). "The prioritization of voice fundamental frequency or formants in listeners' assessments of speaker size, masculinity, and attractiveness". *The Journal of the Acoustical Society of America* 129.4, 2201–2212. DOI: 10.1121/1.3552866.

Pischedda, Pier Simone (2020). "Translating English Sound Symbolism in Italian Comics: A Corpus-Based Linguistic Analysis across Six Decades (1932–1992)". *Arts* 9.4, 108. DOI: 10.3390/arts9040108.

Pitcher, Benjamin J., Alex Mesoudi, and Alan G. McElligott (2013). "Sex-Biased Sound Symbolism in English-Language First Names". *PLOS ONE* 8.6. Ed. by Elissa Z. Cameron, e64825. DOI: 10.1371/journal.pone.0064825.

Pogacar, Ruth, Agnes Pisanski Peterlin, Nike K. Pokorn, and Timothy Pogačar (2017). "Sound symbolism in translation: A case study of character names in Charles Dickens's Oliver Twist". *Translation and Interpreting Studies. The Journal of the American Translation and Interpreting Studies Association* 12.1, 137–161. DOI: 10.1075/tis.12.1.07pog.

Potts, Willard (1986). *Portraits of the Artist in Exile: Recollections of James Joyce by Europeans*. Harcourt Brace Jovanovich.

Pouw, Wim, Jan de Wit, Sara Bögels, Marlou Rasenberg, Branka Milivojevic, and Asli Ozyurek (2021). "Semantically Related Gestures Move Alike: Towards a Distributional Semantics of Gesture Kinematics". *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body, Motion and Behavior*. Ed. by Vincent G. Duffy. Vol. 12777. Cham: Springer International Publishing, 269–287. DOI: 10.1007/978-3-030-77817-0_20.

Pratt, Carroll C. (1930). "The spatial character of high and low tones". *Journal of Experimental Psychology* 13.3, 278–285. DOI: 10.1037/h0072651.

Quine, Willard Van Orman (1960). *Word and Object*. MIT Press.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rácová, Anna (2014). "Ideophones in Bengali". *Asian and African Studies* 23.1, 1–36.

Ramachandran, Vilayanur S and Edward M Hubbard (2001). "Synaesthesia — a window into perception, thought and language". *Journal of Consciousness Studies* 8.12, 3–34.

Rathcke, Tamara, Chia-Yuan Lin, Simone Falk, and Simone Dalla Bella (2021). "Tapping into linguistic rhythm". *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12.1, 11. DOI: 10.5334/labphon.248.

Real-time 3D Spectrogram now available in SignalScope for iOS (2015).

Reiter, Sabine (2011). "Ideophones in Aweti". PhD Thesis. Kiel, Germany: Christian-Albrechts-Universität zu Kiel.

Riede, Tobias and Charles Brown (2013). "Body Size, Vocal Fold Length, and Fundamental Frequency – Implications for Mammal Vocal Communication". *Nova acta Leopoldina: Abhandlungen der Kaiserlich Leopoldinisch-Carolinisch Deutschen Akademie der Naturforscher* 111.380, 295–314.

Riede, Tobias and W. Tecumseh Fitch (1999). "Vocal tract length and acoustics of vocalization in the domestic dog (Canis familiaris)". *Journal of Experimental Biology* 202.20, 2859–2867.

Riordan, Carol J. (1977). "Control of vocal-tract length in speech". *The Journal of the Acoustical Society of America* 62.4, 998–1002. DOI: 10.1121/1.381595.

Roettger, Timo B., Bodo Winter, and Harald Baayen (2019). "Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility". *Journal of Phonetics* 73, 1–7. DOI: 10.1016/j.wocn.2018.12.001.

Roffler, Suzanne K. and Robert A. Butler (1968). "Localization of Tonal Stimuli in the Vertical Plane". *The Journal of the Acoustical Society of America* 43.6, 1260–1266. DOI: 10.1121/1.1910977.

Rogers, Susan K. and Abraham S. Ross (1975). "A cross-cultural test of the maluma-takete phenomenon". *Perception* 4.1, 105–106. DOI: 10.1068/p040105.

