The Differentiation of Early Word Meanings from Global to Specific Categories:

Towards a Verification of the "Semantic Pluripotency Hypothesis"

Hiromichi Hagihara

The Differentiation of Early Word Meanings from Global to Specific Categories:
Towards a Verification of the “Semantic Pluripotency Hypothesis”

> 言語発達初期における語の意味の未分化性と可塑的変化：
> 「胚性詞」仮説の検証に向けて

Hiromichi Hagihara

A dissertation submitted to Kyoto University
in partial fulfillment of the requirements for the degree of
Doctor of Human and Environmental Studies

Summary

In most languages, early nouns that semantically correspond to object categories (e.g., “shoe”) dominate young children’s vocabulary when compared to early verbs that correspond to action categories (e.g., “put on”). Many developmental scientists have been trying to investigate why such object words, rather than action words, are advantageous for learning. In considering this issue, some researchers argue that children’s early words consist of label-meaning connections, which are characteristically different from those of older children and adults. In particular, Werner and Kaplan (1963) theoretically posited that the initial meanings of early words do not sufficiently differentiate into specific categories, but rather, correspond to holistic and global event categories as a total situation, in which multiple components, such as objects and actions, are intimately fused. Thus, it is important to ask whether early noun meanings are inherently specific object categories from the very beginning of learning, or if they are at first undifferentiated event categories and subsequently differentiated into specific categories. However, an observational approach alone cannot directly address this question, because even if the meanings of children’s early words are not differentiated, they might not be reflected in their phonetic features; hence, the undifferentiated and specific word meanings could not be distinguished based only on their corresponding labels.

This thesis proposes, as a revision of Werner and Kaplan’s (1963) theoretical assumption, the “semantic pluripotency hypothesis” to experimentally investigate the semantic contents of early words and their development. This hypothesis consists of two sub-hypotheses: First, the initial meanings of children’s words are pluripotent in nature, as they correspond to the context-bound fusion of various factors that emerge from children’s experiences related to the word. Second, such word meanings have plasticity, as they dynamically differentiate into specific, discrete, and more decontextualized categories with later development. This thesis investigates and discusses these sub-hypotheses with a focus on early nouns through five chapters. Specifically, it is posited that (1) the initial meanings of object words are undifferentiated global event categories that include both objects and actions as a fusion, and (2) they subsequently differentiate into specific object categories that are independent of actions.

Chapter 1 provides an overview of the general characteristics of early vocabulary and how young children learn words. After reviewing theoretical, observational, and experimental studies, we focus on the uniqueness of young children’s words. We discuss

how these words differ from those of older children and adults with reference to Werner and Kaplan's theoretical approach, and the experimental findings that support, though not directly, their theory. We then propose the semantic pluripotency hypothesis, a new and experimentally verifiable hypothesis on the flexibility and plasticity of early word learning, and present an outline of this thesis.

Chapter 2 provides the first experimental findings that the initial meanings of object words do not sufficiently differentiate into specific object categories. Using a two-alternative forced-choice task, we preliminarily investigated whether and how object word meanings changed with development. Japanese monolingual toddlers aged 19–35 months (n = 36) watched two juxtaposed video stimuli, in which a girl was performing an action using certain objects, and were then prompted to choose one of the stimuli according to questions about familiar object words (e.g., which ones are shoes?). Statistical modeling demonstrated that even young participants were able to select the apropos stimulus when objects and object-specific actions were compatible on one side (e.g., "putting shoes on" vs. "rubbing two baskets in front of her"). However, when objects and object-specific actions were incompatible (e.g., "rubbing shoes in front of her" vs. "putting on two baskets as if they were shoes"), the probability of selecting a stimulus that included target objects remained at chance level for participants less than 21 months of age or with a fewer than 140 words vocabulary size; whereas the probability steeply increased for older participants or those with a larger vocabulary size. These results suggest that both objects and object-specific actions are entwined in the initial meanings of object words and, only later, object word meanings are differentiated into specific object categories independent of actions. Although the robustness of these results was unclear, due to a preliminary study, the results supported the semantic pluripotency hypothesis.

