Unimodal and Cross-Modal Iconicity in Japanese Ideophones

A cognitive-semiotic approach

Nils Paulsson

Supervisor: Jordan Zlatev

Centre for Language and Literature, Lund University

MA in Language and Linguistics, Cognitive Semiotics

SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

August 2025

Abstract

Drawing on concepts from cognitive semiotics and phenomenology, this thesis investigates how Japanese ideophones are perceived by non-speakers, aiming to clarify key concepts such as iconicity, cross-modality/unimodality and the ultimately the nature of language as a semiotic system. Three research questions guided the study: (1) What is the relation between primary and secondary iconicity in the perception and interpretation of ideophones? (2) Which sensory modalities are perceived as more iconic? (3) To what extent does Japanese orthography influence perceptions of iconicity? An online experiment was conducted with non-Japanese participants using two tasks: (a) a two-alternative forced-choice Choosing Task, designed to elicit pre-reflective judgments of ideophone-meaning correspondences, and (b) an iconicity Rating Task, eliciting reflective evaluations of iconicity. The results show that participants "guessed" meanings above chance across modalities, confirming the roles of primary iconicity: being able to establish the meaning of unknown words (signs) beyond chance. Accuracy in the Choosing Task was highest for unimodal auditory ideophones, followed by cross-modal visual ideophones, while interoceptive items were less transparent. Interestingly, Rating Task diverged from this, suggesting that reflective judgments of vividness and expressivity do not align directly with pre-reflective ratings of iconicity. Orthographic manipulations (hiragana vs. katakana) produced no significant effects. These findings contribute to the debate on iconicity in language by demonstrating that pre-reflective and reflective levels of sign interpretation do not necessarily coincide, supporting a cognitive-semiotic account where primary and secondary iconicity operate in tandem, and pre-reflective judgements of often have higher validity than introspective ones. The study advances understanding of how iconicity functions in ideophones in Japanese, and more generally, thus clarifying the nature of language as a highly structured and conventional but non-arbitrary semiotic system.

Keywords: Japanese ideophones; cross-modal iconicity; sensory modality; orthography; sound symbolism; cognitive semiotics

Acknowledgements

This thesis would not have been possible without the support of many people, to whom I owe my deepest gratitude.

First and foremost, I would like to thank my supervisor, Professor Jordan Zlatev, for his patience, guidance, and encouragement throughout this project. His feedback and support have been invaluable, both in refining the ideas behind this thesis and in helping me see it through to completion.

I am also deeply indebted to my peer and colleague, Daiya Kuryū, whose inspiration, interest, and crucial literature recommendations helped frame this project and set the stage for its design. His practical vetting of my ideophone list and our many conversations about the topic were indispensable to the development of the study. I also want to congratulate him warmly on his PhD position, which is a well-deserved achievement. I would further like to thank Jahnesta Sjöström for important discussions that helped clarify the framing of this research, and my other fellow MA thesis colleagues, including Natalia Basalaeva, Emin Gaaya, Aspasia Voumvaki, Chen Cheng, and many others, for the dialogue, encouragement, practical assistance, and second-person perspectives that guided me along the way.

My thanks also extend to all those who assisted in recruiting participants: friends, cousins, extended family, and peers. There are too many to name individually, but this project would not have been possible without your efforts. I am equally grateful to all the participants who took the time to take part and give their honest responses.

I also acknowledge the MA Program in Language and Linguistics at Lund University for providing cinema vouchers to support recruitment.

Lastly, I wish to thank my family for their unfailing support. To my parents, for helping me get over the finish line, and to my father in particular for his practical help with pivot tables and charts in Excel. I could not have completed this without you.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Language is often thought of as a semiotic system in which the "bond between the signifier and the signified is arbitrary" (Saussure 1959: 67). However, at least some words seem to challenge this principle by sounding like their meanings, like *bang* or *splash*. Such resemblances have been found in most, and possibly all of the world's languages (Perniss, Thompson & Vigliocco 2010; Dingemanse 2012). For example, the sound of a barking dog can be represented as *voff voff* in Swedish, *wan wan* in Japanese, or *bow-wow* in English. These words are instances of *iconicity* in language (Jakobson 1965) because their forms are perceived to resemble their meaning: the sound of a barking dog, despite that they vary from language to language. Onomatopoeia forms a subset of a broader category of words, known as *ideophones:* words that can imitate a much wider range of sensory experiences than sound, which have been studied as part of the long-lasting debate about the role of iconicity in language (Benveniste 1981; Hamano 1998; Hinton, Nichols & Ohala 1995; Jakobson 1965; Kunihira 1971).

The view that the relation between "form and meaning", or more appropriately from a semiotic perspective, *expression* and *content* (e.g. Sonesson 2010) in language is fundamentally motivated has found increasing support in the recent literature (Dingemanse et al. 2015; Erben Johansson et al. 2020; Perniss, Thompson & Vigliocco 2010). The phenomenon itself, often referred to as *sound symbolism* (Hinton, Nichols & Ohala 1995), has been a topic of debate since antiquity, notably discussed in Plato's *Cratylus* dialogue (Sedley 2003). Experimental support in favor of similarity-based associations between (fictive) sound-forms and visual shapes dates back to at least the early twentieth century, with classical studies conducted and described by Sapir (1929) and Köhler (1929). These demonstrated that people can reliably associate novel words, like *mil* and *mal* or *takete* and *maluma*, with abstract properties such as visual shapes that are either round or soft. More recent research has provided further support for such cross-domain mappings, now commonly known as the "bouba–kiki effect" (Ahlner & Zlatev 2010; Cuskley, Simner & Kirby 2017; Ramachandran & Hubbard 2001). These findings suggest that certain phonetic properties evoke corresponding sensory perceptions, reinforcing the argument that the signs of

language are not entirely arbitrary. However, the influence of orthography on the outcome of these experiments has been the subject of debate, with suggestions that the shape of letters used to represent certain sounds may affect the results (Cuskley, Simner & Kirby 2017; De Carolis et al. 2018; Turoman & Styles 2017), while other studies show that the effect is robust across both cultures and writing systems (Ćwiek et al. 2021). The question how it is possible to identify resemblances across "shapes" in different sensory modalities, for example hearing and vision, has not been fully addressed either (Ahlner & Zlatev 2010; Imai et al. 2025).

Ideophone research has provided another valuable avenue for studying language iconicity and, more generally, the *motivated* rather than arbitrary nature of language as a semiotic system (Dingemanse 2019; Lockwood, Dingemanse & Hagoort 2016; Thompson, Akita & Do 2020). Ideophones are particularly well-suited for this purpose because they exhibit systematic correspondence between their phonetic structure and the sensory experiences they signify. This makes them especially useful for studying the transfer of meaning across sensory modalities, in a similar manner to the experiments that gave rise to the bouba–kiki effect. Studies have shown that listeners unfamiliar with a given language can often "guess" the meaning of ideophones at rates that are significantly above chance, likely owing to generally perceivable sound-meaning correspondences. However, some studies comparing English speaking and Japanese speaking participants show that, for Japanese ideophones, subjective ratings may be affected by conventionalization and frequency effects (Thompson, Akita & Do 2020).

This thesis aims to clarify the role of iconicity in language by examining the extent to which sound-meaning correspondences in Japanese ideophones can be perceived cross-linguistically, and which sensory modalities are involved in cross-modal iconic mappings. In addition, it aims to assess the effects of orthography by comparing the influence of written representation using two phonographic Japanese scripts. It employs methods and frameworks developed within *cognitive semiotics* (Konderak 2018; Sonesson 2012a; Zlatev 2015) to provide a broader understanding of the process of meaning-making by combining concepts and methods from linguistics, semiotics, cognitive science, and phenomenology, making relevant distinctions between different types of iconicity where necessary. This includes differences between *primary* and *secondary*, as well as between *unimodal* and *cross-modal* iconicity (notions that are explained in Chapter 2), and how

these relate to various distinctions that have previously been made in the literature on ideophones. Additionally. the role of *polysemiosis* and *(inter)subjectivity* are taken into consideration, as explained in Chapter 2. More specifically, the study aims to answer the following research questions:

RQ1 What is the relation between primary and secondary iconicity in the perception and interpretation of ideophones?

RQ2 Which sensory modalities are perceived as more iconic?

RQ3 To what extent does Japanese orthography influence perceptions of iconicity?

The structure of the thesis is as follows. Chapter 2 provides the theoretical foundation, defining central concepts and reviewing prior research on iconicity in language, with a particular focus on ideophones. Chapter 3 presents the methodological framework, detailing how the research questions were operationalized, the experimental design, materials, procedures, and hypotheses. Chapter 4 reports the empirical findings from both the accuracy-based choosing task and the subjective rating task, with analyses of the effects of sensory modality and orthography. It also situates the findings within the theoretical framework introduced in Chapter 2 and addresses limitations of the study. Finally, Chapter 5 summarizes the main contributions, outlines directions for future research, and offers concluding remarks.

Chapter 2 Theoretical Background

This chapter presents the conceptual and empirical foundations of the thesis. Section 2.1 introduces the field of cognitive semiotics, outlining its scope, methodological principles, and reliance on phenomenological concepts. Section 2.2 presents one hierarchical analysis of semiosis, focusing on iconicity as a central semiotic ground and distinguishing between primary and secondary forms of iconicity. It further examines how iconicity operates in signs generally and in language in particular, with special attention to ideophones. Section 2.3 reviews empirical studies that address the perception and interpretation of iconic signs, especially Japanese ideophones, highlighting findings relevant to the current research questions. This includes studies of cross-modal iconicity, typological hierarchies of ideophone domains, and the role of orthography in shaping iconic perception. Finally, Section 2.4 summarizes the theoretical background and presents the revised research questions that guide the empirical investigation in Chapter 3.

2.1 Basics of cognitive semiotics

Cognitive semiotics has emerged over the past two decades as a distinct discipline dedicated to the study of meaning-making across different forms of communication and cognition. It arose from the realization that neither traditional semiotics nor cognitive science alone could fully account for the multifaceted phenomenon of meaning. Early proponents identified a need to bridge the gap between the humanities and the sciences in understanding how mind and meaning are intertwined (Zlatev 2015: 1043). In response, cognitive semiotics developed as a transdisciplinary approach that integrates insights from semiotics, linguistics, psychology, and neuroscience, while also drawing on philosophy, and in particular on the school of *phenomenology* (e.g., Sokolowski 2000). The result is a new paradigm for studying meaning that seeks to combine empirical research with deep theoretical reflection. This section introduces cognitive semiotics as such a transdisciplinary science of meaning-making, defining its scope and distinguishing its pluralistic, non-reductive stance from earlier frameworks. We will then delve into two foundations of its methodology: (1)

phenomenological concepts and triangulation, and (2) the conceptual-empirical loop, which together ground the approach adopted in this thesis.

Cognitive semiotics is most generally "the study of mind and meaning, the way meaning exists and works in human minds (and ideally, in animal minds in general)" (Brandt 2011: 49). In other words, if semiotics studies meaning and cognitive science studies the mind, cognitive semiotics explicitly studies meaning in minds. As mentioned above, it is characterized as "a new interdisciplinary, or rather transdisciplinary, field" focusing on meaning, integrating methods and theories from both the cognitive sciences and the humanities (Zlatev 2015: 1043). Based on developments during the past decade, Zlatev (in press) further defines it as "the academic discipline that focuses on meaning-making (semiosis), combining concepts and methods from semiotics, cognitive science, linguistics, as well as phenomenology". Crucially, cognitive semiotics does not limit itself to language or symbols alone; it examines meaning-making in all its forms, in language, gesture, perception, art, and even non-human communication, using whatever disciplinary tools are necessary. As a result, the field is inherently pluralistic. It draws on an expansive range of disciplines: "starting with philosophy of mind, via semiotics and linguistics, cognitive science(s), neuroanthropology, developmental and evolutionary psychology, comparative studies, and ending with robotics" (Konderak 2018). This breadth reflects the core aim of cognitive semiotics: to synthesize perspectives on meaning that have traditionally been separate, and in doing so to offer a more comprehensive understanding of how significance is created and interpreted by human beings and (other) animals.

Integral to this vision is the integration of semiotic and cognitive approaches. Classical semiotics often analyzed static systems of signs (e.g. linguistic structures or cultural codes), whereas cognitive science traditionally reduced mind to computational or neural mechanisms. Cognitive semiotics moves beyond both limitations. It focuses on the dynamics of meaning: meaning is not a fixed product but an ongoing process. As Konderak (2018) explains, cognitive semiotics views meaning not as a static entity but as something fundamentally dynamic and context-dependent, continually reinterpreted by active subjects. Language, for example, is seen "not as [a] ready-to-analyze complete phenomenon, but rather as a process" shaped by individual, social and environmental factors. This contrasts with approaches in traditional semiotics that might

take a "snapshot" view of meaning or treat language as a closed system. By emphasizing meaning-making (semiosis) itself, cognitive semiotics accounts for how meanings emerge, change and circulate across different timescales, from milliseconds of perceptual experience to centuries of linguistic evolution (Konderak 2018). It thereby acknowledges that the subjects of meaning-making are active, not passive receivers, an insight aligned with enactive cognitive science and phenomenology.

At the same time, cognitive semiotics explicitly distances itself from the reductionism found in some cognitive science paradigms. Mainstream cognitive science since the mid-20th century often adopted a physicalist stance, treating mind and meaning as ultimately explainable in terms of neural or computational processes alone. This led to "strong reductionist tendencies, viewing mind and meaning as ultimately physical phenomena" (Zlatev 2015: 1044). Cognitive semiotics, in contrast, maintains a non-reductive stance: it refuses to reduce meaning purely to brain activity or to any single explanatory level. Instead, it espouses an ontological pluralism and methodological openness. Zlatev (2015) notes that cognitive semiotics is "considerably more pluralist in its ontological and methodological commitments" than standard cognitive science, with a "firmer foot in the humanities". This means that multiple kinds of entities (e.g. subjective experiences, social interactions, sign vehicles, neural processes) are recognized as real and relevant for meaning, and multiple methods (from phenomenology to behavioral experiments) are considered valid tools for inquiry. In sum, cognitive semiotics seeks a synthesis that both integrates and transcends its parent disciplines (semiotics, linguistics, psychology/neuroscience), rather than being subsumed by any one of them. It is this non-reductive, transdisciplinary commitment to studying meaning that defines cognitive semiotics and sets it apart from earlier approaches.

2.1.1 Phenomenological concepts in cognitive semiotics

One of the key influences that differentiates cognitive semiotics from other approaches is its incorporation of phenomenology, the philosophical school that privileges human experience and consciousness, addressing fundamental epistemological and ontological questions on this basis (Sokolowski 2000). Phenomenology provides crucial concepts, particularly *subjectivity* and *intentionality*, which inform the cognitive semiotic view of meaning-making. These concepts highlight the role of lived experience and the first-person perspective in any account of meaning,

reinforcing the field's non-reductive stance. In this subsection, I define these terms and explain their relevance for cognitive semiotics, and thus for the present thesis.

Subjectivity in phenomenology refers to the first-personal character of experience, the fact that experiences always occur for someone. The term denotes the qualitative, felt aspect of consciousness (the "what-it-is-like" of experience) and the presence of a perspective or self to whom things appear. As Sokolowski (2000) puts it, subjectivity is the realm of experiencing subjects, for whom experiences are immediately given. Every act of meaning is experienced by a subject as "my" experience in some implicit way. Zahavi (2019) similarly emphasizes that any conscious experience entails a minimal form of self-awareness or for-me-ness, a tacit sense that I am the one undergoing the experience. This does not imply a grand Cartesian ego, but rather a pre-reflective awareness of oneself in the act of experiencing. In cognitive semiotics, recognizing the primacy of subjectivity is essential because it asserts that meaning ultimately depends on qualitative experience. There is no meaning without someone to mean or find meaning. For example, a sign or word only means something insofar as it is perceived or understood by a subject. By foregrounding subjectivity, cognitive semiotics aligns with the phenomenological insight that "our knowledge of the world… arises from a first-person perspective" (Gallagher & Zahavi 2010: 89) and that even scientific knowledge is rooted in such experience. This ensures that cognitive semiotics does not abstract away the lived, experiential dimension of meaning: any theory of signs or cognition must be grounded in how things within the Lifeworld or *Lebenswelt* (Husserl's term for the world as lived). The central importance of subjectivity has been termed the "primacy of subjectivity" in cognitive semiotics (Zlatev in press), meaning that phenomena like signs, language, or thought are ultimately understood in relation to the experiencing agent.

Intentionality is the second fundamental phenomenological concept informing cognitive semiotics. In phenomenology, intentionality means that consciousness is always consciousness of something. Every experience has an intentional object; our thoughts, perceptions, and emotions are directed toward objects or states of affairs in the world. Classic phenomenologists like Husserl and Merleau-Ponty stressed that intentionality is what "opens the window" from the subject's mind to the world, linking subjective experience with objective content. Sokolowski (2000) describes intentionality as the cornerstone of phenomenological theory of mind: "every act of consciousness

we perform, every experience that we have, is intentional… it is essentially consciousness of something". Zahavi (2019) likewise underscores that intentionality prevents a collapse into solipsism by affirming that our experiences are inherently world-involving: whenever we experience, we are engaged with objects, people, or situations beyond ourselves. This concept is highly relevant for cognitive semiotics because meaning is inherently intentional. A sign, for instance, stands for or about something to someone, it exemplifies a triadic intentional relation between a sign vehicle, an object, and an interpretant (in Peircean terms). Cognitive semiotics leverages the notion of intentionality to analyze how meaning is constituted in the relationship between minds and the world. It reminds us that to study meaning-making, we must consider both subjective interpretation and objective reference. Furthermore, intentionality highlights the importance of *intersubjectivity*: because meanings are about a shared world, different subjects can converge on the same meaning. Thus, intentionality leads naturally into questions of how multiple subjects share or coordinate meanings, an issue central to semiotics and communication.

In summary, phenomenological subjectivity and intentionality inform cognitive semiotics by ensuring that any account of meaning remains rooted in first-person experience and the directedness toward a world, in short, meaning is understood as experienced meaning. This perspective guards against purely behaviorist or neuroreductive accounts, insisting that the manifestation of meaning in experience is the ultimate ground of analysis (Zahavi 2019). It also provides cognitive semiotics with a rich conceptual vocabulary (e.g. Lifeworld, embodiment, intersubjectivity) for describing how meaning is lived and shared, which is important in analyzing phenomena like iconicity in language.

Beyond these concepts, cognitive semioticians have adopted phenomenological methodology in the form of *phenomenological triangulation*. This is a key methodological tool that combines insights from first-, second-, and third-person perspectives to study meaning-making in a balanced way. Recently elaborated by Zlatev and Mouratidou (2024), phenomenological triangulation is about examining a phenomenon (such as an act of communication or an experience of an ideophone) from multiple vantage points and integrating the results. It operates along two axes: Horizontally, it acknowledges three irreducible dimensions of the Lifeworld in phenomenology: Self, Others, and Things (i.e. the experiencing subject, other subjects, and the

object or world). Any meaningful event involves all three, a self that is experiencing or expressing something, an intersubjective context with others, and some content or object. Triangulation demands that we consider each of these dimensions in our analysis, rather than, say, studying signs in isolation from their users, or studying cognition in isolation from the world. Vertically, phenomenological triangulation involves three methodological perspectives on any phenomenon: a first-person perspective, a second-person (or socially interactive) perspective, and a third-person (or "objective", but not in the sense of "mind independent", see below) perspective. The first-person perspective is accessed through philosophical reflection and phenomenological methods, for example, asking what an experience feels like or using introspective reports. The second-person perspective involves intersubjective or empathetic/qualitative methods, such as dialogues, interviews, or participant observation, which capture how meaning is negotiated or shared between people. The third-person perspective corresponds to the detached, quantitative methods of science, such as experimental psychology, neuroscience, or behavioral observation, which look at observable correlates of meaning-making (e.g. reaction times, brain activity, communication patterns). By combining these three perspectives, cognitive semioticians aim to "triangulate" on meaning: findings from one perspective are cross-checked and enriched by findings from the others. For instance, the intuition or phenomenological description of how an ideophone sounds vivid (first-person) can be compared with how listeners actually respond in interaction (second-person) and with precise measurements and statistical analysis of the latter (third-person). The horizontal and vertical planes of triangulation together ensure a comprehensive approach, as shown in **Table 1**.

Through phenomenological triangulation, cognitive semiotics guards against biases that come from relying on only one method or viewpoint. In a way, this "operationalizes" the field's pluralism: a meaning phenomenon should be studied subjectively, intersubjectively, and objectively. This methodological principle will be applied in later chapters, for example, by examining iconicity in Japanese ideophones through introspective analysis, cross-cultural elicitation, and the measurement of experimental data, thereby covering the spectrum of first- to third-person perspectives.

Table 1. Extended phenomenological triangulation (Zlatev & Mouratidou 2024).