Romesburg, H. Charles (1984). *Cluster Analysis for Researchers*. Belmont, California: Lifetime Learning Publications.

Rosenberger, Tara (1998). "Prosodic Font: the Space between the Spoken and the Written". MA thesis. Cambridge, MA: Massachusetts Institute of Technology.

Rusconi, Elena, Bonnie Kwan, Bruno L. Giordano, Carlo Umilta, and Brian Butterworth (2006). "Spatial representation of pitch height: the SMARC effect". *Cognition* 99.2, 113–129. DOI: 10.1016/j.cognition.2005.01.004.

Sandler, Wendy (1996). "Phonological features and feature classes: The case of movements in sign language". *Lingua* 98.1-3, 197–220. DOI: 10.1016/0024-3841(95)00038-0.

Sandler, Wendy (1999). "Prosody in Two Natural Language Modalities *". *Language and Speech* 42.2-3, 127–142. DOI: 10.1177/00238309990420020101.

Sandler, Wendy (2013). "Vive la différence: Sign language and spoken language in language evolution". *Language and Cognition* 5.2-3, 189–203. DOI: 10.1515/langcog-2013-0013.

Sandler, Wendy, Gal Belsitzman, and Irit Meir (2020). "Visual foreign accent in an emerging sign language". *Sign Language & Linguistics* 23.1-2, 233–257. DOI: 10.1075/sll.00050.san.

Santoso, Ailsa Agatha and Nina Setyaningsih (2020). "Iconicity Adaptation: An Analysis of Onomatopoeias in Finance Smurfs Comic and its Indonesian Translation Smurf Bendahara". *LITE: Jurnal Bahasa, Sastra, dan Budaya* 16.1, 120–134.

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

Saussure, Ferdinand de (2011). *Course in General Linguistics*. Ed. by Perry Meisel and Haun Saussy. Trans. by Wade Baskin. New York: Columbia University Press.

Schel, Anne Marijke, Simon W. Townsend, Zarin Machanda, Klaus Zuberbühler, and Katie E. Slocombe (2013). "Chimpanzee Alarm Call Production Meets Key Criteria for Intentionality". *PLOS ONE* 8.10, e76674. DOI: 10.1371/journal.pone.0076674.

Schlatter, Ralf (2019). *Margarethe geht*. Innsbruck: Limbus.

Schlobinski, Peter (2001). "*knuddel – zurueckknuddel – dich ganzdollknuddel*: Inflektive und Inflektivkonstruktionen im Deutschen". *Zeitschrift für germanistische Linguistik* 29.2, 192–218.

Schmidt, Christopher M. and Dagmar Neuendorff, eds. (2008). *Sprache, Kultur und Zielgruppen: Bedingungsgrößen für die Kommunikationsgestaltung in der Wirtschaft*. Europäische Kulturen in der Wirtschaftskommunikation. Deutscher Universitätsverlag. DOI: 10.1007/978-3-8350-5491-2.

Schmidtke, David S., Markus Conrad, and Arthur M. Jacobs (2014). "Phonological iconicity". *Frontiers in Psychology* 5. DOI: 10.3389/fpsyg.2014.00080.

Seashore, C. E. (1899). "Localization of sound in the median plane". *Univ. Iowa Stud. Psychol* 2, 46–54.

Sedley, D. N. (2003). *Plato's Cratylus*. Cambridge, UK: Cambridge University Press.

Senghas, Richard J., Ann Senghas, and Jennie E. Pyers (2005). "The Emergence of Nicaraguan Sign Language: Questions of Development, Acquisition, and Evolution". *Biology and Knowledge Revisited*. Routledge, 305–324. DOI: 10.4324/9781410611970-16.

Seyfarth, R. M., D. L. Cheney, and P. Marler (1980a). "Monkey responses to three different alarm calls: evidence of predator classification and semantic communication". *Science* 210.4471, 801–803. DOI: 10.1126/science.7433999.

Seyfarth, Robert M. and Dorothy L. Cheney (2003). "Signalers and Receivers in Animal Communication". *Annual Review of Psychology* 54.1, 145–173. DOI: 10.1146/annurev.psych.54.101601.145121.