Chapter 3 describes the extent to which looking and pointing responses are equivalently interpretable in forced-choice tasks, to prepare for the replication of the findings in Chapter 2 in more children around the critical period when the semantic differentiation of object words likely occurs. In the preliminary study, we observed that toddlers under the age of two often did not provide clear pointing responses in the forced-choice task, although they appeared to spend more time looking at one of the stimuli presented simultaneously. Therefore, it was necessary to measure both pointing and looking responses in subsequent studies to reduce data loss. However, so far, the validity of treating these different indices as equivalent has not been attested. Thus, we aimed to investigate how accurately pointing responses (i.e., left or right) could be predicted from concurrent preferential looking. Using part of the video data of toddlers aged 18–23 months (n = 48), which were obtained in the experiments in Chapter 4, we developed models that predicted pointing from looking responses. The results showed that the prediction accuracy for the proposed models was substantial (85.8–89.7% agreement), indicating that looking responses would be reasonable alternative indices for pointing responses. However, further exploratory analysis revealed that looking responses without pointing responses would be qualitatively different from those with pointing responses. These findings suggest the need of using both pointing and looking indices for analysis, so that the obtained data can be interpreted in more detail. The models proposed in this chapter enable us to apply the same forced-choice task used in Chapter 2 to younger toddlers without increasing missing data. They also allow us to conduct the same statistical analysis for both pointing and looking measurements, and make a direct comparison of the results from these different indices.

Combining the findings obtained in the previous chapters, Chapter 4 examines the semantic pluripotency hypothesis more thoroughly. In addition to confirming the robustness of the previous findings, we further investigated whether toddlers could appropriately understand object word referents solely based on object-specific actions, and how developmental changes in object word meanings were related to concurrent and later vocabulary growth. Using both cross-sectional (n = 69) and longitudinal (n = 16) data of 18–23-month-old toddlers, we found that only younger participants could not choose the correct video stimulus that matched object words when objects and object-specific actions were presented separately (e.g., “rubbing shoes” vs. “putting on two baskets”), despite the success when objects and object-specific actions were matched. Older participants were able to select the appropriate stimulus for both conditions. Although the detected critical period of when the semantic differentiation of object words occurred was a few months earlier than in the preliminary study, it was certain that object word meanings developmentally changed steeply during the latter half of the second year of the children’s lives. The results from additional conditions demonstrated that participants of all age ranges failed to judge object word referents solely by object-specific actions (e.g., “putting on two baskets” vs. “rubbing two baskets”). These results are robust for both the pointing and preferential looking measurements. Taken together, these results indicate that the initial meanings of object words are global event categories comprising both objects and actions as a fusion (e.g., “putting shoes on”), and they later differentiate into specific object categories that were independent of actions (e.g., “shoes” alone). Moreover, the degree of semantic differentiation of object words was positively related to both concurrent and subsequent vocabulary sizes of action words in particular. This suggests that the differentiation of object word meanings encouraged toddlers to develop new label-meaning connections that can be used for specific action categories, apart from objects.

Finally, Chapter 5 summarizes the key findings and refines the semantic

pluripotency hypothesis based on them. Our findings corroborate the semantic pluripotency hypothesis, provide the first experimental support for Werner and Kaplan's (1963) theoretical hypothesis, and set the stage for future research on children's early word learning. Additionally, the semantic pluripotency hypothesis can contribute to the integration of interrelated, but separately explored, research topics such as event categorization, contextual effects on word learning, and cross-situational statistical word learning. We discuss such theoretical implications as well as practical implications for caregivers, educators, and clinicians. Overall, although the semantic pluripotency hypothesis is still in its nascent stage, the experimental exploration of this thesis demonstrates semantic flexibility and plasticity with development and will contribute to illuminating more aspects of the uniqueness of children's early words.

Keywords*: language development, semantic pluripotency, word meaning differentiation, object-specific action, event category, two-alternative forced-choice task*

Acknowledgments

The road to my Ph.D. was quite winding and bumpy, during which a number of people have supported and helped me in various ways. Several unpredictable things happened to me: I transferred from the Faculty of Integrated Human Studies to the Faculty of Medicine at Kyoto University; I quit my job as a pediatric occupational therapist and entered the Graduate School of Human and Environmental Studies; and I further transferred from that department. In fact, deciding to pursue a Ph.D. was unexpected for me. I am quite sure that I would not have been able to stand on this road by myself. I would like to thank all those I encountered along the way.

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Masa-aki Sakagami, for his encouragement, patience, and support. For all the biggest turning points on my road, of both my public and private life, he was there. In this sense, I even feel as if he was my father in Kyoto. Despite his specialty in nonlinear physics, he has an astonishing sensitivity to distinguish and discover what is interesting in any field of science, for which I most admire him. I think it was a brave decision to permit me to belong to his laboratory because of different majors. I was extraordinarily lucky to have worked with him.

I would also like to express my appreciation to Dr. Yusuke Moriguchi for supervising my Ph.D. research. Dr. Moriguchi welcomed me to participate in the regular seminar of his lab despite having a different affiliation, and gave me a very helpful, insightful advice. He always looked after me, giving constructive comments, and correcting my discussions to avoid going off topic. The quality of my thoughts and writing was supported by his assistance. It is an honor to have worked with him.