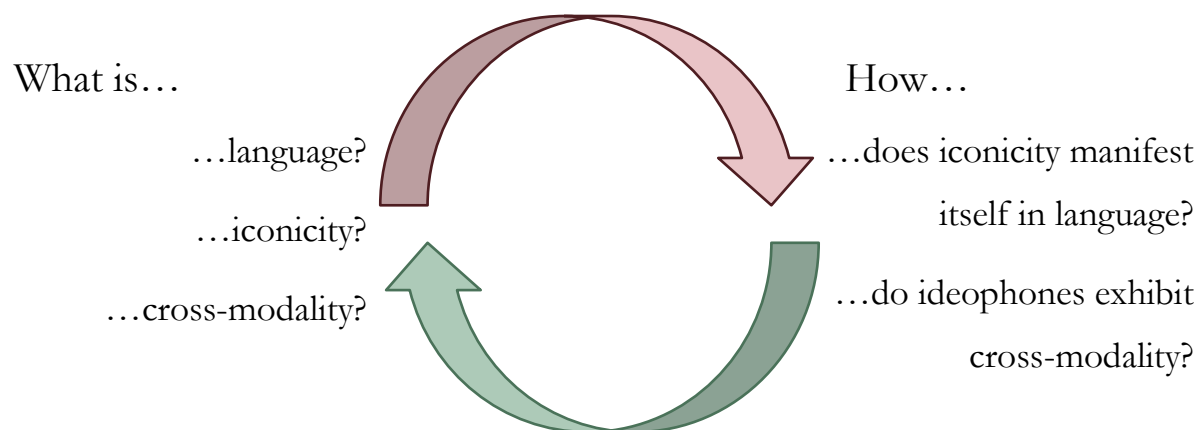| Dimension	Self	Others	Things
Perspective			
First person	Reflection	Empathy	Phenomenological reduction
Second person	Psychotherapeutic dialogue	Interview	Intersubjective validation
Third person	"Third person data" analysis"	Psychological experiment	Causal explanation

2.1.2 The conceptual-empirical loop

A second core methodological principle of cognitive semiotics is the conceptual-empirical loop or "spiral" (Zlatev 2015: 1058; Konderak 2018: 15). This principle captures the idea that research in cognitive semiotics proceeds through an iterative interplay between theoretical concepts and empirical observation. Rather than treating theory and data as separate or hierarchical (with one dominating the other), cognitive semiotics continually integrates conceptual reflection with empirical research in a cyclic process (Zlatev 2012: 4; Konderak 2018). In practice, this means that cognitive semioticians formulate concepts and models of meaning (inspired by philosophy, semiotic theory, linguistics, etc.) and then test or refine them through empirical studies (drawing from psychology, neuroscience, anthropology, etc.), whose results feed back into improved concepts. This looping process prevents purely speculative theorizing on the one hand, and unguided empiricism on the other. Instead, theory is constantly grounded in empirical findings, and empirical work is theory-driven. For example, a cognitive semiotic concept like cross-modal iconicity (a theoretical idea about sound-symbolic links between modalities) can be used to design experiments or observations; the findings from those studies may then lead to a revised definition of iconicity or new distinctions, which can again be tested. Zlatev (2015: 1058) identified "[integration of] theoretical and empirical research" as one of the foundational commitments of the field. Similarly, Konderak (2018: 15) emphasizes that cognitive semiotics "combines conceptual and empirical approaches" in a balanced way. This synergistic method is in contrast to classical semiotics, which often stayed at the level of text analysis and theory, and in contrast to some cognitive science approaches that might accumulate data without a rich concept of meaning. The conceptual-empirical loop thus embodies the transdisciplinary spirit of cognitive semiotics: it

loops between humanities-style interpretation and scientific experimentation to progressively deepen our understanding of meaning-making.

In the context of this thesis, the conceptual-empirical loop provides a guiding framework. The forthcoming chapters will alternate between developing conceptual accounts of key ideas, notably iconicity (how expression can resemble content) and cross-modality (interactions between sensory modalities in language), and presenting empirical investigations of those ideas with respect to Japanese ideophones. Adhering to the cognitive semiotic methodology outlined above, I will first conceptually clarify what cross-modal iconicity entails, drawing on the literature and cognitive semiotic principles; then I will examine empirical evidence (from experiments, corpus analysis, or observational studies) to see how these concepts manifest in real usage of ideophones. This iterative approach not only tests the theoretical ideas against reality but also uses the data to refine and nuance our conceptual understanding. In doing so, it exemplifies how cognitive semiotics seeks a holistic but rigorous account of meaning-making. The phenomenological insights (subjectivity, intentionality, triangulation) and the conceptual-empirical loop together form the methodological foundation for this study of cross-modal iconicity in Japanese ideophones, ensuring that this analysis remains grounded in both human experience and empirical science, true to the transdisciplinary nature of cognitive semiotics.

Figure 1. Conceptual-empirical loop as applied in this thesis.

What is…
…language?
…iconicity?
…cross-modality?

How…
…does iconicity manifest itself in language?
…do ideophones exhibit cross-modality?

2.2 Levels of meaning and iconicity

A common idea within cognitive semiotics is that semiosis (meaning-making) is not a monolithic capacity but is organized in a hierarchy of distinct layers or levels, each with its own characteristics (Sonesson 2012b: 81; Zlatev and Konderak 2023: 170). In Sonesson's account, even before we reach the level of explicit signs, more basic semiotic phenomena exist – for example, perception itself can involve iconic or indexical relationships even without full sign function (Sonesson 2012). Prior to any kinds of signs, is the notion of semiotic ground, proposed by Peirce:

> [a] sign, or **representamen**, is something which stands to **somebody** for **something** in some respect or capacity. […] The sign stands for something, its **object**. It stands for that object, not in all respects, but in reference to a sort of idea which I have sometimes called the **ground** ... (CP 2.229, my emphasis)

One such ground is *iconicity* – the relationship of resemblance or similarity between an expression and what it stands for. Iconicity manifests in different ways at different hierarchical levels, from direct perceptual analogues and mimetic gestures to complex linguistic signs like ideophones. In the following subsections, I examine iconicity hierarchically: first as a ground for signification (2.2.1), then in the context of signs and sign use (2.2.2), and finally within language, especially Japanese ideophones (2.2.3).

2.2.1 Iconicity as ground

In Peirce's triadic theory of signs, as interpreted by Sonesson (e.g. 2010) iconicity is the ground of similarity between the sign's form (expression) and its content (object). Sonesson (2023) notes that iconicity is often simplistically equated with resemblance (e.g. pictures as "iconic signs"), but Peirce's nuanced view divides iconicity into several subtypes (Sonesson 2023: 194). In particular, Peirce distinguished three kinds of iconic grounds: (a) *imagistic* iconicity, where sign and object share simple qualities (as in a photograph or onomatopoeic sound imitation); (b) *diagrammatic* iconicity, where it is the relations among parts of the sign that mirror relations in the object (for example, the arrangement of elements in a diagram resembling the arrangement of parts of a face); and (c) *metaphoric* iconicity, where the similarity in question is not between expression and

content, but between different contents, for example between the literal and metaphorical sense of the word *star* (Jakobson 1965). Sonesson emphasizes that such iconic grounds are not yet full signs in themselves: they are potentials for signification. A ground of resemblance only yields an actual iconic sign when it is integrated with the *sign function* – i.e. when someone uses or interprets the resemblance as standing for (or *signifying*) something (Sonesson 2012a). In other words, the perception of similarity is a cognitive link between two entities, which may motivate a sign, but without a sign-user's intentional linking of expression and content, we have iconicity per se rather than an iconic sign. This insight allows Sonesson to separate the emergence of iconicity in cognition from the emergence of iconic signs in communication (Sonesson 2012a).

Empirical research in developmental psychology supports the idea that human children are sensitive to iconic relations very early, even prior to language. For example, Imai and Kita (2014) found that both adults and young children can detect subtle sound-meaning resemblances in words – a phenomenon known as "sound symbolism" – without needing explicit instruction (Imai and Kita 2014, as cited in Giraldo 2020: 42). In other words, children are able to perceive an iconic cross-modal correspondence between the sounds of certain words and their meanings. This suggests that iconic grounding may facilitate word learning by providing an intuitive link between expression and content. Indeed, such findings challenge the traditional Saussurean doctrine that linguistic signs are arbitrary, as mentioned in Chapter 1, and elaborated in Section 2.2.3. Instead, there is evidence that some and possibly many words in most languages have a motivated, non-arbitrary linkage between expression and content. Thus, iconicity (in this case, resemblance between phonetics and semantics) is available as a semiotic ground that language learners intuitively respond to from early on, and use later when acquiring and using language.

Another line of evidence for the cognitive reality of iconic ground comes from cross-modal iconicity studies. A classic example is the "bouba/kiki" paradigm, where people consistently match a nonce word such as *bouba* to a curvy, rounded shape and another word like *kiki* to a jagged, spiky shape. This mapping occurs above chance without any prior knowledge, indicating a direct cross-modal iconic perception. Ahlner and Zlatev (2010) investigated this effect in detail and introduced a useful distinction (building on Sonesson 1997) between *primary* and *secondary* iconicity. Primary iconicity means that the resemblance is transparent enough that one can

understand the sign's meaning on the basis of this (as when one immediately interprets a picture of a cat as "cat"). Secondary iconicity, by contrast, means the resemblance is only noticed after one knows what the sign is supposed to mean (for instance, once you are told that a certain abstract logo represents a cat, you might then discern stylized whiskers or ears in its design), as in Figure 2.

Figure 2. Logo of *Mistral AI*, creators of "Le Chat," an example of secondary iconicity
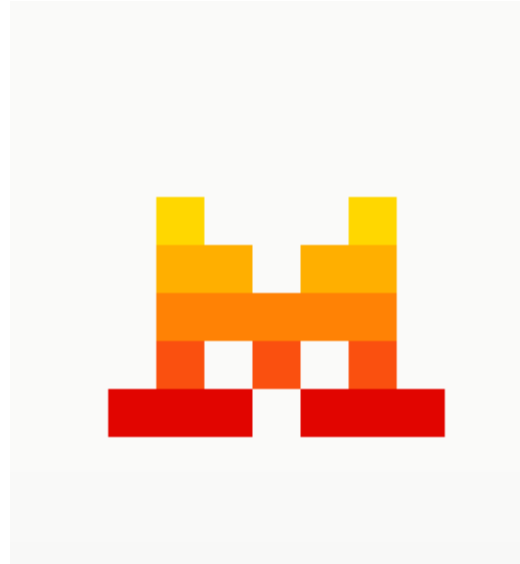
In their cross-modal experiments, Ahlner and Zlatev (2010) found that both types of iconicity work in tandem. Participants were first told that there is a correspondence between certain sound-forms and shapes – effectively establishing a sign relation in their mind (a step of secondary iconicity, since they now know these sounds are meant to refer to these images). However, they were not told which word goes with which shape. To make the correct pairing, participants then relied on the perceived similarity between the word's sound and the shape's visual properties – a step of primary iconicity (using the "natural" sound-shape association). In sum, the interpreter first recognizes an expression–object pairing is intended (the sign function is acknowledged) and then discerns the analogical ground linking them to decide on the mapping. Ahlner and Zlatev's analysis showed that iconic interpretation often involves a sequence: establishing that a sign relation exists (cueing the search for resemblance) and then exploiting resemblance to infer the specific meaning. Crucially, this demonstrates that cross-modal iconicity is not a marginal curiosity, but a cognitive

process combining reflective knowledge of a sign relation with direct cross-sensory resemblance (Ahlner and Zlatev 2010: 314‑319). Even resemblances across sensory modalities (sound → vision, etc.) can be primarily iconic if they are immediately graspable, and research indicates that adults and children alike can make such resemblances before being explicitly taught the meanings. Thus, iconicity as a ground spans a continuum from the obvious and direct to the subtle and discovered, playing a foundational role in how signs come to be established and understood.

2.2.2 Iconic signs

When an iconic ground is elevated to a full sign, it enters into a more complex structural and cognitive arrangement. In cognitive semiotics, a sign is typically defined in triadic terms, following the definition of Peirce quoted above, but with important phenomenological refinements. A useful modern definition is given by Zlatev et al. (2020), who specify that a sign involves at least two interrelated parts, an expression (E) and an object (O), and crucially involves a subject's reflective awareness of the relation between them. In simplified terms, *something (E) signifies something else (O) for someone (S) who recognizes this relation*. Two conditions must be met: (a) the subject becomes aware of an intended object through a perceivable expression, and (b) the subject is (at least potentially) aware that condition (a) is taking place (i.e. aware of the relation between E and O as a directed, meaningful relation). Condition (a) corresponds to the basic *intentionality* of the sign: the expression directs the subject's mind to an object. Condition (b) captures the *reflective, signitive intentionality* that characterizes true sign use: the subject knows that the expression is serving to represent the object. This reflective awareness is what distinguishes a sign from a mere *signal* or *association*.

Figure 3. A diagrammatic view of signification.

Object (O)

iconicity
indexicality
symbolicity

Semiotic Ground

consciousness

Subject (S)

meaning

Expression (E)
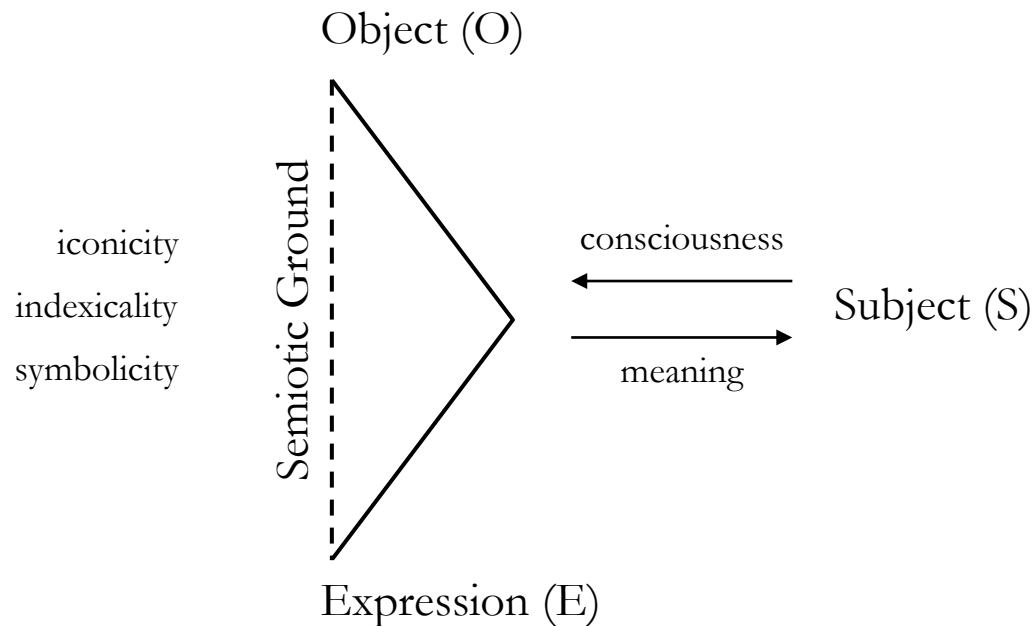
To illustrate, consider an animal alarm call versus a human word. A vervet monkey's leopard alarm call will cause other monkeys to run up a tree. The call is an *indexical signal* – it reliably co-occurs with a leopard's presence (contiguity in space/time), and receivers respond appropriately. However, the monkeys do not appear to be reflectively aware that this sound signifies that predator. They react to the call but do not use it as a sign with conceptual awareness of reference. By contrast, when a human child of few years of age hears the word "leopard" and understands it, she not only perceives the sound and thinks of a leopard (condition a) but also knows that this word is a sign that stands for the animal (condition b). The latter involves what Zlatev et al. call a particular kind of reflective, signitive consciousness".

It is this reflective link – the awareness of the expression–object relationship – that produces what Sonesson calls "the double asymmetry of the sign". In an established sign relation, the expression is typically more immediately present to our senses, while the content (or object) is more salient in our *attention or interest* (Sonesson 2012: 85). For example, when reading a word or looking at a pictorial sign, we tend to focus on what it means (the content) rather than on its physical form, yet that content is accessed through the form which is forefront in our perception. Moreover, the relation is asymmetric in that the expression directs us to the content, not vice versa

(the word leopard brings to mind the animal, but the animal does not inherently call to mind that word). This double asymmetry – the expression being more tangible and the content more meaningful – is a hallmark of signs and arises because a sign user differentiates the sign from its referent. In Husserlian terms, the sign (as expression) is an object of consciousness that "points to" another object beyond itself, and we are aware of this pointing structure (Sonesson 2012: 85; Zlatev and Konderak 2023: 183–184).

With this definition in mind, we can define iconic signs as those signs in which an iconic ground (resemblance) is dominant in the expression–object relation. However, it is important to note that any actual sign may involve multiple semiotic grounds at once. As Zlatev and Konderak (2023) explain, indexicality (based on factual contiguity or cause-effect link) and symbolicity (based on convention or arbitrary rule) can co-occur with iconicity in a single sign. For instance, a pointing gesture is primarily an indexical sign (its efficacy comes from a physical or deictic connection to the referent's location), but it often has an iconic aspect (e.g. the hand shape might mimic a certain directional movement) and a conventional aspect (certain pointing norms are learned socially). Nevertheless, we categorize a sign by the predominant ground: an iconic sign is one whose interpretive success relies mainly on resemblance, an indexical sign (or pure index) relies on direct connection or correlation, and a symbolic sign relies on habitual or agreed convention (Sonesson 1989; Peirce 1931–58). In cognitive semiotic terms, an index or a symbol can be seen as specializations of the sign function where the ground is, respectively, contiguity or convention, rather than similarity. All require a subject's understanding of the expression–object relation, but they differ in how that relation is established. Icons "make sense" by themselves through resemblance (at least to an informed observer), indices often show or indicate something by "closeness", and symbols tell by mutual agreement (requiring a cultural code).

It is the presence of "signitive, reflective intentionality" that separates genuine signs from mere natural indices or symptoms. For example, smoke naturally indicates fire (smoke is causally produced by fire), so smoke is an index of fire; but smoke is not a sign for fire unless some interpreter actively takes it to stand for fire (e.g. a scout reading smoke signals). Likewise, a stylized drawing of fire can be an iconic sign for fire only if we intentionally use that resemblance to denote fire. In human semiosis, the double asymmetry mentioned above arises only when this

reflective stance is present (Zlatev et al. 2020: 160). Without reflective awareness, an organism might notice a similarity or react to a cue, but it would not distance the expression from the referent in the way a sign-user does (Sonesson 2012: 85). This theoretical stance aligns with the view that a degree of reflective consciousness is a precondition for genuine sign use (Eco 1976; Sonesson 2007), while semiosis is more general phenomenon, including signal use and direct perception. In summary, an iconic sign in the full sense is a triadic relation: the expression resembles the object, and the subject knows and utilizes this resemblance to direct attention or meaning – a process rooted in reflective awareness.

2.2.3 Iconicity in language

Applying the above framework to language reveals that iconicity plays a far more significant role than traditionally acknowledged by structuralist theory. Saussure (1916) famously argued that the linguistic sign is "arbitrary", i.e. that there is no motivated relation between the sound of a word and its meaning. However, evidence from sound symbolism and iconic language contradicts this thesis. From a cognitive semiotic perspective, very few linguistic signs are *entirely* unmotivated; in other words, most have some non-arbitrary *ground* (iconic or indexical) in addition to conventionality (Giraldo 2020: 43). Giraldo (2020) emphasizes that language, like other human meaning systems (e.g. gesture and pictorial depiction), relies on iconic and indexical grounds to a considerable degree, and these grounds are "important in all cases" – directly challenging the Saussurean dogma of arbitrariness. Thus, "sound symbolism" is a prime example of linguistic iconicity: certain speech sounds or patterns resemble aspects of meaning (whether by evoking a sensory property, a size, a movement, etc.). For instance, the high front vowel [i] in words like *teeny* or *ci** (this) often correlates with smallness or proximity, whereas an open back vowel like [a] (as in *large* or *that*) may suggest largeness or distance – patterns that appear in many languages (Johansson and Zlatev 2013, and see Ohala 1994 on the "frequency code"). These are iconic *mappings* between the phonetic space and the conceptual space (here, sound frequency and the notion of size or distance). Such sound–meaning correspondences undermine a strict arbitrariness by showing systematic motivation. On may conclude that few (if any) words are purely arbitrary in form; instead, iconic or indexical motivations (however subtle) often lurk beneath the conventional veneer.

18

A particularly vivid domain of linguistic iconicity is that of *ideophones* – also known as "expressives" or "mimetics" – which have been described as "marked words *depictive* of sensory imagery" (Dingemanse 2012: 654), a characterization that reflects their iconic ground, either imagistic or diagrammatic. Japanese, for example, is rich in ideophones that represent, with a combination of primary and secondary iconicity (see Section 2.2.1) events (e.g. *gorogoro* for a rolling sound or state, *pikapika* for something bright/shiny, etc.). These words exhibit sound symbolism by design. Linguistic research (Dingemanse 2012; Dingemanse et al. 2015; McLean 2021) has established ideophones as a legitimate object of study and shown that their distribution across languages follows certain patterns. Dingemanse (2012) proposed an implicational hierarchy of ideophone meanings in various languages, which Giraldo (2020) summarizes and cites. This hierarchy is shown in (1).