Seyfarth, Robert M., Dorothy L. Cheney, and Peter Marler (1980b). "Vervet monkey alarm calls: Semantic communication in a free-ranging primate". *Animal Behaviour* 28.4, 1070–1094. DOI: 10.1016/S0003-3472(80)80097-2.

Sharpsteen, Ben, Larry Morey, Wilfred Jackson, David Hand, William Cottrell, and Perce Pearce (1937). *Snow White and the Seven Dwarfs*.

Shea, S. A. (1996). "Behavioural and arousal-related influences on breathing in humans". *Experimental Physiology* 81.1, 1–26. DOI: 10.1113/expphysiol.1996.sp003911.

Shih, Stephanie S, Jordan Ackerman, Noah Hermalin, Sharon Inkelas, Hayeun Jang, Darya Kavitskaya, Shigeto Kawahara, Miran Oh, Rebecca L Starr, and Alan Yu (2019). "Cross-linguistic and language-specific sound symbolism: Pokémonastics".

Shih, Stephanie S, Jordan Ackerman, Noah Hermalin, Sharon Inkelas, and Darya Kavitskaya (2018). "Pokémonikers: A study of sound symbolism and Pokémon names". *Proceedings of the Linguistic Society of America* 3, 42. DOI: 10.3765/plsa.v3i1.4335.

Shinohara, Kazuko and Shigeto Kawahara (2010). "A Cross-linguistic Study of Sound Symbolism: The Images of Size". *Annual Meeting of the Berkeley Linguistics Society* 36.1, 396–410. DOI: 10.3765/bls.v36i1.3926.

Shintel, Hadas, Howard C. Nusbaum, and Arika Okrent (2006). "Analog acoustic expression in speech communication". *Journal of Memory and Language* 55.2, 167–177. DOI: 10.1016/j.jml.2006.03.002.

Shrum, L. J. and Tina M. Lowrey (2007). "Sounds Convey Meaning: The Implications of Phonetic Symbolism for Brand Name Construction". *Psycholinguistic phenomena in marketing communications*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 39–58.

Sidhu, David, Jennifer Williamson, Velina Slavova, and Penny M. Pexman (2021a). *An Investigation of Iconic Language Development in Four Datasets*. Tech. rep. PsyArXiv. DOI: 10.31234/osf.io/qv9pg.

Sidhu, David M. and Penny M. Pexman (2015). "What's in a Name? Sound Symbolism and Gender in First Names". *PLOS ONE* 10.5, e0126809. DOI: 10.1371/journal.pone.0126809.

Sidhu, David M. and Penny M. Pexman (2019). "The Sound Symbolism of Names". *Current Directions in Psychological Science*, 0963721419850134. DOI: 10.1177/0963721419850134.

Sidhu, David M., Penny M. Pexman, and Jean Saint-Aubin (2016). "From the Bob/Kirk effect to the Benoit/Éric effect: testing the mechanism of name sound symbolism in two languages". *Acta Psychologica* 169, 88–99. DOI: 10.1016/j.actpsy.2016.05.011.

Sidhu, David M., Chris Westbury, Geoff Hollis, and Penny M. Pexman (2021b). "Sound symbolism shapes the English language: the maluma/takete effect in English nouns". *Psychonomic Bulletin & Review* 28.4, 1390–1398. DOI: 10.3758/s13423-021-01883-3.

Sidhu, David Michael (2019). "Explorations of Sound Symbolism and Iconicity". PhD Thesis. Calgary, Canada: University of Calgary.

Simner, Julia, Christine Cuskley, and Simon Kirby (2010). "What Sound Does That Taste? Cross-Modal Mappings across Gustation and Audition". *Perception* 39.4, 553–569. DOI: 10.1068/p6591.

Singer, Tania, Ben Seymour, John O'Doherty, Holger Kaube, Raymond J. Dolan, and Chris D. Frith (2004). "Empathy for Pain Involves the Affective but not Sensory Components of Pain". *Science* 303.5661, 1157–1162. DOI: 10.1126/science.1093535.