It was an immense pleasure to have Prof. Kazumi Taniguchi on my dissertation committee. Prof. Taniguchi gave me valuable opportunities to learn a lot about the theoretical aspects of linguistics before and during my Ph.D. studies. For instance, she encouraged me to give oral presentations at conferences such as the Conference of the Japanese Cognitive Linguistics Association and the Whole Group Meeting of Evolinguistics. She has helped me situate my research properly in the context of language development theories.

I would like to extend my gratitude to my colleagues, Dr. Naoto Ienaga, Dr. Kei Terayama, and Dr. Hiroki Yamamoto for kindly collaborating with me. Dr. Ienaga took charge of Python coding regarding machine learning. Dr. Terayama helped me consolidate unorganized ideas and interact with anonymous referees during the review process. Dr. Yamamoto helped me conduct the Bayesian statistical modeling. All of them were incredibly patient and cheerful, taking time for discussion on research in progress over and over again (even at midnight!). I could not have written this thesis without their engagement and encouragement. I am also grateful to the collaborators of my research. I thank Dr. Kei Shimonishi and Mr. Shingo Sawa for helping me develop a system that could simultaneously show experimental stimuli, record participants via webcam, and record touch responses; Ms. Mana Yada, Mr. Tomohiko Kuzuba, Mr. Tenchi Mizutani, and Mr. Tomoya Suzuki for assistance with data coding and analysis; Ms. Maho Kondo and Mr. Yoshio Tsutsumi (C223) for drawing wonderful illustrations; Ms. Mariko Kida for acting in the video stimuli; and Ms. Tomomi Hirosawa for managing my research budget.

I deeply thank all those who participated in my studies and cooperated in the collection of data. Participants of my studies were recruited from various kindergarten, nursery schools, and daycare centers across Japan. As a practitioner as well as a researcher, I would like to bring my research findings back to these practical or clinical fields in the future. I would also like to continue my research on developmental science in order to improve the lives and environments of all children, families, and teachers in diverse situations.

I cannot mention all the people who supported or encouraged me in various ways here due to the limitation of space; however, I would like to say thank you so much to everyone. I believe my gratitude will reach all of those involved. I would like to mention several groups that deeply affected my Ph.D. research and thesis writing to express my thanks: Sojin no Mikata (an interdisciplinary educational project of the preparing future faculty); Decobocollabo (an interdisciplinary research team); the Body and Language Research Team; the Center for Early Childhood Development, Education, and Policy Research at the University of Tokyo (Cedep); the KIT Language and Motion Research Team; the IRCN Babylab at the University of Tokyo; Kotoboo (an international team for creating enjoyable comic contents of language development); Narakan (an interprofessional team for the improvement of children’s daily activities); the Evolinguistics Young Researchers’ Association; and the Japanese Young Therapist Association. In addition , my future research plan was developed through interactions with my future mentors, Dr. Sho Tsuji and Prof. Sotaro Kita. I am grateful and excited to have the opportunity to work with them.

This research was financially supported by a Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Research Fellows (18J21948), JSPS Grant-in-Aid for Scientific Research on Innovative Areas (20H05002), and Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research. Part of the research in this thesis was also conducted under the Young Scholar Award for Distinguished Research in 2016 by the Japan Association of Applied Psychology.

Finally, I thank my marvelous friends and family for always supporting me. They were patient and kind, even when I changed my major, quit my job, and decided to pursue my Ph.D. course. I was able to keep myself motivated thanks to them. Thank you so much for being with me at any time, in any place.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Based on (but significantly revised and expanded):

Hagihara, H., & Sakagami, M. (2019). Shoki-gengo ni okeru imi no zentaisei to kasoteki henka: Kodomo no kotoba ni hinshi-kouzou wa aruka? [Semantic pluripotency in early word learning: Do children have an explicit part-of-speech structure?]. *Baby Science*, *18*, 14–24. Retrieved from https://www.crn.or.jp/LABO/BABY/LEARNED/18/2018-18_2_Hagihara.pdf

> Finally, we have to notice that when the child first uses and responds to adult words referentially, he is referring not so much to an object—a thing within the situation—as to the situation as a whole. At this stage words are far from being simply names, means of the static representation of objects. A word that the child speaks is to be regarded rather as one means by which he responds to a situation, as an essential part of his total reaction to it. (Lewis, 1936, p. 159)

1.1 Introduction

What does it mean when a child who is just beginning to speak mutters the word "shoe"? Generally speaking, common nouns, such as "shoe," are names of objects that correspond to object categories. These common nouns do not directly allude to actions or activities, such as "to put on" or "to take a walk." However, during early language development, the label "shoe" does not necessarily mean the name of an object, even when only a single noun or one-word utterance is produced (Dewey, 1894; Lewis, 1936; McCune, 2008; Stevenson, 1893; Werner & Kaplan, 1963). At this age, the word "shoe" may have a more complex meaning, such as "I put my shoes on.", "Look at those cool shoes.", or "Let's go out together." Although young children might know that a common noun corresponds to a specific object category, children may use one-word utterances to express more complex thoughts or demands because of their limited vocabulary. In fact, some researchers have distinguished intentional or purposeful meanings of such utterances from referential meanings (e.g., Dore, 1975).