(1)

  SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS <
  INNER FEELINGS/COGNITIVE STATES

It means that if a language has ideophones for a relatively abstract or internal domain (say, mental states or inner feelings), it will also have ideophones for all "lower" (more perceptually concrete) domains like sound and movement. In simpler terms, the presence of less iconic or less directly imitative ideophones (e.g. for emotions or cognitive states) in a language implies the presence of more concrete ones (especially for sound). The most common ideophones universally are those for sound, essentially onomatopoeias (since imitating sounds with sound is a straightforward one-to-one iconic mapping). These are unimodal iconic mappings (sound-to-sound: using spoken sound to represent environmental sound). As one moves up the hierarchy, ideophones increasingly represent one modality via another – for example, using sound to signify visual patterns or movement. Such cases involve cross-modal iconicity: e.g. a word's phonetic features (auditory) might evoke a visual shape or a tactile sensation. According to the hierarchy, ideophones for movement or visual patterns (like a word resembling the manner of motion or the shape/appearance of something) will typically be found only in languages that already have sound-

imitative ideophones. And if a language extends ideophonic representation to inner feelings or abstract states, it will have ideophones for all the "lower" sensory realms as well. This implicational hierarchy highlights that iconicity spans from straightforward intra-modal imitation to more abstract inter-modal analogies. Sound-to-sound mappings (onomatopoeia) are the easiest, hence most widespread.

Cross-modal mappings (sound to vision, sound to touch, etc.) are more complex and are found in languages with a robust ideophone system that has expanded to cover these domains (Dingemanse 2012: 663). Japanese, notably, has ideophones not only for sounds (e.g. *zaza* 'sound of rushing water') and movements (*yukkuri* 'slowly') but even for visual patterns (*chirachira* 'flittering, e.g. of falling snow') and psychological states (*wakuwaku* 'excitement'), exemplifying the full span of the hierarchy.

From a cognitive semiotic view, the existence of ideophones and sound symbolism in language shows that iconicity, indexicality and symbolicity (conventionality) as grounds for linguistic signs coexist and complement each other. Far from being "mere curiosities" or "exceptions" to arbitrariness, iconic phenomena like ideophones suggest that language users exploit resemblance whenever useful to ground meaning. Giraldo (2020: 43) argues that iconicity and indexicality are central in language alongside symbolicity, making language a motivated semiotic system rather than a purely arbitrary code. The Saussurean view, which placed arbitrariness as fundamental, is thus contradicted by the recognition that motivations (natural analogies, experiential correlations) pervade language (Perniss et al. 2010; Dingemanse 2018). Sound-symbolic patterns, such as the "frequency code" where high-frequency sounds connote smallness or brightness and low-frequency sounds connote largeness or darkness (Ohala 1994), likely arise from embodied experiences, and provide a natural basis that languages can conventionalize to various degrees. Even in domains like grammar, one finds one form of diagrammatic iconicity (see Section 2.2.1) – e.g. the ordering of words reflecting the sequence of events or the closeness of concepts (Haiman 1985). The cognitive semiotic framework allows us to see all these as manifestations of iconic grounding operating within the sign system of language.

Japanese ideophones and similar phenomena in other languages demonstrate that iconicity in language operates on multiple levels. While most iconicity in language may be secondary

(Sonesson 2010), the evidence from developmental psychology and experiments such as those summarized in Section 2.2.1, show that it interacts with primary iconicity: the similarity itself places a key role for interpreting the signs.

Some iconic words are unimodal (using a medium to imitate the same kind of sensory experience, as in onomatopoeic sound imitation of sounds), while others are cross-modal (using one medium to suggest a different sensory modality, as when the way a word sounds evokes a visual or tactile sensation). The implicational hierarchy of ideophone types in (1) underscores how languages build outward from the most accessible iconic mappings (sound-for-sound) to increasingly abstract ones (sound-for-feeling), retaining the more concrete mappings as a foundation. Thus, iconicity bridges the gap between embodied experience and linguistic expression. Rather than undermining the systematicity of language, such iconic correspondences enrich language, providing additional layers of meaning-making that are grounded in human perception and cognition. In cognitive semiotic terms, linguistic signs, even highly conventional ones, can be seen as sedimented layers of semiosis that often preserve an "iconic echo" of the experiences they originate from. The study of Japanese ideophones within this framework exemplifies how iconicity can be regarded as a phenomenon that may cross modalities.

2.3 Relevant empirical studies

Recent research has increasingly shown that iconicity is not a peripheral curiosity but a general property of human language (Dingemanse et al. 2015; Perniss, Thompson & Vigliocco 2010). Across languages, certain sounds or sound patterns recur in association with particular meanings, suggesting sound symbolism has both language-specific and potentially universal aspects (Erben Johansson et al. 2020). Ideophones, with their "special semiotic and sound-symbolic properties", have become a key testing ground for these ideas, serving as "an attractive target for studies of multi-modality, iconicity and sensory language" (Dingemanse 2018: 1). This section reviews empirical studies on Japanese ideophones and related sound-symbolic phenomena, focusing on (i) the degree of primary vs. secondary iconicity in ideophones across semantic domains, (ii) cross-linguistic versus language-specific sound-symbolic patterns, (iii) the influence of orthography and the visual modality (particularly Japanese kana scripts), and (iv) norming studies on the perceptual

strength of ideophones. Together, these studies provide the empirical foundation and highlight the research gap addressed in this thesis.

As discussed in section 2.2.1, iconicity can be categorized as *primary*, where the resemblance between form and meaning is transparent enough that one can understand the sign's meaning from the expression alone, and as *secondary*, where the resemblance becomes apparent only once one knows what the sign means (Sonesson 1997; Ahlner & Zlatev 2010: 315). Ideophones and experimental phenomena such as the bouba/kiki effect arguably exhibit a combination of both types, since on the one hand, one cannot understand what the word signifies with certainty; on the other hand, when asked to "guess", preferences are not random. Classic anecdotal evidence suggests that even listeners with no knowledge of a language can often guess the meaning of foreign ideophones at above-chance rates, indicating an element of primary iconicity. For example, many languages use reduplicated forms to depict repeated actions or sensations (e.g. Japanese *dokidoki* `ドキドキ for a heartbeat), a pattern that naïve listeners can intuitively associate with repetition. Experimental studies confirm that such cross-modal mappings are broadly available: when presented with matching tasks, participants often correctly pair unfamiliar ideophones with their meanings (or related images/sounds) more often than chance would predict (Lockwood, Dingemanse & Hagoort 2016; see also experimental overviews in Dingemanse et al. 2015). This suggests that certain ideophonic forms carry direct sensory cues (e.g. a word's acoustic properties resembling a sound or a texture) that can be recognized without prior learning (i.e. primary iconicity).

However, the same studies also reveal a role for secondary iconicity: not all ideophones are likely to be equally as transparent to foreign listeners, as some might be more readily understandable than others. Consistent with the implicational hierarchy (1), Dingemanse (2012; McLean 2021) observed that if a language has any ideophones, it nearly always has sound imitative ones which are unimodal (i.e. sound to sound), whereas ideophones for inner feelings or abstract concepts appear only in languages that already have sound and movement ideophones (see section 2.2.3 (1)). Japanese, notably, has ideophones across this entire spectrum (e.g. *zaza* ザザ "sound of rushing water," *yukkuri* ゆっくり "slowly (movement)," *kirakira* キラキラ "sparkling (visual)," *tsuntsun* ツンツン "prickly (tactile)," *uron* うろん "absent-minded/confused (internal state)"). The

prevalence of sound and motion ideophones suggests these domains are more naturally mapped to speech, presumably because auditory and motor experiences share more structural properties with spoken form, enabling a degree of primary iconicity (McLean 2021). By contrast, internal feelings or mental states might require more cultural convention to encode iconically, reflecting secondary iconicity.

McLean (2021) tested this hierarchy in Japonic languages, finding that speakers readily produced ideophones for stimuli in the auditory, motion, shape, and texture domains, while more abstract domains were less spontaneously ideophonizable. These results mirror the cross-linguistic distribution of ideophones and support the idea that certain perceptual mappings (especially sound-to-sound or sound-to-movement) are inherently easier, essentially grounded in primary iconicity, whereas others rely on extending familiar patterns or metaphors (secondary iconicity, or greater proportions of symbolicity) to be established.

A growing body of quantitative research addresses how much iconic expression-object correspondences are universal tendencies versus language-specific conventions. Cross-linguistic corpus studies have identified some widespread sound-symbolic patterns. For instance, the so-called "frequency code", high-frequency sounds (like /i/) associating with small or bright concepts, and low-frequency sounds (like /u/ or /o/) with large or dark concepts, appears in many languages' vocabularies (Ohala 1994; Dingemanse 2018). Likewise, an analysis of over 6,000 basic vocabulary items by Blasi et al. (2016) found non-arbitrary phonetic associations (e.g. nose frequently contains nasals across languages) beyond what chance would predict. Such findings reinforce that human cognitive biases (or shared experiences) can give rise to similar iconic links independently (Perniss & Vigliocco 2014). Erben Johansson et al. (2020) provide a typological overview, showing that certain broad semantic categories (like animal sounds, bodily actions, etc.) commonly exhibit vocal iconicity across unrelated languages. There is a cross-linguistic substratum of iconic associations, a potential universality of primary iconicity in certain domains. At the same time, each language develops its iconic lexicon in unique ways. Japanese sound symbolism, for example, includes patterns that must be learned to be appreciated. A striking case is the *voicing contrast* in Japanese ideophones (Hamano 1998; Kawahara 2020): many ideophone pairs differ only by a voiced vs. voiceless consonant, yet their meanings follow a consistent

semantic shift: the voiced form tends to indicate something heavier, larger, or more intense than the voiceless form (e.g. *chīsan* チーサン vs. *jīsan* ジーサン in a hypothetical pair) – a pattern sometimes called the "gada-kata" alternation or voice-set.

An English speaker without knowledge of Japanese might not perceive *zarazara* (ザラザ ラ) as "rough" rather than "smooth"; indeed, McLean et al. (2023) found one listener mis-guessed the meaning of *zarazara* because they assumed it meant "smooth," whereas Japanese speakers, knowing the voicing convention and the existence of *sarasara* ("smooth"), would correctly understand *zarazara* as "rough". Here the iconic mapping (fricative /z/ ~ continuous sensation) can be interpreted in opposite ways, hinting at greater degrees of symbolicity (or conventionalization). More generally, Thompson, Akita & Do (2020) collected iconicity ratings for words across the Japanese lexicon (from both Japanese and English speakers) and found evidence of conventionalization: native Japanese speakers tended to rate frequent ideophones as highly iconic (perhaps due to entrenched sound-meaning associations in the lexicon; the authors present some additional caveats), whereas English speakers' ratings were lower or more variable for those same items. Some words that Japanese speakers consider strongly iconic did not sound inherently meaningful to outsiders, a sign that part of the iconic "feel" comes from internalized lexical patterns (secondary iconicity through exposure) rather than perceptually salient auditory resemblance alone.

To disentangle these factors, McLean, Dunn & Dingemanse (2023) conducted a study using two complementary measures of iconicity: a *guessing task* (measuring immediate, naive understanding of a word's meaning) and an *iconicity rating task* (measuring the experienced strength of "form-meaning" resemblance). They tested 304 Japanese words (including many ideophones) with English-speaking participants. The results were illuminating. Firstly, the two measures were positively correlated, indicating they tap into the same underlying phenomenon to an extent (words that were easier to guess tended to be rated as more iconic). However, there was a systematic difference between ideophones and regular words. Ideophones received significantly higher iconicity ratings than non-ideophonic words even when both types were guessed with equal (low) accuracy (McLean et al. 2023: 732), suggesting that participants felt ideophones were somehow meaningful or fitting, despite often failing to guess them correctly. In other words, the

very form of ideophones (e.g. their atypical phonology or reduplication) acted as a cue that "this is an iconic word," leading to high iconicity judgments: a secondary iconicity effect, since it relies on recognizing the sign's evocative intent. Yet those formal cues did not always reveal the actual meaning (limited primary iconicity), which explains the poor guessing performance for many ideophones. Notably, raters were more consistent in their ratings for ideophones than for ordinary words, even when wrong, presumably because the ideophonic forms invited them to impose some interpretation and boosted their conviction that a resemblance exists (McLean et al. 2023: 734). This finding resonates with Ahlner and Zlatev's (2010) two-step model of iconic interpretation (see Section 2.2.1): once listeners suspect a sign is meant to be iconic (the sign function established, akin to secondary iconicity), they find an analogical mapping (primary iconicity) to guess the meaning. McLean et al.'s English participants clearly recognized Japanese ideophones as signifying something (hence high iconicity ratings), but without the cultural knowledge they sometimes mapped the wrong feature (e.g. focusing on one phonetic aspect that they interpreted differently).

From a cross-linguistic perspective, these results highlight that iconicity has both universal and language-specific layers. Universally, people share intuitions that certain word forms ought to carry meaning (e.g. reduplicated, rhythmic words seem descriptive of repeated events, voiced fricatives might suggest something continuous or thick, etc.). But the actual interpretation of those expressions can diverge across languages due to conventional calibration.

As an example, many languages use reduplication iconically (for plurality, repetition, intensity, etc.), a testament to a shared primary iconic insight. Yet the specific patterns (e.g. what reduplication signifies in each language, or what voicing contrasts mean) are conventional. A recent cross-linguistic study by Punselie et al. (2024) introduced the concept of "cumulative iconicity" to quantify how many iconic mappings are stacked in a given word. They analyzed 239 ideophones from five languages (including Japanese) and scored each word on multiple possible correspondences between "form and meaning" (sound imitation, structural repetition, syllable pattern mirroring event structure, etc.). Crucially, words with higher cumulative iconicity scores were rated as more iconic and were more guessable by non-native listeners. For example, a Japanese ideophone like *dokidoki* (ドキドキ "heartbeat") has repetitive syllables matching the

iterative thump of a heart, a bipartite structure aligning with the "lub-dub" cardiac cycle, and onomatopoeic sound quality, multiple layers that may give it a high iconicity score. Punselie et al. found that no ideophone with a high cumulative-iconicity (many structural analogies) went unnoticed as iconic, showing that these features are universally salient. On the other hand, there were ideophones that naive listeners rated as only moderately iconic or had trouble guessing, even though those words are considered iconic by native speakers. These tended to have lower cumulative-iconicity scores in the analysis, suggesting that the study's five structural features didn't fully capture whatever subtle cues natives rely on. Such cases likely involve language-specific iconic conventions outside the coded features, again underscoring that iconicity can be highly evident when multiple perceptual analogies align (primary iconicity), but in other cases it emerges through learned associations and contextual knowledge (secondary iconicity, or by further extension, symbolicity).

In sum, empirical studies of Japanese ideophones demonstrate a continuum from straightforward iconic signs (immediately apparent across audiences) to less straightforward ones (iconic only to the initiated), and this continuum is what the notions of primary and secondary iconicity, operating in "proportions" rather than either-or, meaningfully capture.

2.3.1 Orthography and the visual modality

Another factor that can modulate iconicity is the mode of presentation, spoken vs. written. In experimental settings, researchers noticed that the famous bouba-kiki effect (again, the tendency to match nonsense words like *bouba* with round shapes and *kiki* with spiky shapes) might partly be influenced by the visual shape of the written words. Cuskley, Simner and Kirby (2017) investigated this by separating phonological and orthographic influences. They found that literate adults' bouba-kiki responses were "heavily mediated" by the angularity of the letters: nonwords containing angular letters like 'K' were much more likely to be paired with spiky shapes, whereas those with rounded letters like 'B' tended to be paired with round shapes. In their analysis, orthographic form accounted for over half of the variance in a written task and was the sole significant factor influencing ratings. This suggests that for literate individuals, the written shape itself can create a within-modal iconic link (spiky-looking letters to spiky shape) that functions as a form of secondary iconicity.

However, other studies have shown that orthography does not play a solely decisive role. Research with pre-literate children and with cultures using different scripts still found bouba-kiki-like effects, suggesting a deep auditory-visual mapping that exists independently of knowing letter shapes. Ćwiek et al. (2021), for instance, tested the bouba-kiki effect across numerous languages and their corresponding writing systems (including alphabets, abjads, and syllabaries) and reported that the effect is robust regardless of the specific script. They found consistency even in cultures that are mostly oral, implying an underlying cross-modal association between speech sounds and shape qualities that does not require letter shapes. The likely reconciliation of these findings is that both factors are at play: a genuine, cross-modal correspondence exists between sound and shape, but when written forms are introduced to literate participants, an extra unimodal iconicity may emerge through the visual properties of the letters, which can enhance or even "overshadow" the pure sound-based iconicity (Cuskley et al. 2017: 125).

This issue is particularly pertinent for Japanese ideophones because the language utilizes two phonographic scripts with distinct visual styles. The two kana syllabaries *hiragana* ひらがな (characterized by curvy, flowing strokes) and *katakana* カタカナ (characterized by angular, sharp strokes), are often used for stylistic purposes. Ideophones are frequently written in katakana to mark them as vivid or emphatic. This practice aligns with the "script relativity hypothesis," which argues that the perceptual and cognitive processes involved in reading accommodate the specific visual and structural features of a given writing system (Pae 2020) . For example, the block-like structure of the Korean Hangul script, which combines letters into syllabic units, leads its readers to prefer segmenting syllables into a "body-coda" (CV-C) structure, a different cognitive approach from the "onset-rime" (C-VC) segmentation common among readers of linear alphabets like English (Pae 2020).

Applied to Japanese, a similar hypothesis could be made. Katakana's angular geometry might visually amplify perceptions of sharpness or energy in a word, whereas hiragana's roundness might enhance perceptions of softness, much as Cuskley et al. (2017) found with the letters 'K' and 'B', respectively. There is also a simpler interpretation: because katakana is conventionally used for onomatopoeia and loanwords, readers may be conditioned to treat words written in that script as being more focused on sound-meaning relationships. In this view, the script acts as a

conventional cue to "read this iconically" rather than exerting an influence through its inherent shape. Empirical evidence on how visual features of scripts affect sound-matching is emerging. Turoman and Styles (2017), for instance, presented participants with letter pairs from 56 real, unfamiliar ancient scripts to match with the vowel sounds 'oo' /u/ and 'ee' /i/. They found that participants could guess the correct letter at an above-chance rate. Their analysis revealed that the key visual property driving this effect was not spikiness, but rather visual complexity: the letter with more "ink" (a higher rate of black/white cycles detected across spatial frequencies) was more likely to be guessed as /u/. This suggests that people can latch onto various available visual features, such as line length or complexity, as potential iconic cues for sound.

The present study explicitly tests orthographic influence by comparing participant responses in conditions with audio-only presentation vs. written presentation in different scripts. Based on the literature, the expectation is that seeing an ideophone in the visually angular katakana script might heighten cross-modal iconic judgments relative to hearing it alone or seeing it in the curvier hiragana script, effectively boosting its primary iconicity through the visual modality.

A further body of research indicates that Japanese orthography itself may modulate perceptions of iconicity. In particular, Iwahara, Hatta and Maehara (2003) demonstrate experimentally that script types such as hiragana and katakana evoke distinct semantic associations in readers. Hiragana was found to carry softer and more graceful connotations, whereas katakana was associated with sharper, harder, or more emphatic impressions (Iwahara, Hatta & Maehara 2003: 378–379). These findings suggest that script choice in Japanese is not purely a matter of convention but may reflect systematic correspondences between visual form and semantic impression. This line of evidence is directly relevant to ideophones, which are frequently written in kana and whose perceived iconicity may thus be shaped not only by their phonological form but also by the orthographic script in which they are presented.

2.4 Summary and revised research questions

Chapter 2 has outlined the theoretical and empirical background for this study. We have seen that iconicity in language is pervasive yet multifaceted. Especially in the case of Japanese ideophones, iconic meaning arises through a combination of direct cross-modal resemblance and learned

conventions. Based on the cognitive semiotic literature, I introduced the distinctions between primary iconicity (direct, immediate resemblance) and secondary iconicity (resemblance recognized through knowledge of a sign's meaning or context) and discussed how these apply to cross-modal language signs. I also reviewed how different sensory modalities vary in iconic affordances: auditory and motion events are more readily expressed in sound (hence widely attested ideophone domains), whereas internal sensations rely more on cultural coding or symbolicity. Additionally, I considered the influence of script as an often-overlooked factor that can shape the interpretation of ideophones by providing visual iconic cues or marking of "sound symbolism." Norming studies confirmed that ideophones are among the most perceptually evocative words, supporting a cognitive-semiotic view that they serve as iconic signs within language.