Śledziński, Daniel (2010). "Analiza struktury grup spółgłoskowych w nagłosie oraz w wygłosie wyrazów w języku polskim". *Kwartalnik Językoznawczy* 3-4.3-4, 61–83.

Smith, Ross (2006). "Fitting Sense to Sound: Linguistic Aesthetics and Phonosemantics in the Work of J.R.R. Tolkien". *Tolkien Studies* 3.1, 1–20. DOI: 10.1353/tks.2006.0032.

Sonesson, Göran (1994). "The ecological foundations of iconicity". *Semiotics Around the World: Synthesis in Diversity. Proceedings of the Fifth International Congress of the IASS*. Ed. by Irmengard Rauch and Gerlard F. Carr. Berlin & New York: Mouton de Gruyter, 739–742.

Speed, Laura J, Wessel O van Dam, Gabriella Vigliocco, and Rutvik H Desai (2018). "Movement Speed Affects Speed Language Comprehension". *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 40. Madison, Wisconsin: Cognitive Science Society, 7.

Spence, Charles (2011). "Crossmodal correspondences: a tutorial review". *Attention, Perception, & Psychophysics* 73.4, 971–995. DOI: 10.3758/s13414-010-0073-7.

Spence, Charles (2012). "Managing sensory expectations concerning products and brands: Capitalizing on the potential of sound and shape symbolism". *Journal of Consumer Psychology*. Brand Insights from Psychological and Neurophysiological Perspectives 22.1, 37–54. DOI: 10.1016/j.jcps.2011.09.004.

Spence, Charles (2018). "Multisensory Perception". *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. American Cancer Society, 1–56. DOI: 10.1002/9781119170174.epcn214.

Stel, Mariëlle, Eric van Dijk, Pamela K. Smith, Wilco W. van Dijk, and Farah M. Djalal (2012). "Lowering the Pitch of Your Voice Makes You Feel More Powerful and

Think More Abstractly". *Social Psychological and Personality Science* 3.4, 497–502. DOI: 10.1177/1948550611427610.

Stokoe, William C. (2001). *Language in Hand: Why Sign Came Before Speech*. Gallaudet University Press.

Stolarski, Łukasz (2015). "An Experiment in Size-Sound Symbolism: /i/ versus /a/ in Polish Diminutive Suffixes". *Anglica Wraticlaviensia*. Acta Universitatis Wratislaviensis 53.

Strange, Winifred and Ocke-Schwen Bohn (1998). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies". *The Journal of the Acoustical Society of America* 104.1, 488–504. DOI: 10.1121/1.423299.

Styles, Suzy J. and Lauren Gawne (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. DOI: 10.1177/2041669517724807.

Suire, Alexandre, Alba Bossoms Mesa, Michel Raymond, and Melissa Barkat-Defradas (2019). "Sex-biased sound symbolism in French first names". *Evolutionary Human Sciences* 1. DOI: 10.1017/ehs.2019.7.

Swadesh, Morris (1955). "Towards Greater Accuracy in Lexicostatistic Dating". *International Journal of American Linguistics* 21.2, 121–137. DOI: 10.1086/464321.

Szalontai, Adam, Petra Wagner, Katalin Mady, and Andreas Windmann (2016). "Teasing apart lexical stress and sentence accent in Hungarian and German". *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. München, Germany: Ludwig-Maximilians-Universität München, 215–218.

Taglialatela, Jared P., Jamie L. Russell, Jennifer A. Schaeffer, and William D. Hopkins (2011). "Chimpanzee Vocal Signaling Points to a Multimodal Origin of Human Language". *PLOS ONE* 6.4, e18852. DOI: 10.1371/journal.pone.0018852.

Talmy, Leonakd (1985). "Lexicalization patterns: semantic structure in lexical forms". *Language typology and syntactic description*. Ed. by T. Shopen. Cambridge: Cambridge University Press, 36–149.

Talmy, Leonard (1991). "Path to Realization: A Typology of Event Conflation". *Annual Meeting of the Berkeley Linguistics Society* 17.1, 480–519. DOI: 10.3765/bls.v17i0.1620.