In contrast to this explanation, which is based on a pragmatic view, the use of single word nouns may be further understood with a semantic explanation. This implies that children's initial meanings of common nouns may not sufficiently differentiate into specific object categories. Instead, early words may correspond to holistic, global, and undifferentiated events that emerge from children's experiences (Werner & Kaplan, 1963). For example, the meaning of the word "shoe" for very young children may not reflect the "shoe" object category but an event category of "putting shoes on and going out." Werner and Kaplan (1963) suggest that during early language development, the meanings of children's words differentiate from global event categories into specific categories, such as objects or actions. Nouns used by toddlers would thus differ in terms of meaning from nouns employed by adults, as they are not easily classified by their parts of speech (e.g., nouns or verbs) (Church, 1961; Dewey, 1894; Lewis, 1936; McCune, 2008; Okamoto, 1982; Stevenson, 1893; Tomasello, 2003; Volterra, Bates, Benigni, Bretherton, & Camaioni, 1979; Vygotsky, 1934/1986). This semantic perspective was created based on careful observation of children's utterances. However, as we will see later, no experimental research has directly investigated whether the meanings of early words are global event categories at first or are specific object categories from when first learned. The present thesis therefore experimentally investigates this possibility to deeply understand the semantic flexibility and plasticity of early words.

In its first step, this chapter will provide an overview of the general characteristics of early vocabulary and how young children learn words. First, we will briefly review observational findings on the dominance of nouns in children's early vocabulary over other parts of speech in most languages. Second, after reviewing the theoretical explanations and related empirical discoveries of early noun learning (i.e., object words), we look into the effects of object-specific actions, basically expressed by verbs, on object word and object-category learning. Third, we focus on the uniqueness of young children's words. We discuss how these words differ from those of older children and adults with Werner and Kaplan's (1963) theoretical approach and the experimental findings that support their assumptions. We then propose the "semantic pluripotency hypothesis," a new and experimentally verifiable hypothesis on the flexibility and plasticity of early word learning. Finally, we present the outline of this thesis, which focuses on the developmental semantic change in common nouns, words that appear to universally dominate children's early vocabulary compared to other parts of speech.

1.2 The dominance of nouns in early vocabulary

Typically, children produce their first word around 12 months (Benedict, 1979; Blinkoff & Hirsh-Pasek, 2019; Moore, Dailey, Garrison, Amatuni, & Bergelson, 2019; Nelson, 1973b). While children's vocabulary gradually grows at the start of their second year of life, vocabulary acquisition increases more rapidly during the second half of their second year of life. This rapid growth is known as the "vocabulary spurt"[1] (Goldfield & Reznick, 1996; Mervis & Bertrand, 1995; Nazzi & Bertoncini, 2003; Nelson, 1973b). Although traditional observational approaches examining word comprehension focused on children's word production, word production and comprehension do not occur at the same time. While both aspects compose word learning, word comprehension has been revealed to precede word production[2] (Bergelson & Swingley, 2012; Benedict, 1979; Bornstein & Hendricks, 2012; Gershkoff-Stowe & Hahn, 2013; Goldin-Meadow, Seligman, & Gelman, 1976; Oviatt, 1980; Schafer & Plunkett, 1998).

In most languages, nouns hold the most dominant position in children's early vocabulary compared to other parts of speech when classifying early words into such grammatical components (Bornstein et al., 2004; Caselli, Casadio, & Bates, 1999; Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, in press; Gentner & Boroditsky, 2001), although, in some verb-friendly languages such as Mandarin, verbs are equivalent to nouns or occupy a larger part of early vocabulary than nouns (Brown, 1998; Choi & Gopnik, 1995; Tardif, 1996; Tardif et al., 2008; Tardif, Gelman, & Xu, 1999). These early vocabulary characteristics across languages have been well documented with caregiver-report-based methods such as the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1993, 1994, 2007) because these methods assess the

---

[1] It is also called the "naming explosion." Although the vocabulary spurt is a striking topic that has attracted the attention of many researchers, there is some evidence that casts doubt on the existence of such a phenomenon (Bloom, 2004; Ganger & Brent, 2004).