Despite these advances, there remain gaps in our understanding of how ideophones are perceived by listeners (especially those unfamiliar with the language) and how multiple sensory modalities contribute to their interpretation. In particular, we still lack a clear picture of how primary and secondary iconicity interact in real time: Do listeners rely purely on sound to grasp meaning, or do they immediately seek confirmation from context or script? Does iconicity function similarly across sensory domains, or are some modalities inherently more iconically accessible? And to what extent does Japanese orthography (hiragana vs. katakana) modulate the iconic perception of ideophones?

Drawing on the literature review, we can revisit the same research questions (RQs) with some additional context to better address these gaps.

RQ1 What is the relation between primary and secondary iconicity in the perception and interpretation of ideophones?

In concrete terms, how well can listeners who do not speak Japanese interpret the meanings of ideophones from sound alone, and does providing additional knowledge or context (such as a written form or hint at meaning) change their performance? This question examines to what extent the iconicity of ideophones is primary versus secondary. Primary iconicity would enable

above-chance understanding of some ideophones, whereas predominantly secondary iconicity (knowing the sign's meaning) would imply that non-Japanese speakers should be at chance level, without strong additional cues. The interplay of the two will be observed by comparing tasks that involve naive guessing vs. those that involve informed rating or feedback.

RQ2 Which sensory modalities are perceived as more iconic?

More specifically, are ideophones linked to certain sensory modalities (e.g. sound or movement) consistently easier for people to identify or find iconic than those depicting visual patterns or internal feelings? This question is inspired by the implicational hierarchy of ideophone meanings (Dingemanse 2012) and related findings that, for example, auditory ideophones might be more universally guessable than interoceptive ideophones. This thesis investigates whether participants perform better (or report stronger iconic impressions) with ideophones from some domains over others. A cognitive-semiotic account would predict that ideophones closer to bodily experience, especially those imitating sounds or direct sensorimotor patterns, have a more immediate iconic ground (primary iconicity), whereas those that rely on more abstract or interoceptive experiences may require more interpretive work and conventionality (secondary iconicity). By comparing response accuracy and iconicity ratings across semantic categories (sound, motion, visual texture, internal state), one could identify if there is a modality-based hierarchy of iconicity accessibility in the experimental data.

RQ3 To what extent does Japanese orthography influence perceptions of iconicity?

Specifically, does seeing an ideophone written in a particular script (katakana vs. hiragana) or not seeing it in writing at all affect how participants perceive its sound–meaning connection? This question probes the polysemiotic interaction between the visual form of language and its vocal form, in the perception of iconicity. Presenting an ideophone in katakana (as opposed to just hearing it, or seeing it in hiragana) might enhance listeners' inclination to treat it as an

iconic sign, possibly improving their confidence or altering their interpretation. This would manifest as differences in guessing accuracy or rated iconicity between an audio-only condition and a text-present condition, and between different scripts. The outcome will tell us whether the visual sign vehicle in language can modulate cross-modal iconic perception – a finding with implications for how polysemiotic and multimodal the sign function can be (e.g., whether the written sign invites a form of iconic insight, as suggested by Cuskley et al. 2017). One can interpret any orthographic effects in light of the script relativity hypothesis (Pae 2020) and the convention of using katakana to represent ideophones, in line with the script's historical origin as a pronunciation guide.

In the line with the above, the original research questions can be revised as follows:

RQ1   What is the relationship between primary and secondary iconicity in the perception and interpretation of Japanese ideophones?

RQ2   Which sensory modalities are perceived as more (unimodally and cross-modally) iconic in Japanese ideophones?

RQ3   To what extent does Japanese orthography influence the perception of iconicity in Japanese ideophones among non-speakers?

Chapter 3 Methodology

This chapter describes the methodological approach taken to investigate the research questions outlined in Chapter 2. The study adopts a cognitive-semiotic framework, applying phenomenological triangulation to integrate three perspectives: first-person reflection, second-person participant reports, and third-person experimental data. The first-person perspective establishes the conceptual grounding for the study, the second-person perspective is addressed, above all, through post-experiment questionnaires that capture participants' subjective strategies, and the third-person perspective is operationalized through the controlled design of the choosing and rating tasks. The chapter is structured as follows. Section 3.1 introduces the first-person methods, Section 3.2 presents the second-person methods, and Section 3.3 details the third-person methods, including participants, materials, and experimental design. Section 3.4 concludes with a summary and the hypotheses that guided the empirical investigation.

3.1 First-person methods

In line with the phenomenological approach outlined in section 2.1, I used the first-person perspective to inform the research design. Researcher intuition and introspection played a key role in generating hypotheses and selecting stimuli. Before formal data collection, I reflected on my personal impressions of Japanese ideophones, listening to their sounds, considering their meanings, and gauging the iconic resonance between sound and sense. This reflective process helped identify potential sound-meaning correspondences and oppositions. For example, I felt that *kirakira* (キラキラ, 'glittering, sparkling') conveys a bright, shimmering quality, whereas *donyori* (どんより, 'dark, gloomy') feels heavy and dull. Such intuitions informed the pairing of these ideophones as opposites, hypothesizing that their forms might iconically reflect their contrastive meanings.

These first-person insights were not purely "subjective", in the negative sense of the term; they were continuously checked against theory and later against input from others (co-researchers and participants) thus complementing them with second-person methods, in ultimately deciding on the stimuli and the experimental design (e.g. which words to include and how to categorize

them), while the anticipated findings would in turn be used to reflect back on and refine the initial intuitions. First-person analysis provided an essential starting point for operationalizing iconicity in an experiment, ensuring that the materials and tasks felt relevant from my own experiential point of view.

3.2 Second-person methods

Second-person methods involved dialogue and collaboration with others to refine the study. I consulted with co-researchers – for instance, Kuryū Daiya and Jahnesta Sjöström – who provided valuable feedback from their perspectives. These collaborators, one of whom is a native Japanese speaker, helped verify and fine-tune the stimulus materials and interpretations. Through discussions, we ensured that the chosen ideophones and their English translations were appropriate and clear. For example, potential stimulus words and their pairings (identified via first-person intuition) were reviewed together to confirm that the intended meanings were accurate and culturally valid. The English translations of each ideophone were produced and cross-checked in this collaborative manner, with a native speaker vetting the translations for accuracy and nuance. All final word pairings and translations were vetted by a native Japanese speaker for authenticity and correctness, incorporating this second-person feedback loop into the stimulus preparation.

Additionally, a debriefing interaction with participants was built into the methodology. After the experimental tasks, participants were given an opportunity to share their experiences and strategies (e.g. via a short questionnaire or interview). In these debriefings, participants could reflect on how they made their choices in the Choosing Task or what they thought about when rating the words. This two-way communication provided qualitative insights into the participants' thought processes and confirmed that they understood the tasks. It also allowed me to empathetically consider the participants' subjective experiences when interpreting the results.

Finally, engaging with existing literature can be seen as a form of second-person method (treating the published insights of other researchers as a dialogue). I extensively read and incorporated findings from previous studies on sound symbolism and ideophones (as reviewed in 0). This ensured that the design was informed by others' observations, for example, being aware of prior debates on whether written letters' shapes influence sound symbolism judgments . By

"listening" to the scholarly community, I adjusted the methodology (e.g. including orthographic conditions, described below) to address gaps and leverage best practices. In summary, second-person input from colleagues, participants, and literature, and my supervisor helped shape a robust experimental design that is sensitive to multiple perspectives.

3.3 Third-person methods

To address the research questions from a more comprehensive standpoint, I designed and conducted a controlled experiment. This third-person method provided quantitative data on how non-Japanese speakers perceive iconic relationships between sound and meaning. The experiment was designed to incorporate features of earlier studies on iconicity (e.g. McLean, Dunn & Dingemanse 2023; Ahlner & Zlatev 2010; Giraldo 2020) and cross-modality, as discussed in Section 2.3, while extending them within a cognitive-semiotic framework.

In brief, English-speaking participants with no knowledge of Japanese were recruited to take part in an online experiment. They were asked to perform a series of tasks that operationalize the research questions into measurable outcomes. In these tasks, participants were presented with Japanese ideophones and their meanings under various controlled conditions, and their responses were recorded for analysis. The tasks were constructed to examine both accuracy (whether participants can infer meaning from sound) and their subjective ratings (how iconic the sound-meaning link is perceived to be). By systematically varying the presentation of stimuli (e.g. with or without written forms) and the type of stimuli (covering different sensory modalities), the experiment aimed to probe the role of sensory modality and orthography in iconicity perception. The following subsections detail the participants involved (3.3.1), the materials and stimuli used (3.3.2), and the experimental design and procedure (3.3.3), including task structure, conditions, and instructions.

3.3.1 Participants

A total of 36 participants (N = 36) took part in the study. They were recruited through a combination of in-person referrals, public advertisements, and online postings. All participants were adults who provided informed consent and reported at least moderate proficiency in English (the language of instructions and translations). A key inclusion criterion was no prior knowledge

of Japanese – individuals who had ever significantly studied or spoken Japanese were not eligible, in order to ensure that any success in the tasks would be due to primary iconicity rather than prior vocabulary knowledge, and secondary iconicity. Participants who initially volunteered but failed to meet the criteria (for example, those who did not consent or who disclosed Japanese language knowledge) were excluded from the final sample.

Those who qualified and completed the experiment were compensated for their time with a cinema ticket voucher (worth roughly 175 SEK; ≈15€), provided by the MA Program in Language and Linguistics at Lund University. After giving consent and meeting the criteria, participants received a web link to the experiment hosted on Gorilla Experiment Builder (https://www.gorilla.sc), an online platform for running behavioral experiments. They could complete the study remotely using their own device (a laptop computer, smartphone, or tablet) with audio enabled. Participation was typically done in a quiet setting of each person's choosing, given the online nature. The Gorilla platform handled the presentation of stimuli and recording of responses.

Participants were randomly assigned to one of three experimental groups (see 3.3.3 for details on conditions). Each group experienced a different mode of stimulus presentation, but the tasks and stimuli were otherwise identical across groups. The assignment was roughly equal, with about one-third of participants (≈12 per group) in each condition, though exact numbers per group varied slightly (Condition 1: N = 11; Condition 2: N = 13; Condition 3: N= 12). These conditions are explained in Section 3.3.3 below.

Data collection took place over a defined period (between 29 April 2025 and 17 July 2025), during which all 36 participants completed the experiment. **Table 2** summarizes the linguistic background of the participants, confirming that none had Japanese as a first or second language.

Table 2. Reported languages by role (L1 vs. L2).

Language	L1 Count	L2 Count	Total
Swedish	24	5	29
German	2	11	13
French	1	5	6
Italian	0	6	6
Spanish	0	4	4
Chinese	2	1	3
Finnish	0	3	3
Greek	2	1	3
Danish	0	2	2
Hungarian	1	1	2
Russian	2	0	2
Bulgarian	1	0	1
Czech	1	0	1
Hebrew	0	1	1
Icelandic	1	0	1
Indonesian	1	0	1
Malay	0	1	1
Portuguese	0	1	1
Total	**38**	**42**	**80**

As shown in **Table 2**, the participants were linguistically diverse: their self-reported native languages included Swedish (the most common, N = 24), Chinese, German, Russian, Greek (N = 2 each), and several others (Bulgarian, Czech, French, Hungarian, Icelandic, Indonesian; N = 1 each). Many were multilingual, but English was a second language for all and served as the common medium for instructions and responses. This diversity underscores that any systematic patterns in responses are unlikely due to a particular language background, especially since Japanese was not known to any participant.

3.3.2 Materials

The primary materials for this experiment were Japanese ideophones (sound-symbolic words) chosen to cover a range of sensory meanings and to manipulate written form usage. A total of 60 Japanese ideophones were selected as critical stimuli, with accompanying English translations for

each. These stimuli were obtained and refined through a multi-step process. An initial pool of candidate ideophones was gathered from a large Japanese language corpus. Specifically, the NINJAL-LWP for BCCWJ (NLB), an online search interface for the Balanced Corpus of Contemporary Written Japanese, was used. The NLB's dedicated "onomatopoeia" search function was employed to retrieve a list of common Japanese ideophones. The top 600 most frequent ideophonic words in the corpus were compiled into a list.

Each ideophone from the list was transcribed in hiragana and katakana (the two Japanese phonographic scripts) to examine how frequently it appears in each script. Using Sketch Engine (a web-based corpus analysis tool), the frequency of each word in hiragana vs. katakana was measured on a large web corpus of Japanese (JaTenTen11). This step produced a quantitative measure of script preference for each ideophone – some words occur predominantly in katakana, others in hiragana. These frequencies were merged with the NLB frequency data, yielding a dataset indicating not only how common each ideophone is, but also which script it is most commonly written in. From this dataset, a subset of words was chosen with the following criteria in mind:

a) words representing three sensory modalities were selected – specifically, auditory sounds, visual patterns, and internal feelings/cognitive states – corresponding to positions along an implicational hierarchy of ideophone meanings (Dingemanse 2012; McLean 2021).

b) words predominantly written in katakana, since one aim was to examine orthographic influence (katakana is often used for more salient or emphatic ideophones, see below). For each of the three modality categories (sound, visual, internal state), about 20 high-frequency katakana-preferring ideophones were shortlisted.

c) I ensured the meanings were monosemous (having a specific, primary meaning) and modality-specific.

d) Using the JMdict Japanese–Multilingual Dictionary, I checked each candidate's definitions. Any ideophone with multiple distinct meanings or meanings spanning multiple sensory modalities was excluded. For example, an ideophone like *garan* (ガラン) which can mean both "clanging (sound)" and "empty/deserted (visual scene)" was dropped due to cross-modal polysemy. Likewise, if the English translation of a word was polysemous or

ambiguous, that word was excluded to avoid confusing participants. This vetting ensured each stimulus had a clear, singular image or concept associated with it.

The final selection of stimuli was structured into pairs of contrasting ideophones. For each predominantly-katakana ideophone on the shortlist, I looked for a counterpart ideophone (often one more commonly written in hiragana) with an opposite or qualitatively contrasting meaning in the same sensory domain. As explained in Section 3.1 and 3.2, researcher intuition (first-person method) and co-researcher consultation (second-person method) were used to identify these pairs. The goal was to create pairs that differed along two dimensions simultaneously: meaning contrast (e.g., bright vs. dark, fast vs. slow, lively vs. sluggish) and script usage. For example, in the visual domain, *kirakira* (キラキラ, sparkling) – a word often written in katakana and evoking brightness – was paired with *donyori* (どんより, gloomy) – a word more often in hiragana, evoking darkness. This pair contrasts a vivid, bright visual image with a dull, dark one, and importantly, the former word happens to be conventionally written in katakana much more than the latter. In the auditory domain, similarly, an ideophone for a loud or sharp sound might be paired with one for a soft or muffled sound (with one typically in katakana and the other in hiragana). In the internal state domain (feelings/cognitive states), a positive or excited state ideophone could be paired with a negative or lethargic state ideophone, again ensuring one is katakana-dominant vs. hiragana-dominant. These pairings were finalized by manually verifying script frequency differences using the earlier corpus data and ensuring the contrasts made intuitive sense.

The outcome of this process was a set of 30 contrasting pairs of ideophones (60 individual words total). Each pair consisted of two words from the same sensory category that have opposite or strongly different meanings. Across the 30 pairs, there were 10 pairs per sensory modality (10 auditory pairs, 10 visual pairs, 10 interoceptive/internal pairs). Thus, 20 ideophones fell into each modality category. Every pair also reflected a contrast in orthographic frequency: typically, one word in the pair is predominantly written in katakana (often the one with a sharper, more salient or high-energy meaning) while the other is more often written in hiragana (often the one with a softer or more subdued meaning). All ideophones were matched with an English gloss (a concise translation) based on dictionary definitions (primarily JMdict) and refined by the research team. A

native Japanese speaker reviewed all pairs and translations to ensure that the English word chosen accurately captured the sense of the Japanese ideophone and that the two words in each pair were a fair and meaningful opposite or contrast. The final list of stimuli, therefore, comprised well-defined sound-symbolic words covering auditory, visual, and internal sensory imagery, each with a clear English equivalent.

In addition to the 60 main experimental words, a small number of ideophones were set aside for practice trials. Eight ideophones (forming 4 additional pairs) were selected for use in practice sessions prior to the main tasks, see Table 3. These were chosen to be similar in style to the main stimuli (sound-symbolic and monosemous) but were not included in the main dataset so as not to confound the results. Participants encountered these practice items during instruction phases to familiarize themselves with the task format. Responses to practice trials were not recorded for analysis. No other filler or non-ideophone items were included – the experiment focused solely on the curated ideophone stimuli.

Table 3. Practice items

Pairing	Modality	Hiragana word	Hiragana reading	Hiragana meaning	Katakana word	Katakana reading	Katakana translation
p0	auditory	そっと	[sot:o]	softly, gently	ガタン	[gataɴ]	with a bang
p1	visual	じっと	[dʑit:o]	motionless, fixedly	パッと	[pat:o]	suddenly, in a flash
p2	auditory	そろそろ	[soɾosoɾo]	slowly, quietly	ガンガン	[gaɴgaɴ]	clanging, banging
p3	visual	そっくり	[sok:uɾi]	exactly alike, spitting image	ガラリ	[gaɾaɾi]	completely, totally changed

Audio files were generated through the Google Cloud Text-to-Speech API (https://cloud.google.com/text-to-speech) with the female Japanese Wavenet voice "ja-JP-Wavenet-B" to match McLean Dunn and Dingemanse (2023: 719). These were then converted into .mp3 format and combined with an image (the equivalent katakana, hiragana, and symbols "△"or "▽") to yield three unique files for each word. All stimuli (audio files and text translations) were uploaded to the Gorilla platform in the form of .mp4-files. The platform was programmed such that each trial would call the correct file (with a built-in player for participants to hear it),

inherently displaying the relevant text or image (English meanings and Japanese script, depending on condition). The full list of tested items is found in Appendix A. Having prepared the materials in this careful way, I next designed the experimental procedure to leverage these stimuli in two complementary tasks outlined in the following section.

3.3.3 Experimental design

The main experiment consisted of two tasks performed in sequence: a Choosing Task followed by Rating Task. This dual-task structure was chosen to gather two different measures of iconicity: (a) measure of mapping accuracy (whether participants can guess meanings from sounds) and (b) subjective iconicity rating (how well participants feel a sound matches a meaning), inspired by the paradigm of McLean, Dunn and Dingemanse (2023). The combination provides a richer understanding than either measure alone, in line with previous research (McLean, Dunn & Dingemanse 2023; Iida & Akita 2023). In terms of cognitive-semiotic theory and methodology, (a) corresponds to spontaneous, pre-reflective judgements and (b) to reflective judgements. The overall design in the present thesis can be characterized as a mixed factorial design (cf. Keppel & Wickens 2004), with one between-subjects factor (stimulus presentation condition, 3 levels) and two within-subjects factors (stimulus modality, and measure type/task):
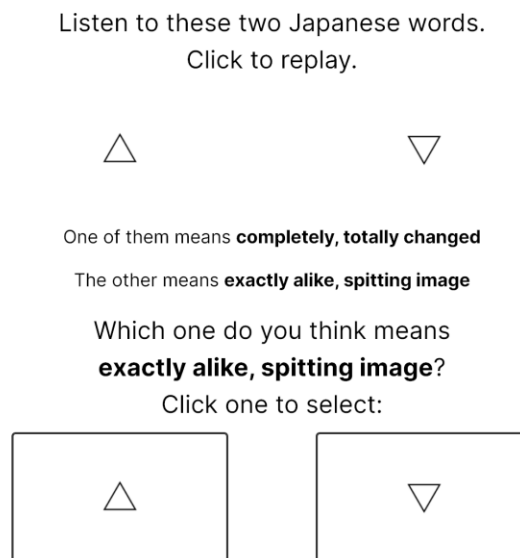
3.3.3.1 Conditions

Different groups of participants experienced different visual presentation modes for the Japanese words. Each participant was randomly assigned to one of the three conditions and remained in that condition for both tasks. The conditions are described as follows:

- **Condition 1 – Audio-Only (C1)**

Participants heard the Japanese ideophones without any written text on screen. This served as a baseline condition to measure how well sound alone can convey meaning.

Figure 4. Audio-Only Condition (C1)

Listen to these two Japanese words.
Click to replay.

△   ▽

One of them means **completely, totally changed**

The other means **exactly alike, spitting image**

Which one do you think means
exactly alike, spitting image?
Click one to select:

| △ | ▽ |

- **Condition 2 – Congruent Script (C2)**

Participants heard each ideophone with a Japanese text label in kana displayed, using the script that *conventionally or intuitively matches the word's meaning*. In practice, this meant hiragana was used to display ideophones associated with "soft", "gentle", or slower concepts, and katakana was used for ideophones conveying "sharp", "hard", loud or more forceful concepts. This mapping was based on a general observation that Japanese writers often use hiragana for softer or more emotive mimetic words and katakana for harsher or more emphatic ones (a phenomenon sometimes described in linguistics and language pedagogy), see Section 2.3.1. In this stimulus set, each word had an identified tendency: e.g., *kirakira* (sparkling, a bright/sharp visual) is typically written in katakana, whereas *donyori* (gloomy, a dull quality) is usually in hiragana; in the congruent condition these were shown in those respective scripts. The Congruent Script condition tests whether seeing the "expected" orthography reinforces participants' iconic impressions of the words.