Taub, Sarah F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge University Press.

Taylor, Insup Kim and Maurice M. Taylor (1962). "Phonetic symbolism in four unrelated languages". *Canadian Journal of Psychology/Revue canadienne de psychologie* 16.4, 344–356. DOI: 10.1037/h0083261.

Taylor, Kevin J. (2007). *Ka-BOOM!: A Dictionary of Comic Book Words, Symbols & Onomatopoeia*. Lulu Enterprises Incorporated.

Terken, Jacques (1991). "Fundamental frequency and perceived prominence of accented syllables". *The Journal of the Acoustical Society of America* 89.4, 1768–1776. DOI: 10.1121/1.401019.

Teuber, Oliver (1998). "fasel beschreib erwähn–Der Inflektiv als Wortform des Deutschen". *Germanistische Linguistik* 26.6, 141–142.

Tham, Diana S.Y., Alison Rees, J. Gavin Bremner, Alan Slater, and Scott Johnson (2019). "Auditory information for spatial location and pitch–height correspondence support young infants' perception of object persistence". *Journal of Experimental Child Psychology* 178, 341–351. DOI: `10.1016/j.jecp.2018.05.017`.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie (2001). "Estimating the number of clusters in a data set via the gap statistic". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2, 411–423. DOI: `https://doi.org/10.1111/1467-9868.00293`.

Timmers, Renee and Shen Li (2016). "Representation of pitch in horizontal space and its dependence on musical and instrumental experience". *Psychomusicology: Music, Mind, and Brain* 26.2, 139–148. DOI: `10.1037/pmu0000146`.

Tolkien, J.R.R. (1991). *The Lord of the Rings*. London, UK: HarperCollins.

Tomasello, Michael (2008). *Origins of human communication*. Ed. by Francois Recanati. Cambridge, Massachusetts; London, England: The MIT Press.

Tomaszewski, Piotr (2006). "From iconicity to arbitrariness: How do gestures become signs in peer-group pidgin". *Psychology of Language and Communication* 10.2, 27–59.

Trask, R. L. (2004). *A Dictionary of Phonetics and Phonology*. Routledge.

Traunmüller, Hartmut (1996). "Sound symbolism in deictic words". *Tongues and texts unlimited. Studies in honour of Tore Jansson on the occasion of his sixtieth anniversary*, 213–234.

Tremblain, Antoine and Johannes Ransijn (2015). *Package 'LMERConvenienceFunctions'. Model Selection and Post-hoc Analysis for (G)LMER Models*.

Trimble, Otis C. (1934). "Localization of Sound in the Anterior, Posterior and Vertical Dimensions of "Auditory" Space". *British Journal of Psychology* 24.3, 320–334.

Ultan, Russell (1969). *Some General Characteristics of Interrogative Systems. Working Papers on Language Universals, No.1.* Tech. rep. Stanford, CA: Stanford Univ., Calif. Committee on Linguistics.

Ultan, Russell (1978). "Size-sound symbolism". *Phonology*. Ed. by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik. Vol. 2. Universals of Human Language. Stanford: Stanford University Press, 525–567.

Updates to the OED (2020).

Ussishkin, Adam and Andrew Wedel (2009). "Lexical access, effective contrast and patterns in the lexicon". *Phonology in Perception*. Ed. by Paul Boersma and Silke Hamann. Berlin & New York: Walter de Gruyter.

Usunier, Jean-Claude and Janet Shaner (2002). "Using linguistics for creating better international brand names". *Journal of Marketing Communications* 8.4, 211–228. DOI: `10.1080/13527260210146000`.

Verhoef, Tessa, Simon Kirby, and Bart de Boer (2014). "Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals". *Journal of Phonetics* 43, 57–68. DOI: 10.1016/j.wocn.2014.02.005.

Voeltz, Erhard Friedrich Karl and Christa Kilian-Hatz, eds. (2001). *Ideophones*. Typological Studies in Language 44. Amsterdam: John Benjamins Publishing Company.