[2] Recent experimental studies showed that even 6–9-month-old infants have begun to connect object words to their corresponding meanings, such as food or body parts (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012). But label-meaning connections at this period appear to be more fragile than those of adults because such links depend on who (e.g., mothers or an experimenter) uttered the words (Parise & Csibra, 2012).

presence of a large vocabulary and are easy and inexpensive to administer (Frank, Braginsky, Yurovsky, & Marchman, 2017). Japanese is generally classified as a verb-friendly language (Imai, Haryu, & Okada, 2005; Imai et al., 2008; Waxman et al., 2013) because, for example, it allows one to omit both the subject and object of a sentence and verb meanings are narrower in Japanese than in English, as different verbs like "*kiru*," "*kaburu*," "*kakeru*" are used to express wearing shirts, a hat, and glasses, respectively. Nevertheless, nouns make up the bulk of Japanese early vocabulary (Ogura, 2007; Ogura, Watamaki, & Inaba, 2016). Using the CDI standardized data of children in Japan (Ogura & Watamaki, 2004) and the U.S. (Fenson et al., 1994), Ogura (2007) classified and compared the first 50 most common early acquired words by parts of speech (**Figure 1.1**). She reported that, like children in the U.S., the highest proportion of both comprehension and production vocabulary was occupied by common nouns, followed by social words, such as greetings, and then action words for children in Japan. Japanese early vocabulary consisted of rich onomatopoeic words (i.e., baby talk), but these words were reclassified according to their semantic categories. These data therefore support the view that the "noun bias" is a language-general characteristic of children's early vocabulary.

In languages such as English, most words are primarily divided into two classes, namely nouns and verbs, based on their grammatical and logical properties (Whorf, 1940/1956). Thus, in research on early language development, noun learning has traditionally been contrasted with verb learning (Gentner, 1982; Gentner & Boroditsky, 2001; Gleitman & Gleitman, 1992; Gogate & Hollich, 2016; Goldin-Meadow et al., 1976; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011; Twomey & Hilton, 2020; Waxman et al., 2013). Since nouns learned early generally correspond to object categories, such as "cup," and verbs learned early typically correspond to action categories, such as "eat," (Gogate & Hollich, 2016) researchers tried to theoretically and empirically explain why word-object connections are more advantageous to learn than word-action connections[3].

One of the most accepted theories is the "natural partitions hypothesis," proposed by Gentner (1982). This hypothesis suggests that nouns are learned earlier because concrete objects as referents of object words are more easily isolated from their environment than other elements such as actions, processes, and attributes. On the other hand, the "relational relativity hypothesis" suggests that predicates such as verbs are difficult to learn because

[3] Gogate and Hollich (2016) carefully avoided the use of the terms "noun-object" or "verb-action" connections and used the term "word" to refer to such linguistic labels when discussing preverbal children because it is unclear that children could distinguish these two grammatical categories.