Figure 5. Congruent Script Condition (C2)

Listen to these two Japanese words.
Click to replay.

パッと   じっと

One of them means **motionless, fixedly**

The other means **suddenly, in a flash**

Which one do you think means
suddenly, in a flash?
Click one to select:

パッと   じっと

- **Condition 3 – Incongruent Script (C3)**

Participants heard each ideophone with a Japanese text label in the *opposite* kana script than would be intuitively expected by its meaning. That is, the hiragana/katakana assignment for each word was swapped relative to Condition 2. Words denoting softer, gentler concepts were presented in katakana, and words for sharper or more intense concepts were presented in hiragana. Using the previous example, *kirakira* would be shown in hiragana (キラキラ) and *donyori* in katakana (ドンヨリ) in this condition. The Incongruent Script condition provides a contrast to test whether a mismatched orthographic form can *hinder* or alter the perception of iconicity. If participants perform worse here than in the congruent condition, it suggests that the visual form of the word (its letters) has an influence on their intuition of the meaning. All three groups heard identical audio stimuli; only the presence/absence and type of written feedback differed. By comparing these groups, I could isolate the effect of orthography (no text vs. congruent text vs. incongruent text) on performance and judgments.

Figure 6. Incongruent Script Condition (C3)

Listen to these two Japanese words.
Click to replay.

ぱっと   ジット

One of them means **motionless, fixedly**

The other means **suddenly, in a flash**

Which one do you think means
suddenly, in a flash?
Click one to select:

| ぱっと | ジット |

3.3.3.2 Within-subjects factors: Stimulus Category (Modality)

Each participant encountered stimuli from three sensory domains: Auditory, Visual, and Interoceptive (inner feelings/states). As described in Section 3.3.3 Materials, there were 10 trials (word pairs) per category, making 30 trials total per task. The category of each trial served as a repeated measure factor; for example, I could later compare how the same participant performed on auditory vs. visual items, etc.

3.3.3.3. Within-subjects factors: Task (Measure Type)

Although not a traditional factorial "factor" in the statistical sense (since the two tasks yield different dependent variables), the design involves each participant doing two kinds of tasks one after the other: a choosing task and a rating task. The tasks function as complementary parts of the design but every participant contributes to both measures.

- **Choosing Task**

In the first part of the experiment, participants completed a two-alternative forced-choice task modeled after the "meaning-to-word" paradigm used by e.g. McLean et al. (2023). Each trial of the Choosing Task proceeded as follows: the participant was presented with a target English gloss and had to choose which of two Japanese ideophones corresponded to that meaning. Concretely, on a given trial the screen would display an English word or short phrase (e.g., "sparkling" or "bright and glittery") as the prompt, and the participant would hear two Japanese audio clips in succession (or have the ability to play two audio buttons labeled, "▽" and "△",or the words in Japanese). These two audio clips were the recordings of a contrasting pair of ideophones from the stimulus set. The participant's task was to decide which of the two Japanese words means the given English prompt. They would indicate their choice by clicking on the corresponding button (or text, if visible).

No feedback was given during this task about whether the choice was correct or not. Participants had to rely on their intuition (of primary iconicity) to make a guess. The trial order was fully randomized for each participant, so the 30 trials (covering all 30 pairs, each pair appearing once) came in an unpredictable sequence mixing the modalities. Each of the 60 ideophones was thus tested exactly once per participant – across 30 trials, each pair appeared and each word from that pair was a correct answer for those who chose it (or incorrect for those who didn't). This one-pair-per-trial structure ensured that each of the 60 words was presented exactly once in the Choosing Task (spread across participants' trials). This means that each participant effectively attempted to "guess" 30 meanings (one for each pair), covering the whole stimulus set without repetition. The order of the two options (which side each word was on, or which played first/second) was counterbalanced or randomized to avoid systematic position biases.

Participants were instructed at the start of this task that the words they would hear are from a foreign language (Japanese) and that each trial has one correct answer (the actual translation) and one decoy. They were encouraged to make their best guess based on how the words sound, even if they had no knowledge of Japanese. The instructions emphasized that sometimes a word's sound might "suggest its meaning". Participants were also informed that they could replay the sounds if needed (e.g., Gorilla allows clicking the audio button again). Multiple playbacks were

allowed, but they should follow their first impression. There was no strict time limit per trial, though participants generally took only a few seconds to respond once they heard the options. After making a choice, they proceeded to the next trial automatically.

Prior to the real trials, participants completed a couple of practice trials in this format using the practice items, so they understood the task. These practice trials mimicked the procedure (an English prompt, two Japanese audio options) and gave participants a chance to familiarize themselves with clicking and listening. Again, no feedback was given on practice trials beyond an explanation afterward of the correct answers, to illustrate the idea (e.g., "「○○」 actually means X, so that was the correct choice."). Practice trial answers were not revealed.

The Choosing Task thus yielded, for each participant, a set of 30 responses: each either correct or incorrect. From this I could calculate an accuracy score (out of 30) for each participant and further break it down by modality (out of 10 per category) and by condition group. Because guessing randomly would produce about 50% correct on average (15/30), a group-level accuracy significantly above chance would indicate that participants can intuit the meaning from the sound more often than not – possible evidence of primary iconicity at work.

- **Rating Task**

After completing all 30 trials of the Choosing Task, participants immediately moved on to the second part, the Rating Task. In this task, participants were asked to rate how well each Japanese word's sound matched its meaning, on a Likert scale. Each trial of the Rating Task presented one ideophone at a time (as opposed to pairs) along with its English translation, and the participant had to judge the degree of sound-meaning resemblance. Specifically, the screen would show an English word/phrase (the meaning) and play a Japanese ideophone (audio of the word); in conditions with text, the Japanese word was also displayed in the assigned script on screen. The participant then indicated their judgment on a 7-point scale, where 1 meant "no resemblance between the word's sound and its meaning" and 7 meant "very high resemblance" (with the mid-point 4 being neutral or "moderate resemblance"). Participants were instructed to use the full range of the scale and that there were no right or wrong answers in this part – it was about their subjective impression of each word. They were told to consider "How much does this word sound like what it means?" and then click the number that best fits their impression. As with the first task, four practice trials (with the

same practice ideophones) were given so participants could experience how to use the rating scale before the real task.

The Rating Task included one half of each ideophones pair in the previous task so that every pairing also got an iconicity rating. The trials were again randomized in order for each participant. Participants saw the English meaning and could listen to the Japanese audio as needed before responding. They experienced the same presentation conditions as before: e.g., if a participant was in the Congruent Script group, they saw each word in its congruent kana while rating, whereas an Audio-Only participant saw no Japanese text, only hearing it. This consistency was maintained so that any influence of seeing or not seeing the word would also affect the ratings.

The purpose of the Rating Task was to capture a subjective iconicity judgment for each word, complementing the objective measure from the Choosing Task. For instance, even if a participant guessed a particular word's meaning wrong, they might still report that "it sounded like it should mean something bright" (if they gave a high rating for a word that indeed means something bright). Conversely, a correct guess with a low resemblance rating might indicate the guess was more of a lucky guess or that the word did not sound iconic even if they got it right. By having both measures, I could cross-check findings: do words that are guessed correctly at above-chance rates also tend to have high iconicity ratings? Does one condition make people feel the words are more iconic (ratings) as well as improve their guessing accuracy, or do those diverge? These questions could be answered by correlating the two data sets.

3.3.3.4. Experimental procedure

The two tasks were implemented back-to-back in a single session. There was a brief instruction screen before each task and a short break between them if needed. I deliberately placed the Choosing Task first and the Rating Task second. This ordering ensured that participants approached the guessing task naively, without yet being exposed to all the actual meanings of the words. Only after attempting to guess through all pairs did participants learn the meanings (during the rating task, each trial explicitly shows the meaning). Had the order been reversed, the rating task would effectively teach participants the answers, which could invalidate the guessing task. Thus, the structure preserved the integrity of the first task's measure of naive performance.

Before each main task started, an audio check was performed by Gorilla to confirm the participant could hear sounds properly (e.g., playing a sample sound and asking if they heard it). Participants had to indicate that the sound was audible before proceeding, to avoid data loss due to technical issues. The experiment was self-paced; on average, participants took about 20 minutes on average to complete both tasks. Finally, upon finishing the Rating Task, participants answered any debrief questions and were shown a thank-you message.

3.3.3.5. Summary

The experimental design combined a forced-choice "Choosing" task (30 trials, across 3 modalities) and an iconicity rating task (each word individually rated), under three different presentation conditions (between subjects). This design allowed me to examine multiple facets of the research questions. The contrastive pairs and 2AFC paradigm enable testing whether participants can infer meaning from sound at above-chance levels, which would indicate primary iconicity effects. The division into sensory categories (auditory, visual, interoceptive) allowed me to observe whether iconicity is more evident in some domains than others (e.g., sound-meaning mappings might be easier than feelings-meaning mappings), speaking to the proposed implicational hierarchy (1) (Dingemanse 2012; McLean 2021) that auditory iconicity is more universally accessible than, say, internal state iconicity. The inclusion of orthographic conditions (audio-only vs. congruent vs. incongruent script) addresses the role of iconicity through writing – essentially testing if "shape symbolism" or written form bias influences iconicity perception. If the Congruent script group (C2) outperforms the others, it would suggest that seeing the expected visual form of the word reinforces the iconic link. If the Incongruent group underperforms (or rates words as less fitting), that would suggest a conflict between sound and visual form can reduce the sense of iconicity. By comparing the two tasks, I could also determine if pre-reflective judgments of iconicity aligns with reflective assessments. This layered approach provides a comprehensive basis from which to answer the research questions.

3.4 Summary and hypotheses

In this chapter, I detailed how the study was carried out: participants with no knowledge of Japanese completed two tasks involving Japanese ideophones, under different conditions of

stimulus presentation. The methodology was designed to triangulate the concept of iconicity from different angles – through measures based on pre-reflective judgments and introspective ratings – and to isolate the factors of sensory modality and writing system influence. Based on the literature and the theoretical framework (Chapter 2) the following hypotheses were formulated prior to data collection, guiding expectations.

H1: Above-Chance Guessing

Non-Japanese speakers will be able to recognize the meanings of ideophones at a rate significantly above chance. In other words, in the Choosing Task participants will select the correct meaning more often than the 50% expected by random guessing. This hypothesis (H1) reflects the expectation of primary iconicity – that the sound of an ideophone carries clues to its meaning even without prior knowledge of the meaning of the ideophone. Success in the Choosing Task would mirror classic findings like the bouba-kiki effect, generalized to natural language ideophones.

H2: Modality Effect

Iconicity effects will differ by sensory modality of the ideophone. Specifically, I predicted that auditory ideophones would yield the highest performance since the iconicity is in this case unimodal and most transparent, followed by visual ideophones, and that interoceptive (internal state) ideophones would yield the lowest performance in terms of both Choosing Task accuracy and Rating task. This hypothesis (H2) is derived from the implicational hierarchy, which suggests that sound-based meanings are most universally accessible (most iconic), while internal feelings are more abstract and might rely more on convention, and thus on secondary iconicity. Thus, I expected participants to guess sound-related words (e.g., onomatopoeic ones) more easily than purely visual or internal state words. Similarly, when rating, I expected them to perceive stronger sound-meaning resemblance for auditory ideophones, and the weakest for words describing internal sensations.

H3: Orthography Influence

The general expectation was that the presence and congruence of written kana would modulate participants' performance and judgments. I hypothesized that Condition 2 (Congruent Script) would produce the highest Choosing Task accuracy and the highest iconicity ratings on average.

Condition 1 (Audio-Only) was expected to be slightly lower, representing baseline performance with no orthographic influence. Condition 3 (Incongruent Script) was expected to produce the lowest accuracy and ratings. In short, seeing a matching script (hiragana or katakana in line with the word's "feel") should enhance iconicity, while seeing a mismatching script may confuse or dampen iconic cues. For example, if a word that sounds "sharp" is also seen in the spikier-looking katakana script, participants might be even more confident in its meaning and rate it as highly iconic; if that same word is shown in the softer-looking hiragana, it might create a slight dissonance and lower performance. This set of predictions (H3) aligns with the idea of iconicity in written form and is informed by prior hints in the literature that orthography can influence sound symbolism judgments.

With these hypotheses in mind, the study was conducted, and data were collected. In the next chapter, I present the results of the Choosing Task and Rating Task, analyzing them with respect to the sensory modalities and orthographic conditions. I then discuss how the findings bear on the research questions and the hypotheses outlined above.

Chapter 4 Results and Discussion

This chapter presents the empirical findings from the experimental study detailed in Chapter 3. The primary goal of this investigation was to explore the nature of both unimodal and cross-modal iconicity in Japanese ideophones among non-speakers of Japanese, focusing on the interplay between primary and secondary iconicity, the role of different sensory modalities, and the influence of orthography. The results are organized into four main sections. In Section 4.1, the outcomes of the two-alternative forced-choice Choosing Task are presented, providing a measure of participants' pre-reflective judgements of the meanings of ideophones. This section analyzes overall accuracy and performance across the independent variables: sensory modality (auditory, visual, interoceptive) and experimental condition (Audio-Only, Congruent Script, Incongruent Script). In Section 4.2, the findings from the Rating Task are presented. This provides a measure of the reflective judgments on how iconic participants perceived the sound–meaning relationship of each ideophone to be. These results are examined across modalities and conditions to assess participants' impressions of iconicity. In Section 4.3, a comparative analysis examines the relationship between the accuracy scores from the Choosing Task and the reflective judgments from the Rating Task. In Section 4.4, these quantitative results, supplemented by qualitative observations from the post-experiment questionnaires, are then discussed in relation to the three hypotheses (H1-H3) established at the end of Chapter 3, setting the stage for the final conclusions in Chapter 5.

4.1 Choosing Task results

Participants' overall success on the guessing task was above chance. Across all conditions the mean number of correct guesses per participant was 19.25 out of 30 (SD = 3.20, range 14–27). This corresponds to a mean accuracy rate of 64%, well above the 50% chance level. This finding supports the claim that the form of Japanese ideophones provides non-Japanese speakers with cues sufficient to guide interpretation. Further, as shown in **Figure 7**, performance varied systematically by modality. Unimodal auditory ideophones were guessed most accurately (M = 6.86 out of 10,

SD = 1.26), followed by visual ideophones (M = 6.42, SD = 1.10), with interoceptive ideophones guessed least accurately (M = 5.97, SD = 1.17). This suggests that auditory expressions, which signify sounds directly, and thus unimodally, were most transparent to participants.

However, accuracy varied little by orthographic condition (**Figure 8**); Participants in the Audio-Only, Congruent, and Incongruent Script conditions performed at similar levels, suggesting that orthography did not strongly influence guessing accuracy at the group level.
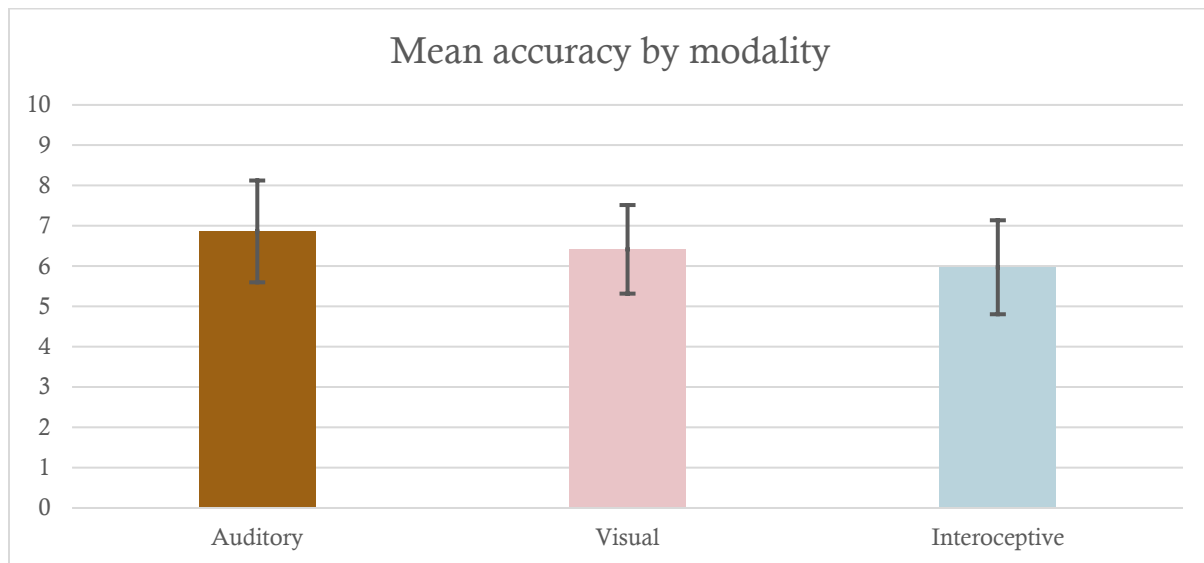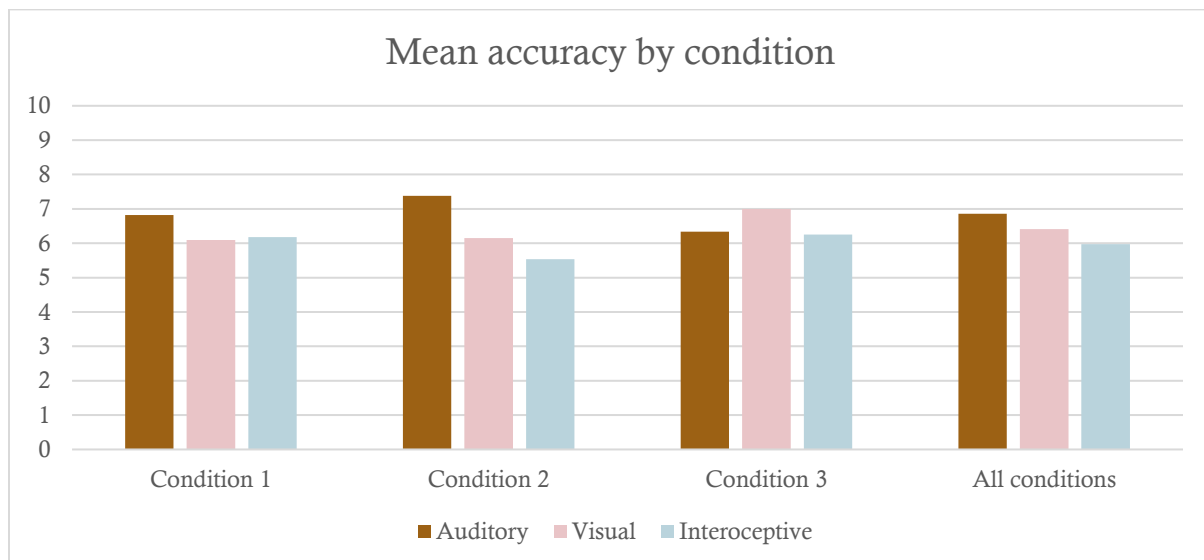
Figure 7. Mean accuracy by modality

Figure 8. Mean accuracy by condition

4.1.1 Mean accuracy by participant and ideophone

Individual differences were marked (**Figure 9**). The best-performing participants achieved scores close to 27 out of 30, corresponding to nearly 90% accuracy, whereas the lowest-performing scored between 14 and 16 ($\approx$47–53%). This wide spread indicates that some individuals were far more attuned to sound–meaning clues than others. Notably, participants' success did not cluster by condition: high and low performers were found in all three groups (Audio-Only, Congruent, Incongruent). This reinforces the conclusion that orthographic presentation had little systematic effect: an attentive participant could do well even without seeing any script, and conversely a less attuned participant might struggle even with a "congruent" script available.