Voronin, Stanislav V. (2006). *Osnovy Phonosemantiki (The Fundamentals of Phonosemantics)*. Moscow, Russia: Lenand.

Wacewicz, Sławomir and Przemysław Żywczyński (2017). "The multimodal origins of linguistic communication". *Language & Communication*. The multimodal origins of linguistic communication 54, 1–8. DOI: 10.1016/j.langcom.2016.10.001.

Wagner, Petra (2016). "Beat it! – Gesture-based Prominence Annotation as a Window to Individual Prosody Processing Strategies". *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. München, Germany: Ludwig-Maximilians-Universität München, 211–214.

Wagner, Petra, Aleksandra Ćwiek, and Barbara Samlowski (2019). "Exploiting the speech-gesture link to capture fine-grained prosodic prominence impressions and listening strategies". *Journal of Phonetics* 76, 100911. DOI: 10.1016/j.wocn.2019.07.001.

Wagner, Petra, Zofia Malisz, and Stefan Kopp (2014). "Gesture and speech in interaction: An overview". *Speech Communication* 57, 209–232. DOI: 10.1016/j.specom.2013.09.008.

Wallman, Joel (1992). *Aping Language*. Cambridge University Press.

Weiss, Jonathan H. (1966). "A Study of the Ability of English Speakers to Guess the Meanings of Nonantonym Foreign Words". *The Journal of General Psychology* 74.1, 97–106. DOI: 10.1080/00221309.1966.9710313.

Westbury, Chris, Geoff Hollis, David M. Sidhu, and Penny M. Pexman (2018). "Weighing up the evidence for sound symbolism: distributional properties predict cue strength". *Journal of Memory and Language* 99, 122–150. DOI: 10.1016/j.jml.2017.09.006.

Westbury, J. and P. Keating (1986). "On the naturalness of stop consonant voicing". *undefined*.

Westermann, Diedrich (1927). "Laut, Ton und Sinn in Westafrikanischen Sudansprachen". *Festschrift Meinhof*, 315–228.

Westermann, Diedrich (1937). "Laut und Sinn in einigen westafrikanischen Sprachen". *Archiv Für Vergleichende Phonetik* 1, 154–72, 193–211.

Wetzels, Ruud and Eric-Jan Wagenmakers (2012). "A default Bayesian hypothesis test for correlations and partial correlations". *Psychonomic Bulletin & Review* 19.6, 1057–1064. DOI: 10.3758/s13423-012-0295-x.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Wickham, Hadley (2017). *tidyverse: Easily install and load the "tidyverse."* https://CRAN.R-project.org/package=tidyverse.

Wierzchowska, Bożena (1967). *Opis fonetyczny języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Willems, Roel M. and Peter Hagoort (2007). "Neural evidence for the interplay between language, gesture, and action: A review". *Brain and Language*. Gesture, Brain, and Language 101.3, 278–289. DOI: 10.1016/j.bandl.2007.03.004.

Wilson, Margaret (2002). "Six views of embodied cognition". *Psychonomic Bulletin & Review* 9.4, 625–636. DOI: 10.3758/BF03196322.

Winter, Bodo (2016). "The sensory structure of the english lexicon". PhD Thesis. Merced, California: UC Merced.

Winter, Bodo (2019). *Sensory linguistics: language, perception and metaphor*. Vol. 20. Converging Evidence in Language and Communication Research. Amsterdam: John Benjamins Publishing Company. DOI: 10.1075/celcr.20.

Winter, Bodo (2020). *Statistics for Linguists: An Introduction Using R*. New York and London: Routledge.

Winter, Bodo, Marcus Perlman, Lynn K. Perry, and Gary Lupyan (2017). "Which words are most iconic? Iconicity in English sensory words". *Interaction Studies* 18.3, 443–464. DOI: 10.1075/is.18.3.07win.

Witzigmann, Véronique and Caren Zacharias Bjarke (2010). *Die kleine Marmeladenfee. Um die Wette - fertig, los!* Köln: Baumhaus.