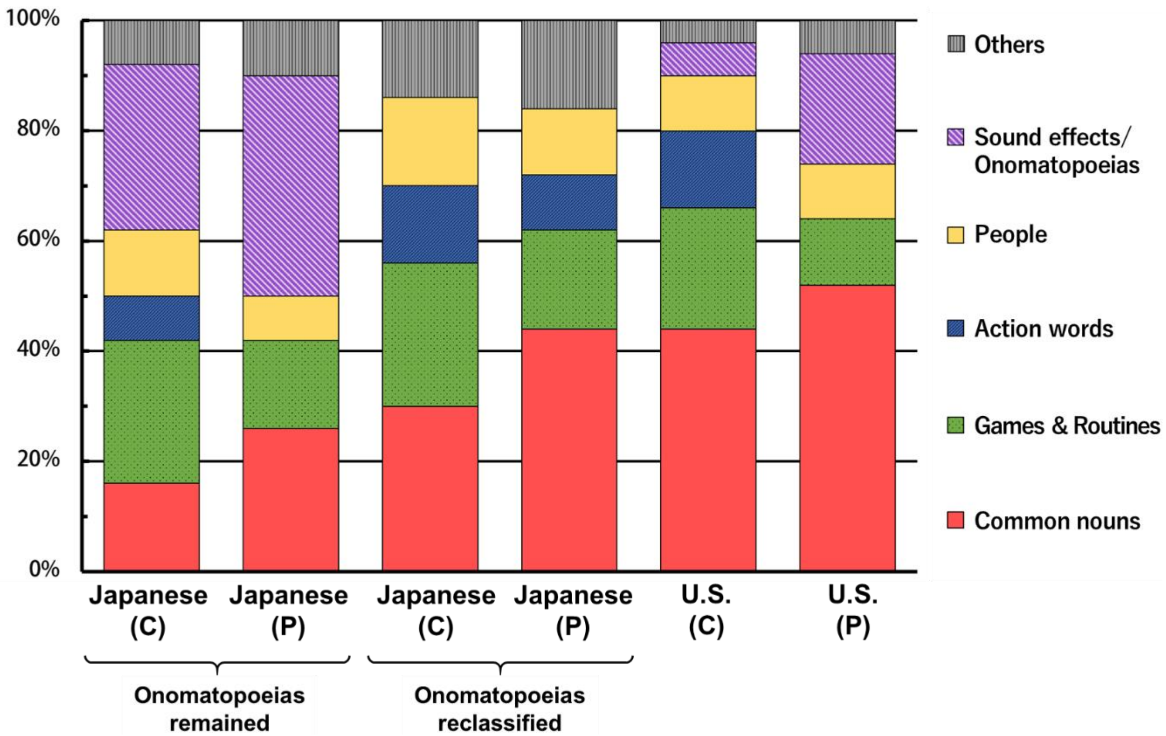

Figure 1.1 Classification of the 50 early words acquired by young children in Japan and the U.S. This figure was made based on Ogura (2007) with permission from The Linguistic Society of Japan. From the standardized data of the MacArthur-Bates Communicative Development Inventories (CDI) in Japan and the U.S., words were arranged according to the age in months in which the occurrence rate exceeded 50 % for each of the comprehension (C) and production (P) vocabularies. The earliest 50 words were extracted and classified. For early words in Japan, onomatopoeic words (i.e., baby talk) accounted for 30% and 40% of word comprehension and production, respectively (the left two bar graphs). However, when such onomatopoeic words were reclassified according to their semantic categories, common nouns had the highest percentage of word comprehension (30%) and production (44%) (the middle two bar graphs). The U.S. data showed a relatively larger proportion of common nouns for both comprehension (44%) and production (52%) vocabularies than the Japanese data (the right two bar graphs). The age at which children learn 50 words was several months earlier for the U.S. children for both comprehension (8–13 months) and production (12–18 months) compared to Japanese children (10–15 months for comprehension; 15–21 months for production).

their referents vary cross-linguistically and are more variable in lexicalizing relationships between objects than in lexicalizing objects themselves (Gentner, 1982; Gentner & Boroditsky, 2001). For example, concrete objects (e.g., cup) have coherent spatial boundaries and are easily perceived, while actions (e.g., kicking) have fluid temporal boundaries, and how actions are segregated from entire events depends on the language.

Recently, Maguire, Hirsh-Pasek, and Golinkoff (2006) proposed the SICI continuum to provide a unified account of word learning of both nouns and verbs because the difficulty of word learning varied not only across, but also within, parts of speech. For example, the difficulty of word acquisition varies from “spoon” (easy) to “peace” (hard) for nouns and from “jumping” (easy) to “believing” (hard) for verbs. SICI is an acronym for Shape, Individuation, Concreteness, and Imageability. According to this theory, these cardinal factors characterize the degree of difficulty in learning individual words. Maguire et al. (2006) pointed out that concrete nouns are easier to learn than action verbs for children because the former is explicitly shaped, easily individuated, highly manipulable, and easily connected to a certain mental image while the latter is not.

These theories seem to accurately explain why nouns (or object words) are learned earlier and with more ease compared to verbs (or action words) during early language development. However, children do not merely connect sounds to the very object or action in front of them when learning words. Even when learning concrete nouns, children have to determine the correct referent of a word they hear and must generalize word meanings according to certain rules[4]. In the next section, we briefly review the theoretical and empirical explanations of how children achieve these complex processes during word learning, specifically when learning object words.

[4] Besides semantics, there are other things that children have to deal with when learning a spoken language. For example, children have to distinguish linguistic sounds from environmental ones, segment sound patterns that make up a sentence, and find which sound patterns in the sentence correspond to a certain word. They also have to abstract sound patterns from what they hear so that they can understand that the words uttered by another person in another place are the same words they have heard before. Generally, these phonetical skills are learned earlier than semantic or grammatical skills even though infants already know several words at that time (Bergelson & Swingley, 2012). Children around 7- or 8-month-olds can segment words from continuous speech (Estes & Lew-Williams, 2015; Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999) and such word segmentation becomes more robust to changes in speakers or speakers’ emotions around the age of 10 months (Houston & Jusczyk, 2000; Schmale & Seidl, 2009; Singh, Morgan, & White, 2004). Research on word segmentation was a precursor to studies on infant statistical learning in language development (a landmark study is Saffran, Aslin, & Newport, 1996; see also Black & Bergmann, 2017; Saffran & Kirkham, 2018; Sandhofer & Schonberg, 2020).

1.3 How do young children learn object words?

1.3.1 Different accounts for word learning mechanisms

Regardless of what types of words are learned, it is difficult for a young child to determine the referent of a novel label she hears. When a child first hears the word “shoe,” there can be an almost infinite number of candidates for its meaning. For example, it may mean an object (e.g., a solid thing with a hole in the top), a part of the object (e.g., strings), an action (e.g., putting on), an attribute (e.g., soft to the touch or red color), etc. In the field of language development research, the difficulty of inductively inferring the correct referent of a word in an uncertain situation is known as “Quine’s problem” or the “Gavagai problem,”[5] being named after an American philosopher Quine who theoretically raised this problem (Quine, 1960).