Performance also varied greatly across ideophones themselves (**Figure 10**). Some ideophones were consistently identified correctly, particularly those with strong onomatopoetic qualities (e.g. auditory words mimicking other sounds). Others, especially interoceptive ideophones referring to internal states, elicited near-chance performance, demonstrating that certain meanings are less transparent to naïve listeners. In fact, accuracy ranged from almost 94% correct for the easiest item down to 36% for the hardest. The most guessable ideophones tended to be those with very transparent sound–meaning relationships, often classic onomatopoeic sounds or ideophones with clear sensory imagery. For example, several auditory ideophones such as *sakutto* (a9) had accuracy rates above 80%, indicating that even without knowing Japanese, participants could infer their meanings from the sound alone. By contrast, the items that elicited the poorest performance were largely interoceptive ideophones items like *dokidoki* (i9), were not consistently identified (despite high iconicity ratings, see Section 4.2). These yielded accuracy around chance (50% or below).
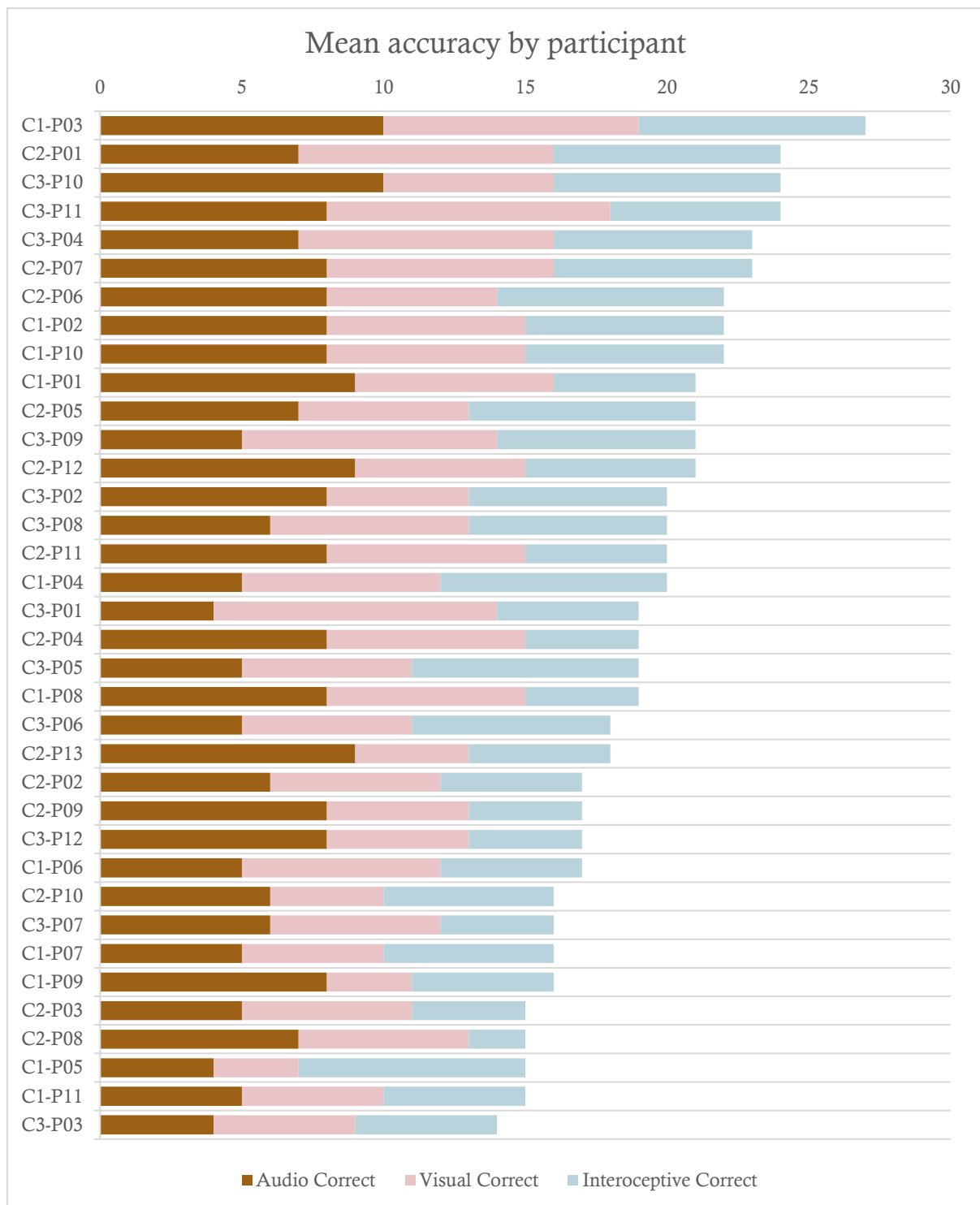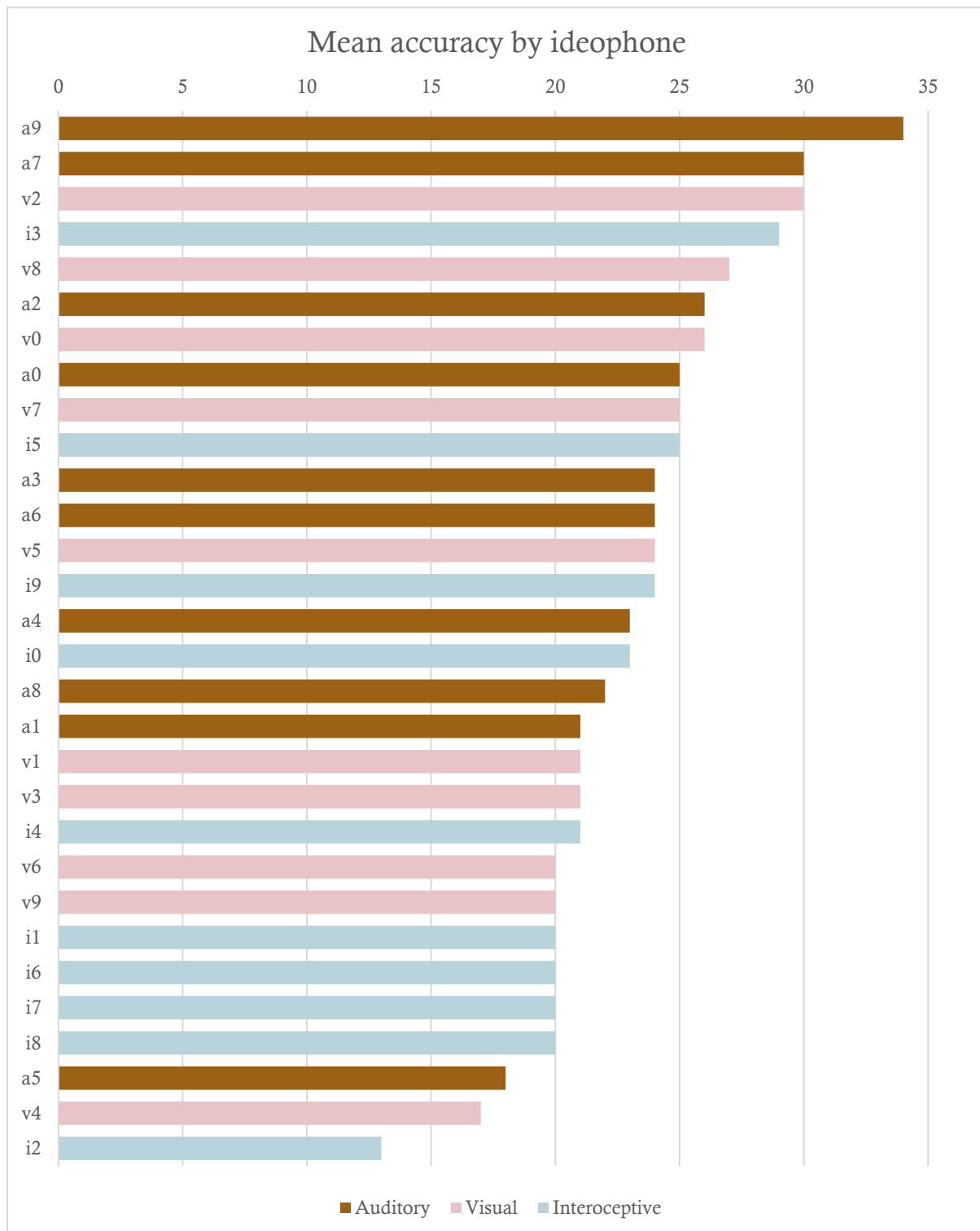
Figure 9. Mean accuracy by participant

Mean accuracy by participant

Audio Correct ■ Visual Correct ■ Interoceptive Correct

Figure 10. Mean accuracy by ideophone


Mean accuracy by ideophone

4.2 Rating Task results

The Rating Task provides a complementary perspective. Across all participants and stimuli, the grand mean iconicity rating was 4.26 on a 7-point scale, suggesting that participants perceived ideophones to be moderately iconic on average. Unlike the Choosing Task, ratings varied across conditions. Participants in the Audio-Only condition provided the highest ratings ($M \approx 4.50$), while those in the Congruent and Incongruent Script conditions produced lower ratings ($M \approx 4.15$). This pattern may reflect a relative advantage for auditory presentation when evaluating iconicity directly, perhaps because visual script provided either no additional cues or distracting, incongruent cues (see **Figure 11**).

Thus, merely seeing the ideophones written (in either kana script) tended to reduce participants' judgments of iconicity relative to hearing alone. This outcome is intriguing: one might have expected the Congruent script to enhance the iconic impression, but instead the presence of any written script coincided with a modest drop in perceived iconicity. One explanation is that the audio-only presentations provided a purer iconic experience, allowing participants to focus solely on sound–meaning resemblances, whereas adding written forms introduced noise or cognitive load. In the Congruent condition, the script was "matched" to the ideophone's typical form, yet it may not have added any meaningful cues for non-readers of Japanese, functioning essentially as a distraction. In the Incongruent condition, the mismatch between sound and an unexpected script could have created a subtle dissonance. The net effect in both cases was that participants with audio-only stimuli felt slightly more confident about iconicity. that for participants with no knowledge of kana, seeing the characters did not boost iconicity, and may even have dampened it.

Figure 11. Mean rating by condition
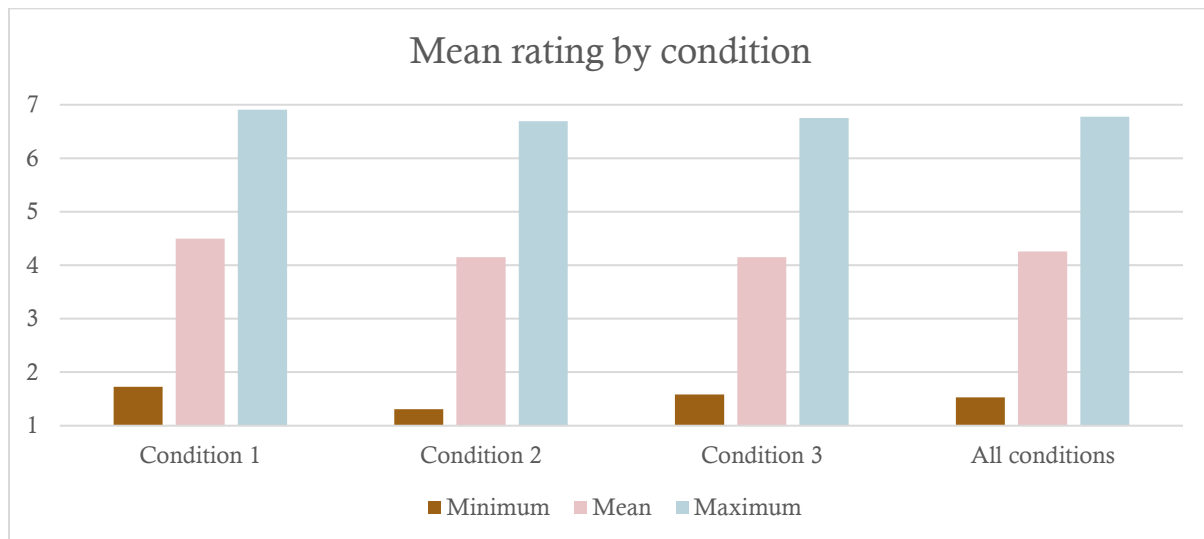
Ratings also differed by modality (**Error! Reference source not found.**). In contrast to the Choosing task, however, interoceptive ideophones received the highest ratings (M = 4.45, SD = 0.78), followed by visual (M = 4.24, SD = 0.28) and auditory ideophones (M = 4.08, SD = 0.60). Individual variation in rating behavior was observed (**Figure 13Error! Reference source not found.**). Some participants used the full 1–7 range, while others clustered ratings narrowly around the mid-scale, suggesting differences in interpretive strategy. Nonetheless, the modality pattern (interoceptive > visual > auditory) held consistently across individuals. Also, there was considerable variation in the ratings of individual ideophones (**Figure 14**).
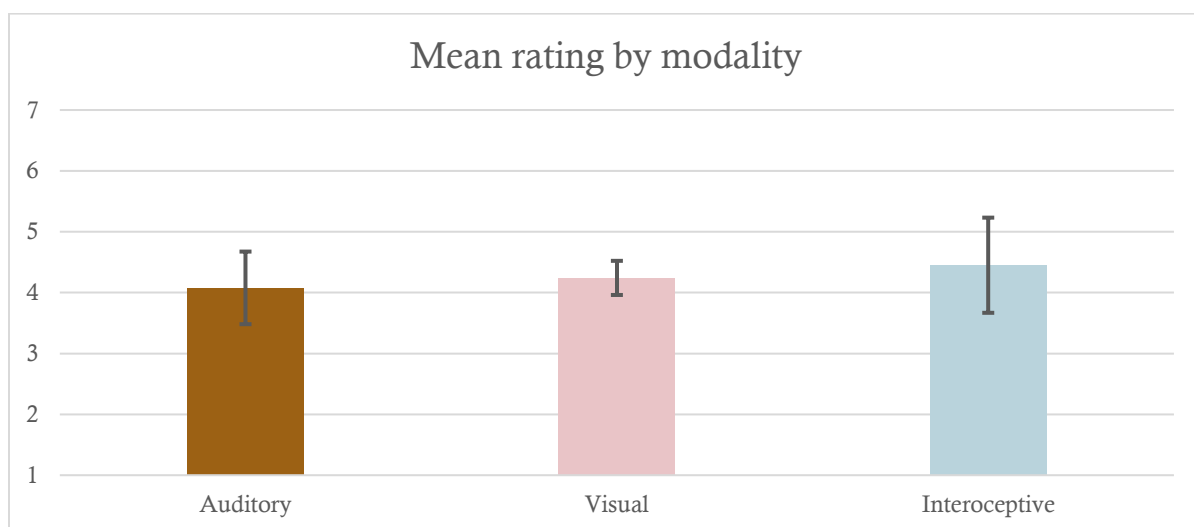
Figure 12. Mean rating by modality

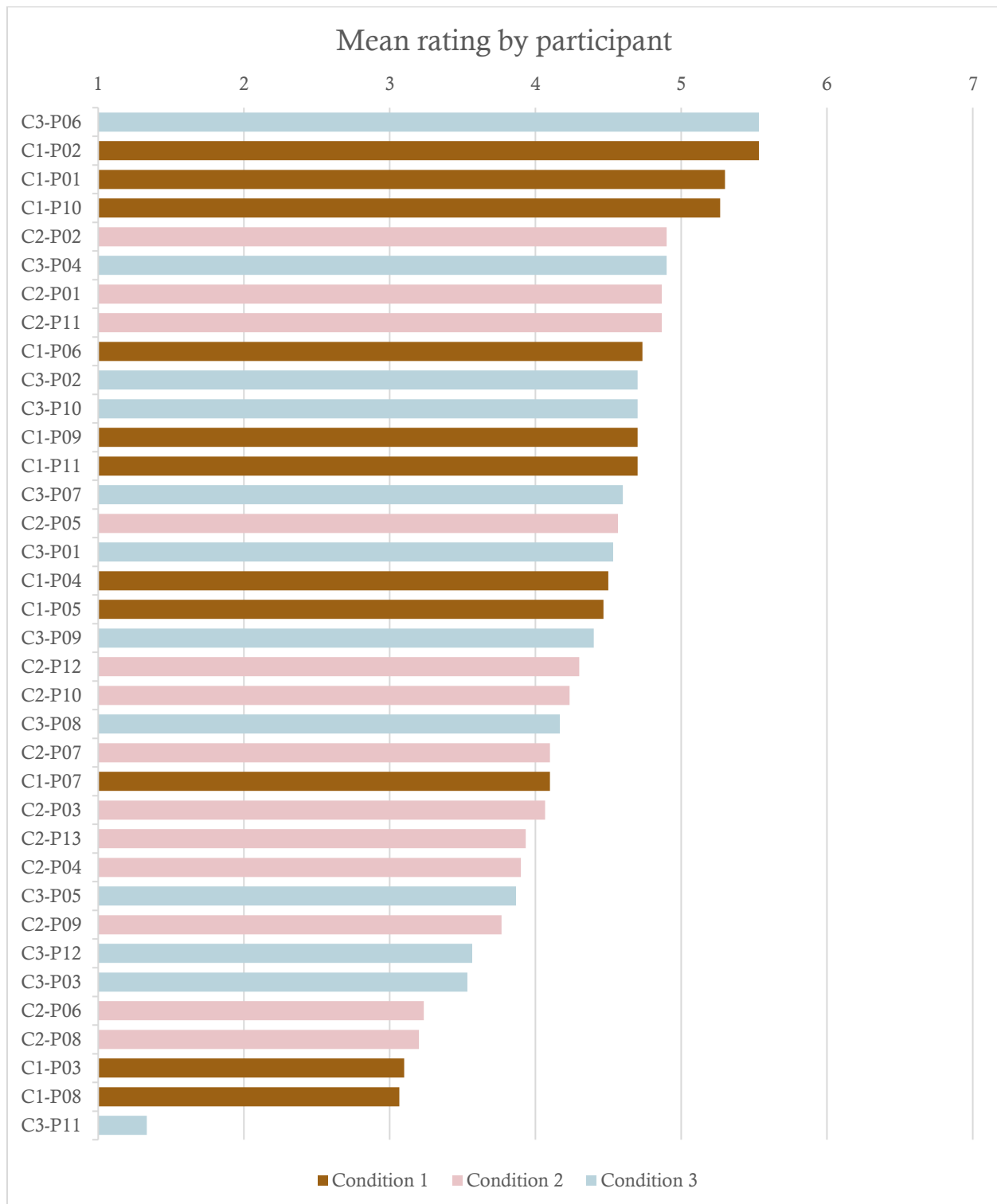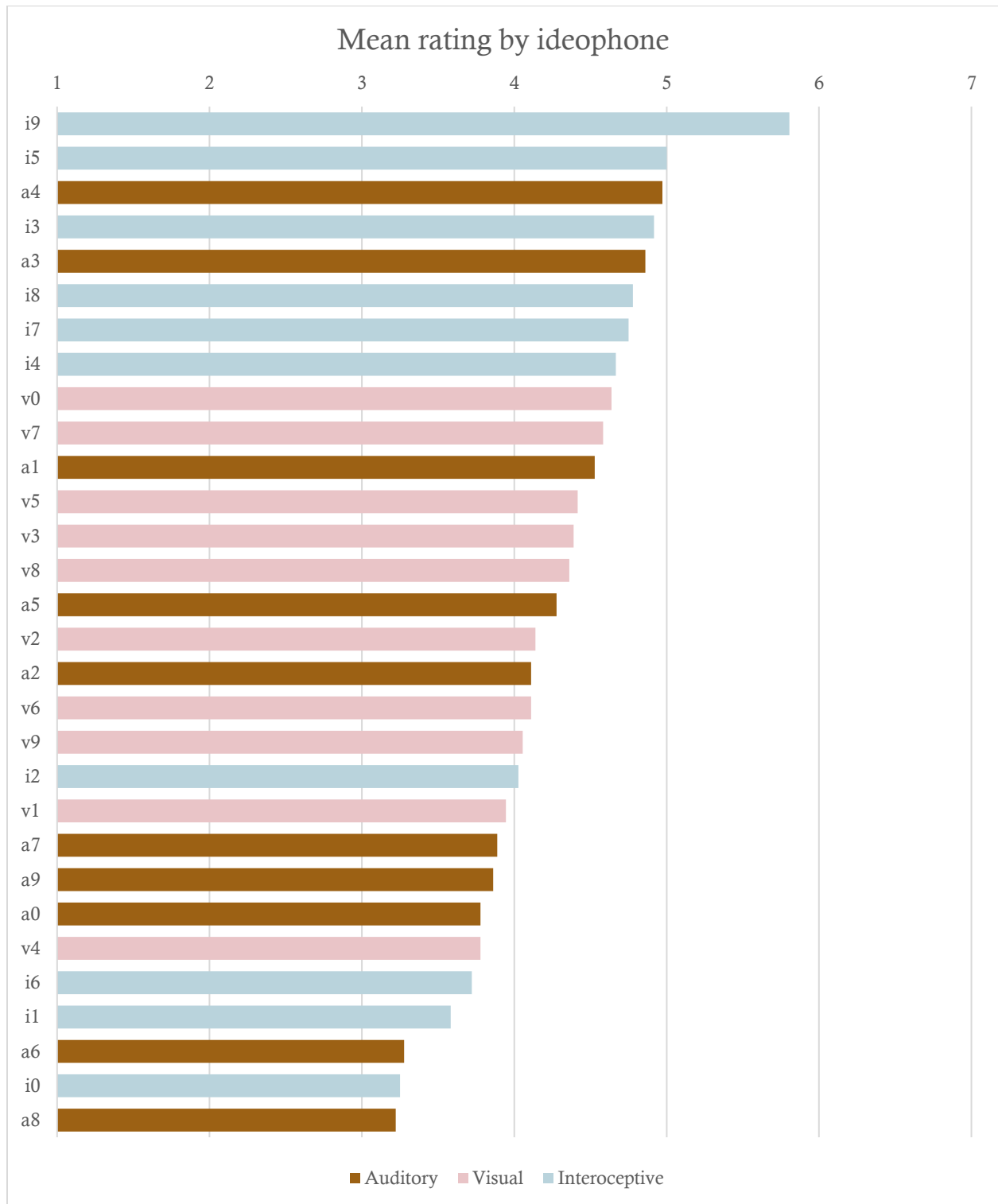**Figure 13.** Mean rating by participant

Figure 14. Mean rating by ideophone

4.3 Cross-task comparison

Comparing Choosing Task accuracy and Rating Task iconicity scores reveals both convergence and divergence (see **Figure 15–Figure 16**). Some ideophones, such as *sakutto* (a9), scored high in both accuracy and ratings. Others, however, diverged sharply. The most striking example is *dokidoki* (i9), which received the highest iconicity rating but only modest accuracy in the Choosing Task. These findings indicate that (pre-reflective) Choosing Task performance and reflective iconicity judgments, while to some degree correlated, are not equivalent. As shown in **Figure 7** and **Figure 12**, the orders are reversed, with unimodal auditory ideophones being judged correctly most easily in the Choosing Task, while receiving the lowest score in the Rating Task. One interpretation is that high accuracy scores diverging from iconicity ratings seem to reflect an immediately perceptible resonance between expression and content, whereas reflective ratings capture an additional dimension of vividness and expressivity.