Wolk, Lesley, Nassima B. Abdelli-Beruh, and Dianne Slavin (2012). "Habitual Use of Vocal Fry in Young Adult Female Speakers". *Journal of Voice* 26.3, e111–e116. DOI: 10.1016/j.jvoice.2011.04.007.

Woodworth, Nancy L. (2009). "Sound symbolism in proximal and distal forms". *Linguistics* 29.2, 273–300. DOI: 10.1515/ling.1991.29.2.273.

Woolard, Kathryn A. (2005). "Codeswitching". *A Companion to Linguistic Anthropology*. John Wiley & Sons, Ltd, 73–94. DOI: 10.1002/9780470996522.ch4.

Wormer, Eberhard J. (2017). "Entenhausen an der Saale". *Orthopädie & Rheuma* 20.5, 60–61. DOI: 10.1007/s15002-017-1230-3.

Wundt, Wilhelm (1973). *The Language of Gestures*. Original work published in 1901. The Hague: Mouton.

Yorkston, Eric and Geeta Menon (2004). "A Sound Idea: Phonetic Effects of Brand Names on Consumer Judgments". *Journal of Consumer Research* 31.1, 43–51. DOI: 10.1086/383422.

Yoshida, Hanako (2012). "A Cross-Linguistic Study of Sound Symbolism in Children's Verb Learning". *Journal of Cognition and Development* 13.2, 232–265. DOI: 10.1080/15248372.2011.573515.

Yuasa, Ikuko Patricia (2010). "Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women?" *American Speech* 85.3, 315–337. DOI: 10.1215/00031283-2010-018.

Zbikowski, Lawrence M. (1998). "Metaphor and Music Theory: Reflections from Cognitive Science". *Journal of the Society for Music Theory* 4.1, 1–8.

Zimmer, Carl (2021). "Ancient Footprints Push Back Date of Human Arrival in the Americas". *The New York Times*.

Zlatev, Jordan (2014). "Human Uniqueness, Bodily Mimesis and the Evolution of Language". *HUMANA.MENTE Journal of Philosophical Studies* 7.27, 197–219.

Żywiczyński, Przemysław, Marta Sibierska, Sławomir Wacewicz, Joost van de Weijer, Francesco Ferretti, Ines Adornetti, Alessandra Chiera, and Valentina Deriu (2021). "Evolution of conventional communication. A cross-cultural study of pantomimic re-enactments of transitive events". *Language & Communication* 80, 191–203. DOI: 10.1016/j.langcom.2021.07.002.

Żywiczyński, Przemysław and Sławomir Wacewicz (2019). *The Evolution of Language: Towards Gestural Hypotheses*. Peter Lang International Academic Publishers. DOI: 10.3726/b15805.

Selbständigkeitserklärung zur Dissertation

Name: Ćwiek, Vorname: Aleksandra

Ich erkläre ausdrücklich, dass es sich bei der von mir eingereichten Dissertation mit dem Titel

"Iconicity in Language and Speech"

um eine von mir erstmalig, selbständig und ohne fremde Hilfe verfasste Arbeit handelt.

Ich erkläre ausdrücklich, dass ich *sämtliche* in der oben genannten Arbeit verwendeten fremden Quellen, auch aus dem Internet (einschließlich Tabellen, Grafiken u. Ä.) als solche kenntlich gemacht habe. Insbesondere bestätige ich, dass ich ausnahmslos sowohl bei wörtlich übernommenen Aussagen bzw. unverändert übernommenen Tabellen, Grafiken u. Ä. (Zitaten) als auch bei in eigenen Worten wiedergegebenen Aussagen bzw. von mir abgewandelten Tabellen, Grafiken u. Ä. anderer Autorinnen und Autoren (Paraphrasen) die Quelle angegeben habe.

Mir ist bewusst, dass Verstöße gegen die Grundsätze der Selbständigkeit als Täuschung betrachtet und nach § 16 der Promotionsordnung der Philosophischen Fakultät II (jetzt: Sprach- und literaturwissenschaftliche Fakultät) vom 27. April 2016 (Amtliches Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 26/2016) entsprechend geahndet werden.

Datum: ______________________________

Unterschrift: ______________________________