To overcome this referential indeterminacy problem and exhaustively explain children’s early word learning, several different mechanisms have been proposed, although they remain controversial (Twomey & Hilton, 2020). Roughly speaking, some researchers suggest that children have innate mechanisms for language acquisition (e.g., Markman, 1989), while others argue that word learning can be explained by a simple associative learning theory (e.g., Samuelson & Smith, 1998), and others claim that socio-pragmatic factors play a crucial role in word learning (e.g., Tomasello, 2003).

As a theoretician who adopted the innate mechanisms approach, Markman (1989, 1992, 1994) proposed the “constraints hypothesis.” This hypothesis states that when young children infer the meaning of a novel word, its potential meaning is limited or biased by the very early phase of language development. In other words, children are constrained to narrow down the candidates of possible word meanings to a few plausible ones by default, enabling them to start with good first guesses when solving the problem of referential indeterminacy. Markman argued for three principle constraints: children likely infer that a novel word refers to an object as a whole rather than to its parts, color, or other properties (the “whole-object assumption”); children tend to avoid two labels for the same object (the “mutual exclusivity assumption”); and children likely extend word referents to objects of a similar kind rather than thematically related ones (the “taxonomic assumption”). These three

[5] “Gavagai” is an example of an unknown word used by Quine (1960). Before facing this problem, children also have to discover that linguistic sounds are symbols that indicate something, and that humans use linguistic sounds to communicate, but we don’t go any further into those prerequisites here.

constraints should enable children to make efficient and rapid inferences of word meanings despite being provided with few examples of the referent (Haryu, 2006; Heibeck & Markman, 1987), which is referred to as "fast mapping" (Carey, 1978; Carey & Bartlett, 1978). Yet, because the term "constraint" gives the impression that these skills are innate, inflexible, and have an all-or-none characteristic, the term "principles" is often used instead of constraints to reflect that these biases are developmentally flexible and more similar to heuristics (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Golinkoff, Mervis, & Hirsh-Pasek, 1994).

Researchers with the view that word learning can be best explained by simple associative learning, as seen in other cognitive domains, claim that constraints or biases for word learning can arise as a product of learning from previous experience. For example, unlike the top-down taxonomic assumption, young children could first generalize a single label to several objects based on their perceptual similarities, such as shape similarity (Imai, Gentner, & Uchida, 1994; Landau, Smith, & Jones, 1988). For instance, although Markman and Hutchinson (1984) showed that young children classified taxonomically related objects into one group (e.g., cups and glasses) rather than thematically related ones (e.g., cups and kettles) in a word-learning task, Landau et al. (1988) argued that this happened not because children had the abstract taxonomic constraints but because they grouped objects by similarity in shape (i.e., cups and glasses are more similar in shape than are cups and kettles). This perception-based tendency for children to generalize the same labels to objects with similar shapes is known as the "shape bias". Numerous studies support the presence of this bias and have now demonstrated that the emergence of this bias is strongly related to early noun learning (Gershkoff-Stowe & Smith, 2004; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Yee, Jones, Smith, 2012). More recently, this approach has fueled research on cross-situational statistical learning (Chen & Yu, 2017; Scott & Fisher, 2012; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007, 2011; Yurovsky, Yu, & Smith, 2013) and has shown that even 12-month-olds can connect novel labels and referents based on co-occurrence probabilities (Smith & Yu, 2008).

Adopting the socio-pragmatic approach, Tomasello (2003) criticized both the constraints-based and associationist explanations. He argued that these accounts underestimated "the informational richness of the socio-interactive environment in which children learn language" (p. 87), including routines, social games, and other patterned cultural interactions. Well-structured and highly predictable characteristics embedded in social games (e.g., peek-a-boo) are thought to help children infer the meanings of uttered words (Bruner, 1983; Ratner & Bruner, 1978). The socio-pragmatic account emphasizes the role of children's social capacities (e.g., shared attention or understanding intention) in word learning[6] (Tomasello, 2003; Tomasello & Farrar, 1986), to explain that young children can learn other kinds of words as easily or early as object words in some situations (Benedict, 1979; Bloom, Tinker, & Margulis, 1993; Kauschke & Hofmeister, 2002; Nelson, 1995; Nelson, Hampson, & Shaw, 1993; Oviatt, 1980; Tomasello & Akhtrar, 1995), and that there are cultural differences in the constitution of early vocabulary[7] (Frank et al., in press). Similarly, the "natural pedagogy theory" assumes that children are sensitive to ostensive signals (e.g., eye contact and infant-directed speech) and utilize these cues efficiently for social learning (Csibra & Gergely, 2009, 2011). Since the effectiveness of ostensive signals on social learning cannot be explained by mere attention getting (Okumura, Kanakogi, Kobayashi, & Itakura, 2020), children are likely to use social signals as more salient cues for learning than other perceptually salient signals (see also Diesendruck, Markson, Akhtar, & Reudor, 2004).