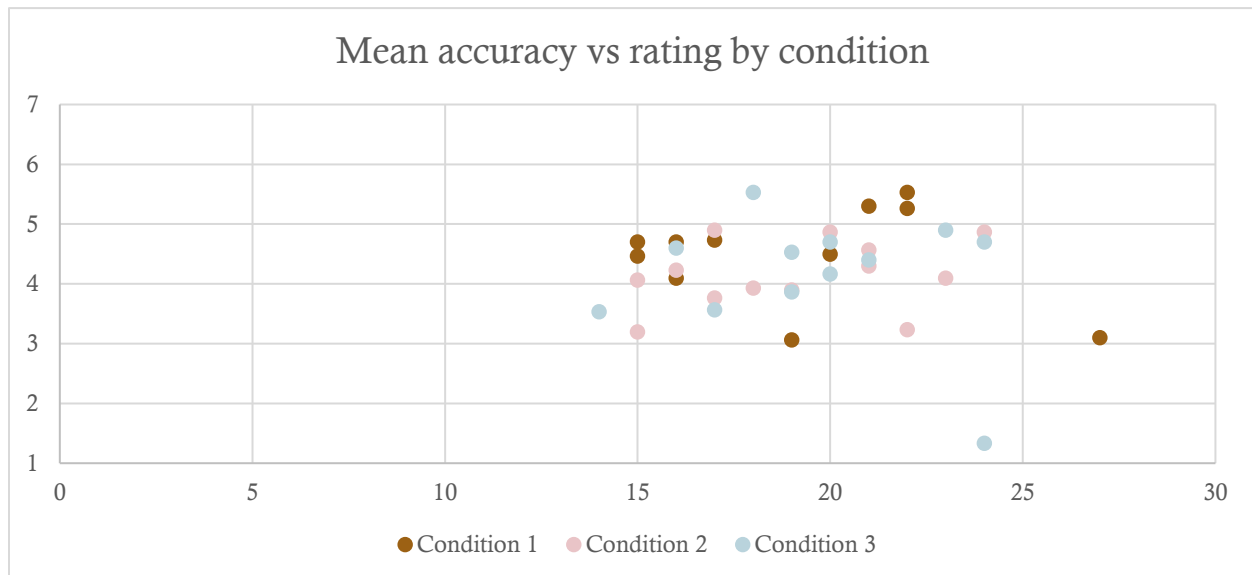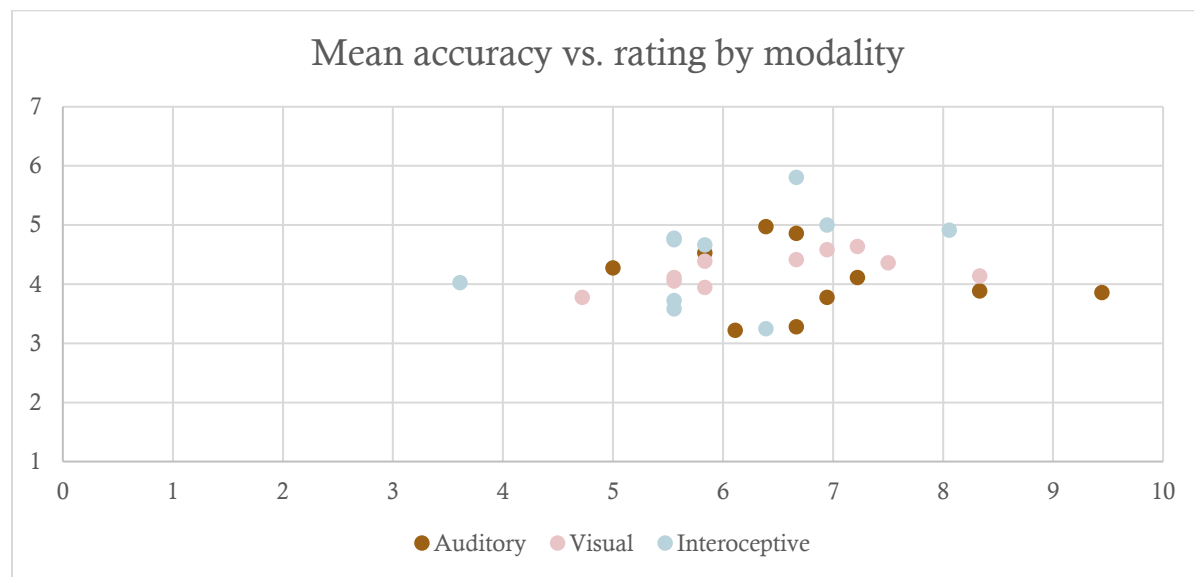**Figure 15.** Mean accuracy vs. rating by condition

Figure 16. Mean accuracy vs. rating by modality

4.4 Hypothesis Evaluation

H1 (Above-Chance Guessing): Non-Japanese participants will guess ideophone meanings at a rate significantly above chance (50%). This hypothesis was supported. As reported in Section 4.1, the mean accuracy in the Choosing Task was ~64%, well above chance, and all 36 participants exceeded the 50% benchmark (even the lowest score was 14 out of 30 ($\approx$ 47%), essentially at chance level, while most were much higher). **Figure 7** illustrates that in each modality, the average correct out of 10 was clearly above 5 (chance). A one-sample comparison would confirm that 19.25/30 correct is statistically greater than 15/30 (chance), validating H1. This finding corroborates previous research and suggests a large degree of iconicity (arguably both primary and secondary) in ideophones: even without knowing Japanese, Choosing Task accuracy was significantly higher than the 50% chance level, demonstrating sensitivity to iconic expressions. This reinforces the claim that ideophones are not experienced as arbitrary by naïve listeners.

H2 (Modality Effect on Guessing): Iconicity effects will differ by sensory modality; specifically, auditory ideophones will yield the highest Choosing Task accuracy, followed by visual, and interoceptive ideophones the lowest. This hypothesis was confirmed by the Choosing Task results. As detailed earlier, participants' accuracy means did follow Auditory > Visual > Interoceptive

(6.86, 6.42, 5.97 out of 10, respectively; see **Figure 7**). This ordering is consistent with the prediction derived from the implicational hierarchy of ideophone iconicity, where sound-based meanings are hypothesized to be the most universally accessible, and internal state meanings the least. The data supports the notion that unimodal iconicity (sound to sound mappings) is more readily perceived than cross-modal iconicity involving internal sensations. It's worth noting that while the pattern matches H2, the magnitude of differences was moderate. Participants still got a non-trivial number of visual and interoceptive items correct, just fewer than auditory. Nonetheless, H2 holds: the modality of the ideophone significantly influenced guessing success. This outcome dovetails with patterns in other studies. This indicates that participants perceived them as highly expressive, even without being able to identify their meanings. Further, the dissociation between the Choosing and Rating tasks aligns with McLean, Dunn and Dingemanse (2023), who found that explicit iconicity ratings and "guessing" accuracy (i.e. pre-reflective judgements) often diverge, apparently since in the case of explicit rating, participants are affected by as yet undefined factors, such as expressivity.

H3 (Orthography Effect on Guessing): Seeing an ideophone's usual script will enhance iconicity performance, while seeing it in an incongruent script will diminish performance (Congruent script > Audio-only > Incongruent in Choosing Task accuracy). This hypothesis was not supported. As seen in Section 4.1, there was no significant advantage or disadvantage conferred by orthographic condition on guessing accuracy (**Figure 8**). Participants in the Congruent script condition did no better than those in the Audio-only condition. In fact, their mean total scores were virtually identical (~19/30 each). Likewise, the Incongruent script group did not underperform the others; their mean was slightly, but not meaningfully, higher, and any minor differences were attributable to chance. Thus, contrary to H3, providing a "matched" kana script did not boost participants' ability to guess meanings, nor did a mismatched script impair it. This suggests that orthographically mediated iconicity, if it exists for these stimuli, was too subtle for non-Japanese participants to leverage. The assumption that hiragana's roundness versus katakana's angularity might modulate perceived meaning was not borne out in behavior, at least not in this guessing measure. This finding contrasts with some prior evidence from native Japanese readers (e.g.

Iwahara, Hatta & Maehara 2003), who found that Japanese speakers associated certain scripts with certain imagery. In this case, for participants who could not actually read the scripts, those visual forms did not translate into better guessing. The lack of an orthography effect in guessing aligns with the idea that the iconicity relied on in this experiment was primarily based on the auditory modality, whereas any cues from writing require cultural familiarity and being more accustomed to the specific semiotic system to be effective. H3's predicted ranking (C2 > C1 > C3) did not appear; orthography had negligible impact on Choosing Task performance.

In sum, the experimental study investigated unimodal and cross-modal iconicity in Japanese ideophones through two complementary behavioral tasks and post-experimental reflections. Three central findings can be highlighted. First, participants performed above chance in the Choosing Task, confirming that even non-speakers are able to exploit iconic form–meaning correspondences. This supports Hypothesis 1 and reinforces the notion that primary iconicity, particularly salient in auditory ideophones, provides a direct motivational link between sound and meaning. Second, modality significantly shaped performance and perception. Participants "guessed" auditory ideophones most accurately, visual ideophones somewhat less, and interoceptive ideophones least in the Choosing Task. In contrast, reflective iconicity ratings revealed the opposite trend: interoceptive ideophones were perceived as most iconic, while auditory items were rated lowest. This divergence between behavioral accuracy and reflective evaluation mirrors McLean, Dunn and Dingemanse's (2023) demonstration that guessing and rating tasks capture different but complementary dimensions of iconicity. Third, orthographic presentation produced minimal effects at the group level, though some individual participants reported being influenced by script shape. This suggests that Japanese kana orthography exerts only a weak role in cross-modal iconic judgments among non-speakers, though its potential interaction with cultural familiarity should not be dismissed.

Chapter 5 Conclusions

In this thesis, drawing on concepts from cognitive semiotics and phenomenology, I have examined cross-modal iconicity in Japanese ideophones through guessing accuracy and iconicity measures. Following the principle of the conceptual-empirical loop, the aim has been to integrate experimental findings with theoretical perspectives on iconicity, within a general cognitive semiotic approach. In this final chapter, I summarize the contributions of the study by answering the research question (5.1), addressing briefly some limitations of the study (5.2) and finally returning to the conceptual side of the conceptual-empirical loop (see Section 2.1), addressing how the study provided new insights into the pivotal concepts (5.3).

5.1 Answering the research questions

5.1.1 RQ1: Primary and Secondary Iconicity in Ideophones

The first research question (RQ1) asked about the relationship between primary and secondary iconicity in the perception of ideophones. The results demonstrate that since participants who do not speak Japanese could "guess" the meanings of Japanese ideophones above chance level, their iconicity is not only secondary, but to a considerable degree also primary, consistent with the conclusion of Ahlner and Zlatev (2010) concerning fictive words. The Choosing Task performance is arguably the best measure of this, as it reflects pre-reflective judgements, which are closer to the more basic levels of intentionality (see Section 2.2.1).

Rating Task judgments, by contrast, involve introspection to considerable extent, and are thus less reliable, as they appear to be sensitive to factor like vividness and expressivity, rather than iconicity (i.e. resemblance between expression and content) per se. The discrepancy between tasks was a central finding of the study. The Choosing Task primarily reflected primary iconicity, since successful performance required participants to map sound directly onto meaning. Here, auditory ideophones, with transparent imitative qualities, yielded the highest accuracy, confirming that primary iconicity is strongest when perceptual resemblance is direct and unimodal (Ramachandran & Hubbard 2001; Dingemanse et al. 2015). The Rating Task, by contrast, revealed

that starting from secondary iconicity – asking participants to explicitly judge the resemblance between a given expression and what it signifies – provided conflicting results. That is, participants often judged ideophones as highly iconic even when they failed to guess their meanings correctly, particularly with interoceptive items. This echoes McLean, Dunn and Dingemanse's (2023) demonstration that iconicity ratings can remain high regardless of guessing accuracy.

5.1.2 RQ2 Which sensory modalities are perceived as more iconic?

In the Choosing Task, unimodal auditory ideophones were identified most accurately, followed by cross-modal visual and interoceptive ones. This confirms that ideophones imitating external sounds provide the most transparent cues for non-speakers, consistent with cross-linguistic findings on the robustness of unimodal iconicity, when the mapping is from auditory expressions to auditory meanings (Ćwiek et al. 2021). In contrast, interoceptive ideophones were the hardest to guess correctly, reflecting the relative opacity of internal bodily states to iconic representation. As pointed out above, the Rating Task reversed this hierarchy. Participants judged interoceptive ideophones as most iconic, followed by visual and auditory ones. This suggests that while interoceptive ideophones lack transparent mappings that facilitate guessing, their structural features (e.g., repetition, vowel quality, prosody) nevertheless evoke a strong introspective "sense of iconicity", even if this was in a sense illusory. This aligns with Iida and Akita's (2023) perceptual-strength norms, which indicate that interoceptive concepts are often rated as abstract and difficult to verbalize.

Second-person reports shed further light on these patterns. Many participants described basing their decisions on "feelings" or "tonality", particularly for words that did not transparently resemble a sound. For example, one participant explained: "I would consider the meaning and how it feels, and then choose the word that most resembled that feeling." Another noted that interoceptive items required imagining how an internal sensation should sound, leading them to rely on general impressions of softness, harshness, or rhythm. This second-person data illustrates how participants projected cross-modal associations onto forms even when direct resemblance was absent, helping explain why interoceptive ideophones received high iconicity ratings despite lower guessing accuracy.

Together, these results support the conclusion that auditory ideophones are most transparent in terms of primary iconicity. Visual ideophones consistently occupy a middle position, benefiting from partial resemblance but lacking the transparency of sound imitation or the expressive force of interoceptive forms. This modality hierarchy provides nuanced evidence for the multi-layered nature of iconicity.

5.1.3 RQ3: To what extent does Japanese orthography influence perceptions of iconicity?

The role of orthography was less clear-cut. In terms of accuracy, orthographic condition had little systematic effect. Participants in the Audio-Only, Congruent Script, and Incongruent Script conditions performed at comparable levels, suggesting that the script in which ideophones were written did not substantially enhance or impair their ability to guess meanings. This finding contrasts with Iwahara, Hatta and Maehara (2003), who showed that script-type can influence perceived compatibility in native readers. For non-speakers, script appears to have been a weaker cue.

However, subtle effects emerged in the Rating Task. Participants in the Audio-Only condition provided the highest iconicity ratings overall, while those in the Congruent and Incongruent Script conditions rated ideophones slightly lower. This suggests that script did play a role, but not in the expected direction. Rather than enhancing iconic perception, the presence of script may have introduced noise or distraction, reducing participants' confidence in their judgments. A similar finding was reported by Turoman and Styles (2017), who showed that script-level visual features can bias sound-symbolic judgments in ways that do not always align with phonological cues.

Qualitative data further illuminate these effects. Several participants spontaneously reported being drawn to the shape of the kana, describing certain characters as "rounded for softer words" or "sharp/angular for harsher words". One participant remarked: "I attempted to match the image to the sound that I felt most appropriately fitting to the shape of the letters. This proved challenging due to my lack of familiarity." These comments demonstrate that even naïve participants used orthographic form as a source in making their introspective judgments , attempting to align visual features of kana with perceived meanings.

At the same time, others stated explicitly that they ignored the script, focusing entirely on auditory cues. One interviewee explained: "I never really focused on the visual representation. I only concentrated on the audio recordings." This variation in reported strategy aligns with the quantitative results: script did not reliably boost performance across the group, but for some individuals, it provided an additional, but inconsistent, dimension of perceived iconicity.

Overall, the findings indicate that orthography exerted weak and inconsistent effects for non-speakers. While script shapes were occasionally recruited into participants' judgments, they did not significantly alter group-level accuracy or ratings. This contrasts with findings among native speakers (Iwahara, Hatta & Maehara 2003) but supports the claim that orthographic iconicity is language-specific and requires cultural familiarity to function systematically. Ideophones may provide expressive resources that render them "vivid", but as shown by the study, this does not necessarily correspond to the strength of iconicity: the degree of resemblance between expression and content,

5.2 Limitations and further research

Several limitations temper these conclusions. The participant sample, drawn from non-Japanese speakers, may not generalize to populations with varying degrees of Japanese exposure or different orthographic literacies. The experiment employed a restricted stimulus set of thirty ideophones, which, although balanced across modalities, cannot capture the full diversity of the Japanese ideophone lexicon. The orthographic manipulation used only hiragana and katakana, excluding kanji-based forms and handwritten variation. Further, the study did not systematically analyze response times, which could provide further insight into processing differences between modalities. Finally, inferential statistics were not employed. These are all aspects that are to be improved in future research.

5.3 Returning to the conceptual side

Stepping back from the specifics, these findings offer a valuable opportunity to revisit the conceptual-empirical loop, including the three "what" questions (see **Figure 1**).

What is iconicity?

Iconicity is best understood not merely as resemblance but as a semiotic ground that motivates the relation between expression and content. Following Peirce and further elaborated by Sonesson, iconicity is a way in which the expression signifies its content by virtue of perceivable similarity, either on the levels of simple features, in imagistic iconicity, or by sharing structural qualities that one sees as analogous, in diagrammatic iconicity (Sonesson 2023: 205). Unlike the Saussurean assumption of arbitrariness, iconicity provides motivation to sign use and interpretation, showing how signs can be experienced as meaningful even prior to or beyond convention. At the same time, iconicity rarely exists in isolation. Signs typically combine all three Peircean grounds, iconicity, indexicality, and symbolicity, where iconicity supplies resemblance-based motivation, indexicality supplies contiguity-based motivation, and symbolicity provides a more conventionality-based meaning. In language, this means that iconicity is not opposed to conventionality but combines with it, as a motivating force that interacts with sedimented linguistic norms to form meaningful content.

Primary and secondary iconicity interact in any context and experimental task, including a Choosing Task-like paradigm, as explained by Ahlner and Zlatev (2010) (see Section 2.3). Judgments from the Choosing show that expression and content can resonate directly, while reflective ratings capture a sense of vividness and expressivity that does not necessarily reflect iconicity. This aligns with phenomenological accounts of intentionality, where more basic layers of intentionality involve immediate embodied experience, and reflective intentionality entails conscious interpretation, which is always fallible.

What is unimodality and cross-modality in iconicity?

Unimodal iconicity refers to resemblance-based mappings within the same sensory modality: for example, onomatopoeic words in speech imitating auditory experiences. Cross-modal iconicity, by contrast, involves mappings across different modalities, such as associating sound with shape or movement (Ahlner & Zlatev 2010). In many cases, visual mappings are robustly perceived, as shown by the bouba–kiki effect (Ramachandran & Hubbard 2001), and related findings, where auditory forms are mapped to visual shapes above chance levels. However, when the mappings involve more elusive experiential domains such as, the arguably vague "modality" referred to in

this thesis as *interoceptive*, which cannot be reduced to a discrete sensory modality, the basis for cross-modal iconicity becomes less clear. This aligns with the present findings that ideophones signifying interoceptive or affective states were the hardest for participants to guess, suggesting that such mappings verge on the illusory: they are perceived as iconic only retrospectively, once conventional associations have been learned. In this way, unimodal iconicity is often primary and directly guessable, while cross-modal iconicity, particularly in the interoceptive domain, tends to be secondary and more dependent on cultural and linguistic sedimentation.

What is language?

Language cannot be defined in terms of arbitrariness, as Saussure insisted (Saussure 1959: 67). Such a position on the centrality of arbitrariness fails to account for the pervasive presence of motivated signs in natural languages, from ideophones to sound-symbolic patterns and gesture-speech ensembles (Dingemanse 2012; Zlatev 2018). Instead, language should be defined as a semiotic system, emerging from and transcending predominantly iconic (and indexical) semiotic systems such as mimesis and gesture (Zlatev 2018: 5–6). It can be thought of as *a highly structured and conventional, but non-arbitrary semiotic system in which iconic, indexical, and symbolic grounds interact, stabilized by conventions but continually renewed through embodied experience and intersubjective negotiation at actual situations*. Language is thus not an arbitrary but a motivated system in which iconicity plays a central role in grounding meaning, particularly in ideophones, where resemblance and convention are fused in a way that illustrates the dynamic interplay of primary motivation and secondary sedimentation.

References

Ahlner, Felix & Jordan Zlatev (2010): Cross-modal iconicity: A cognitive semiotic approach to sound symbolism. *Sign Systems Studies* 38(1/4). 298–348. doi:10.12697/SSS.2010.38.1-4.11.

Benveniste, Émile (1981): The semiology of language. (Trans.) Genette Ashby & Adelaide Russo *Semiotica* 37(s1) doi:10.1515/semi.1981.37.s1.5.

Brandt, Per Aage (2011): What is cognitive semiotics? A new paradigm in the study of meaning. *Signata. Annales des sémiotiques / Annals of Semiotics*. Université de liège. (2). 49–60. doi:10.4000/signata.526.

Breen, James (2004): JMdict: a Japanese-multilingual dictionary. *Proceedings of the Workshop on Multilingual Linguistic Ressources - MLR '04*, 71. Geneva, Switzerland: Association for Computational Linguistics. doi:10.3115/1706238.1706248.

Cuskley, Christine, Julia Simner & Simon Kirby (2017): Phonological and orthographic influences in the bouba–kiki effect. *Psychological Research* 81(1). 119–130. doi:10.1007/s00426-015-0709-2.

Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, et al. (2021): The bouba/kiki effect is robust across cultures and writing systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society. 377(1841). 20200390. doi:10.1098/rstb.2020.0390.

De Carolis, Léa, Egidio Marsico, Vincent Arnaud & Christophe Coupé (2018): Assessing sound symbolism: Investigating phonetic forms, visual shapes and letter fonts in an implicit bouba-kiki experimental paradigm. (Ed.) Veronica Whitford *PLOS ONE* 13(12). e0208874. doi:10.1371/journal.pone.0208874.

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 (2019): Chapter 1. 'Ideophone' as a comparative concept. In Kimi Akita & Prashant Pardeshi (Hrsg.), *Ideophones, Mimetics and Expressives* (Iconicity in Language and Literature), 13–34. John Benjamins Publishing Company. doi:10.1075/ill.16.02din.

Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan (2015): Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* 19(10). 603–615. doi:10.1016/j.tics.2015.07.013.

Eco, Umberto (1976): *A theory of semiotics* (Advances in Semiotics). Bloomington and Indianapolis: Indiana University Press.

Erben Johansson, Niklas, Andrey Anikin, Gerd Carling & Arthur Holmer (2020): The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 24(2). 253–310. doi:10.1515/lingty-2020-2034.

Gallagher, Shaun & Dan Zahavi (2010): *The phenomenological mind: an introduction to philosophy of mind and cognitive science*. Repr. London: Routledge.

Giraldo, Veronica (2020): Referential iconicity in music and speech. *Public Journal of Semiotics* 9(1). 41–77. doi:10.37693/pjos.2019.9.20283.

Haiman, John (1985): *Natural Syntax*. Cambridge: Cambridge University Press.

Hamano, Shōko (1998): *The sound symbolic system of Japanese* (Studies in Japanese linguistics). Stanford, Calif: CSLI Publ. [u.a.].

Hinton, Leanne, Johanna Nichols & John J. Ohala (Hrsg.) (1995): *Sound Symbolism*. 1. edn. Cambridge University Press. doi:10.1017/CBO9780511751806.

IACS About IACS. https://iacs.dk/index.html (letzter Zugriff 27.05.2025).

Iida, Hinano & Kimi Akita (2023): Perceptual Strength Norms for 510 Japanese Words, Including Ideophones: A Comparative Study with English. *Proceedings of the Annual Meeting of the Cognitive Science Society* 45(45).

Imai, Mutsumi & 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, Kimi Akita, Noburo Saji, Masato Ohba & Miki Namatame (2025): Does sound symbolism need sound?: The role of articulatory movement in detecting iconicity between sound and meaninga). *The Journal of the Acoustical Society of America* 157(1). 137–148. doi:10.1121/10.0034832.

Iwahara, Akihiko, Takeshi Hatta & Aiko Maehara (2003): The effects of a sense of compatibility between type of script and word in written Japanese. *Reading and Writing* 16(4). 377–397. doi:10.1023/A:1023668804595.

Iwasaki, Noriko, P Vinson David & Vigliocco Gabriella (2007): What do English Speakers Know about gera-gera and yota-yota?: A Cross-linguistic Investigation of Mimetic Words of Laughing and Walking. *国際交流基金日本語事業部企画調整課*. doi:10.20649/00000380.

Jakobson, Roman (1965): Quest for the Essence of Language. *Diogenes* 13(51). 21–37. doi:10.1177/039219216501305103.

Keppel, Geoffrey & Thomas D. Wickens (2004): *Design and analysis: a researcher's handbook*. Upper Saddle River, N.J: Pearson Prentice Hall.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: ten years on. *Lexicography* 1(1). 7–36. doi:10.1007/s40607-014-0009-9.

Kita, Sotaro (1997): Two-dimensional semantic analysis of Japanese mimetics. De Gruyter Mouton. 35(2). 379–416. doi:10.1515/ling.1997.35.2.379.

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

Konderak, Piotr (2018): *Mind, Cognition, Semiosis: Ways to Cognitive Semiotics*. Lublin, Polska: Maria Curie-Sklodowska University Press.

Kunihira, Shirou (1971): Effects of the expressive voice on phonetic symbolism. *Journal of Verbal Learning and Verbal Behavior* 10(4). 427–429. doi:10.1016/S0022-5371(71)80042-7.

Li, Wenjing & Jordan Zlatev (2022): Intersemiotic translation from fairy tale to sculpture: An exploration of secondary narrativity. *Sign Systems Studies* 50(2–3). 317–345. doi:10.12697/SSS.3.

Lockwood, Gwilym, Mark Dingemanse & Peter Hagoort (2016): Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(8). 1274–1281. doi:10.1037/xlm0000235.

McLean, Bonnie (2021): Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*. De Gruyter Mouton. 25(3). 507–549. doi:10.1515/lingty-2020-2063.

McLean, Bonnie, Michael Dunn & Mark Dingemanse (2023): Two measures are better than one: combining iconicity ratings and guessing experiments for a more nuanced picture of iconicity in the lexicon. *Language and Cognition* 15(4). 716–739. doi:10.1017/langcog.2023.9.

Mendoza-Collazos, Juan & Jordan Zlatev (2022): A Cognitive-Semiotic Approach to Agency: Assessing Ideas from Cognitive Science and Neuroscience. *Biosemiotics* 15(1). 141–170. doi:10.1007/s12304-022-09473-z.

Pae, Hye K. (2020): Linguistic Evidence for Script Relativity. *Script Effects as the Hidden Drive of the Mind, Cognition, and Culture* (Literacy Studies), vol. 21, 147–174. Cham: Springer International Publishing. doi:10.1007/978-3-030-55152-0_8.

Peirce, Charles S. (1985): The Icon, Index, and Symbol. In Charles Hartshorne & Paul Weiss (Hrsg.), *Collected papers of Charles Sanders Peirce*. 5. [printing]. Cambridge, Mass.: Belknap Press of Harvard Univ. Press.

Perniss, Pamela, Robin L. Thompson & Gabriella Vigliocco (2010): Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology* 1 doi:10.3389/fpsyg.2010.00227.

Punselie, Robin, Bonnie McLean & Mark Dingemanse. 2024. The bouba–kiki effect and beyond: Measuring iconicity in the lab. *Glossa: A Journal of General Linguistics* 9(1). 1–27.

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

Sapir, E. (1929): A study in phonetic symbolism. *Journal of Experimental Psychology* 12(3). 225–239. doi:10.1037/h0070931.

Saussure, Ferdinand de (1959): *Course in general linguistics*. (Ed.) Charles Bally & Albert Reidlinger. (Trans.) Wade Baskin New York: Philosophical Library.

Sedley, D. N. (2003): *Plato's Cratylus* (Cambridge studies in the dialogues of Plato). Cambridge, UK ; New York: Cambridge University Press.

Sokolowski, Robert (2000): *Introduction to Phenomenology*. 1. edn. Cambridge University Press. doi:10.1017/CBO9780511809118.

Sonesson, Göran (1997): The ecological foundations of iconicity. In Irmengard Rauch & Gerald F. Carr (Hrsg.), *Semiotics around the World: Synthesis in Diversity: Proceedings of the Fifth Congress of the IASS Berkeley, June 12-18, 1994*, 739–742. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110820065.

Sonesson, Göran (2007): From the meaning of embodiment to the embodiment of meaning. In Zlatev Jordan, Zimke Tom & Frank R (Hrsg.), *Body, Language and Mind. Vol 1. Embodiment*, vol. I, 85–128. Mouton de Gruyter.

Sonesson, Göran (2010a): Semiosis and the elusive final interpretant of understanding. *Semiotica* 2010(179) doi:10.1515/semi.2010.023.

Sonesson, Göran (2010b): From mimicry to mime by way of mimesis: Reflections on a general theory of iconicity. *Sign Systems Studies* 38(1/4). 18–66. doi:10.12697/SSS.2010.38.1-4.02.

Sonesson, Göran (2012a): The Foundation of Cognitive Semiotics in the Phenomenology of Signs and Meanings. *Intellectica. Revue de l'Association pour la Recherche Cognitive* 58(2). 207–239. doi:10.3406/intel.2012.1105.

Sonesson, Göran (2012b): Semiosis Beyond Signs. On Two or Three Missing Links on the Way to Human Beings. In Theresa Schilhab, Frederik Stjernfelt & Terrence Deacon (Hrsg.), *The Symbolic Species Evolved* (Biosemiotics), vol. 6, 81–93. Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-2336-8_5.

Sonesson, Göran (2023): Iconicity and Semiosis. In Jamin Pelkey (Hrsg.), *Bloomsbury Semiotics Volume 1: History and Semiosis*, 193–215. 1. edn. London,: Bloomsbury Academic.

Thompson, Arthur Lewis, Kimi Akita & Youngah Do (2020): Iconicity ratings across the Japanese lexicon: A comparative study with English. *Linguistics Vanguard* 6(1). 20190088. doi:10.1515/lingvan-2019-0088.

Turoman, Nora & Suzy J. Styles (2017): Glyph guessing for 'oo' and 'ee': spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts. *Royal Society Open Science* 4(9). 170882. doi:10.1098/rsos.170882.

Zahavi, Dan (2019): *Phenomenology: the basics* (The Basics). London New York: Routledge, Taylor & Francis Group.

Zlatev, Jordan. (2012). Cognitive semiotics: An emerging field for the transdisciplinary study of meaning. *Public Journal of Semiotics* 4(1). 2–24.

Zlatev, Jordan (2015): Cognitive Semiotics. In Peter Pericles Trifonas (Hrsg.), *International Handbook of Semiotics*, 1043–1067. Dordrecht: Springer Netherlands. doi:10.1007/978-94-017-9404-6_47.

Zlatev, Jordan (2018): Meaning making from life to language: The Semiotic Hierarchy and phenomenology. *Cognitive Semiotics* 11(1). 20180001. doi:10.1515/cogsem-2018-0001.

Zlatev, Jordan (2023): The Intertwining of Bodily Experience and Language: The Continued Relevance of Merleau-Ponty. *Histoire Épistémologie Langage*. Société d'histoire et d'épistémologie des sciences du langage. (45–1). 41–63. doi:10.4000/hel.3373.

Zlatev, Jordan (in press): Five pillars of cognitive semiotics.

Zlatev, Jordan, Simon Devylder, Rebecca Defina, Kalina Moskaluk & Linea Brink Andersen (2023): Analyzing polysemiosis: language, gesture, and depiction in two cultural practices with sand drawing. *Semiotica*. De Gruyter Mouton. 2023(253). 81–116. doi:10.1515/sem-2022-0102.

Zlatev, Jordan & Piotr Konderak (2023): Consciousness and Semiosis. In Jamin Pelkey (Hrsg.), *Bloomsbury Semiotics Volume 1: History and Semiosis*, 169–192. 1. edn. London,: Bloomsbury Academic.

Zlatev, Jordan & Alexandra Mouratidou (2024): Extending the Life World: Phenomenological Triangulation Along Two Planes. *Biosemiotics* 17(2). 407–429. doi:10.1007/s12304-024-09576-9.

Zlatev, Jordan, Marta Sibierska, Przemysław Żywiczyński, Joost Van De Weijer & Monika Boruta-Żywiczyńska (2024): Chapter 5. Can pantomime narrate?: A cognitive semiotic approach. In Przemysław Żywiczyński, Johan Blomberg & Monika Boruta-Żywiczyńska (Hrsg.), *Advances in Interaction Studies*, vol. 12, 115–138. Amsterdam: John Benjamins Publishing Company. doi:10.1075/ais.12.05zla.

Zlatev, Jordan, Przemysław Żywiczyński & Sławomir Wacewicz (2020): Pantomime as the original human-specific communicative system. *Journal of Language Evolution* 5(2). 156–174. doi:10.1093/jole/lzaa006.

Appendix

Appendix A   Practice and Trial Items

pairing	modality	hiragana word	hiragana reading	hiragana meaning	katakana word	katakana reading	katakana translation
p0	auditory	そっと	sotto	softly, gently	ガタン	gataN	with a bang
p1	visual	じっと	zitto	motionless, fixedly	パッ	paQ	suddenly, in a flash
p2	auditory	そろそろ	sorosoro	slowly, quietly	ガンガン	gaNgaN	clanging, banging
p3	visual	そっくり	sokkuri	exactly alike, spitting image	ガラリ	garari	completely, totally changed

pairing	modality	hiragana word	hiragana reading	hiragana meaning	katakana word	katakana reading	katakana translation
a0	auditory	ごそごそ	gosogoso	with a rustling sound	カタカタ	katakata	clattering, rattling
a1	auditory	しとしと	sitosito	splashing	バチャバチャ	batyabatya	splashing
a2	auditory	じゃーじゃー	zyaazyaa	noisily gushing	ポタポタ	potapota	dripping, trickling
a3	auditory	じりじり	ziriziri	crisp, crunchy	シャキシャキ	syakisyaki	crisp, crunchy
a4	auditory	ずるずる	zuruzuru	with a slurp	ポリポリ	poripori	munching, crunching
a5	auditory	そろりそろり	sororisorori	noisily, with heavy feet	ドタバタ	dotabata	noisily, with heavy feet
a6	auditory	どさり	dosari	with a thud	カチッ	katitto	click, snap
a7	auditory	どしどし	dosidosi	with a creak, squeak	ギシギシ	gisigisi	with a creak, squeak
a8	auditory	ぼそり	bosori	in a whisper, in a murmur	キャーキャー	kyaakyaa	shrieking, squealing
a9	auditory	もごもご	mogomogo	with a crunch	サクッ	sakutto	with a crunch
v0	visual	くっきり	kukkiri	clearly, distinctly, sharply	チラチラ	tiratira	flickering, fluttering
v1	visual	くりくり	kurikuri	jagged, serrated	ギザギザ	gizagiza	jagged, serrated
v2	visual	どんより	doNyori	dark, gloomy	キラキラ	kirakira	glittering, sparkling
v3	visual	ぴたり	pitari	fluttering, dangling	ヒラヒラ	hirahira	fluttering, dangling
v4	visual	びっしり	bissiri	closely lined up, densely	バラバラ	barabara	scattered, in pieces
v5	visual	ぶすり	busuri	smilingly, with a grin	ニコニコ	nikoniko	smilingly, with a grin
v6	visual	ふんわり	huNwari	gently, airily, fluffy	ゴツゴツ	gotugotu	rugged, scragged, angular
v7	visual	ぼんやり	boNyari	crisp appearance, stiffly	キリッ	kiritto	crisp appearance, stiffly
v8	visual	まるまる	marumaru	plump, rotund, chubby	ガリガリ	garigari	very skinny, emaciated

v9	visual	むくむく	mukumuku	ruffled, disheveled	ボサボサ	bosabosa	ruffled, disheveled
i0	interoceptive	うんざり	uNzari	boredom, tedious, fed up with	ワクワク	wakuwaku	excitement, nervous
i1	interoceptive	げんなり	geNnari	ravenously, greedily	ガツガツ	gatugatu	ravenously, greedily
i2	interoceptive	しょぼん	syoboN	downhearted, dejected	ルンルン	ruNruN	happy, elated, euphoric
i3	interoceptive	すっきり	sukkiri	to get irritated, annoyed	イライラ	iraira	to get irritated, annoyed
i4	interoceptive	ちゃんと	tyaNto	attentive, proper attitude	ドギマギ	dogimagi	flurried, nervous
i5	interoceptive	のびのび	nobinobi	anxious, nervous	ハラハラ	harahara	anxious, nervous
i6	interoceptive	のんびり	noNbiri	at leisure, in a relaxed manner	ギリギリ	girigiri	just barely, at the last moment
i7	interoceptive	ほっこり	hokkori	shivering, feeling chilly	ゾクゾク	zokuzoku	shivering, feeling chilly
i8	interoceptive	ほっ	hotto	with a feeling of relief	ドキリ	dokiri	being startled, getting a shock
i9	interoceptive	ゆったり	yuttari	with a rapid heartbeat	ドキドキ	dokidoki	with a rapid heartbeat

Appendix B   Informed Consent Form

JOINT FACULTIES OF HUMANITIES AND THEOLOGY

LUND UNIVERSITY

Informed Consent to Participate in a Study on Iconicity

This study is a part of a master's thesis in cognitive semiotics at the Joint Faculties of Humanities and Theology, at Lund University, supervised by Prof. Jordan Zlatev. It consists of two tasks and takes about 25 minutes to complete in total.

The purpose is to study how well people can guess words in languages they don't know. In the first task, you'll be given 30 pairs of two Japanese words and asked to match each with their English translation out of two options. In the subsequent task you will be asked how well some specific Japanese words resemble their meanings.

To participate in the study requires that you do not speak or understand Japanese. You will also need to be able to hear the words, so please make sure you sit in a quiet place, preferably with headphones and on a laptop or tablet. We will check for audio at the start, and you won't be able to continue without it.

Participation is entirely voluntary, and your answers will be anonymous and confidential. You have the right to cease participation at any time and withdraw from the study, should you choose to do so. If you meet these requirements and want to participate, please check the following boxes and click '**Next**' to continue.

Researcher:
Nils Paulsson
ni7728pa-s@lu.se
+46 702089400

Supervisor:
Jordan Zlatev
jordan.zlatev@ling.lu.se

☐ I am age 18 or older

☐ I have read and understand the information above

☐ I understand that I am free to withdraw from engaging with the study at any time, by closing the browser tab, without giving any reason

☐ I consent to participate in this research and want to continue

Appendix C   Task Instructions

Choosing Task Instructions

In this task, you will be presented with 30 word pairs. It takes about 10–15 minutes in total to complete. In each trial, you will see two Japanese words and their English meanings. Your task is to **match one** translation with the Japanese word you think best fits that meaning.

First, two words will **play automatically** and appear side by side. Click either to replay as needed.

Then, **read their translations** and decide which one is which. One will be highlighted as the target meaning.

Click to select the Japanese word below that you feel best fits that meaning. This will start the next trial.

When ready, press '**Next**' to begin a practice run.

Next

Rating Task Instructions

In this task, you will rate 30 words from the previous task. It takes about 5–10 minutes in total to complete. In each trial, you will be shown a Japanese word together with its English meaning. Your task is to rate how much you think the word resembles its meaning on a scale from **1** to **7**.

1 indicates **no resemblance**
while **7** indicates **strong resemblance**.

Click on a number (1–7) to select your rating, then press 'Next' to submit your response.

When ready, press '**Next**' to begin a practice run.

Next

Listen to these two Japanese words.
Click to replay.

△   ▽

One of them means **completely, totally changed**

The other means **exactly alike, spitting image**

Which one do you think means
exactly alike, spitting image?
Click one to select:

△	▽

Listen to these two Japanese words.
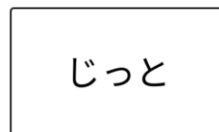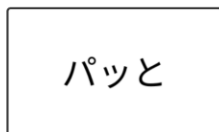Click to replay.

パッと   じっと

One of them means **motionless, fixedly**

The other means **suddenly, in a flash**

Which one do you think means
suddenly, in a flash?
Click one to select:

パッと	じっと

Listen to these two Japanese words.
Click to replay.

ぱっと   ジット

One of them means **motionless, fixedly**

The other means **suddenly, in a flash**

Which one do you think means
suddenly, in a flash?
Click one to select:

| ぱっと | ジット |

Listen to the Japanese word below.
Click to replay.

Replay

△

It means **exactly alike, spitting image**

Do you think there is a resemblance
between the word and its meaning?

No
resemblance | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Strong
resemblance

Next