Although there is currently no agreement on the mechanism that best explains the referential indeterminacy problem, we believe that these theories are complementary as they address different components of early word learning. Moreover, these theories seem to overlap and approach each other recently[8]. For instance, the constraints-based explanation allows for developmental changes in word learning principles (Golinkoff et al., 1992, 1994; Hollich et al., 2000); the associationists quantitively assess how children interact with

---

[6] More recently, Tomasello has used the term "shared intentionality," which refers to collaborative interactions in which participants share mental states, to account for both ontogeny and evolution (e.g., Moll & Tomasello, 2007; Tomasello & Carpenter, 2007; Tomasello, 2014; Tomasello & Rakoczy, 2003). He claims that humans, unlike other primates, have distinctive skills and motivations in cooperative tasks involving shared intentionality (e.g., shared goals, joint attention, and cooperative communication).

[7] Cultural differences in word-learning strategies were also reported, such that compared to English-speaking children, Spanish-speaking children showed less shape bias and produced less shape-based nouns despite equivalent vocabulary sizes (Hahn & Cantrell, 2012). There are also cultural differences in the caregivers' input to their children. While English-speaking caregivers tend to emphasize concrete nouns over verbs, Chinese-speaking caregivers likely emphasize verbs over nouns (Tardif, Shatz, & Naigles, 1997). Interestingly, visual attention when seeing dynamic scenes also differed culturally between young children in the U.S. and China although the evidence was not strong (Chen et al., 2015; Waxman et al., 2016). Cultural differences in adults were also observed in categorization (Ji, Zhang, & Nisbett, 2004), memory retrieval (Masuda & Nisbett, 2001), event segmentation (Swallow & Wang, 2020), and even fundamental visual search (Ueda et al., 2018).

[8] This might be a somewhat aggressive summarization.

physical and social environments (Pereira, Smith, & Yu, 2014; Suanda, Barnhart, Smith, & Yu, 2019; West & Iverson, 2017; Yamamoto, 2020; Yoshida & Smith, 2008 Yu & Smith, 2012); and the socio-pragmatic approach explains children's sensitivity to social cues as innate traits unique to humans (Csibra & Gergely, 2011; Tomasello & Rakoczy, 2003). What is clear, at least for now, is that various types of cues affect children's word learning (Christiansen & Monaghan, 2006; Frank, Tenenbaum, & Fernald, 2013; Golinkoff & Hirsh-Pasek, 2006; Hollich et al., 2000; Twomey & Hilton, 2020; Waxman & Gelman, 2009; Wildt, Rohlfing, & Scharlau, 2019; Yu & Ballard, 2007). However, the cues that are the most important for early word learning vary across theories. For example, Hollich et al. (2000) classified multiple cues into attentional (e.g., perceptual salience and temporal contiguity), social (e.g., social eye gaze and social context), and linguistic (e.g., grammar and prosody) cues. They argued that children proceed with word learning by combining and changing the weight of these factors, and that all of these cues play a crucial role in word learning[9].

1.3.2 Actions affect object word learning

For several decades, many researchers have considered that perceptually stable and static characteristics of objects play a central role in early noun learning (e.g., Gentner, 1982; Gentner & Boroditsky, 2001; Gershkoff-Stowe & Smith, 2004; Hollich et al., 2000; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith et al., 2002; Yee et al., 2012). However, the dynamic information relevant to objects, such as object-specific actions or object functions (e.g., "putting on" for shoes), is also important for learning both early nouns (Booth & Waxman, 2002; Kemler Nelson, 1995; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson, Russell, Duke, & Jones, 2000; Kobayashi, 1997, 1998; Nelson, 1973a; Ross, Nelson, Wetstone, & Tanouye, 1986) and object categories (Booth, 2006; Booth, Schuler, & Zajicek, 2010; Hernik & Csibra, 2009; Horst, Oakes, & Madole, 2005; Madole, Oakes, & Cohen, 1993; Perone & Oakes, 2006; Träuble & Pauen, 2007, 2011), although such actions are essentially expressed as verbs. For example, after seeing the experimenter perform a particular action using a novel object with a novel label,

[9] Although accepting the effects of both perceptual and social cues on word learning, Yurovsky and Frank (2017) claimed that developmental changes in children's use of such potential cues could not be explained by the naïve weighted cue combination but by the development of domain-general cognitive processes (e.g., attention, memory, inhibitory control).

two-year-old children generalized the same label to other objects based on object-specific actions despite differences in shape, and also categorized objects according to such actions even when labels were not presented (Kobayashi, 1997). Ware and Booth (2010) suggested that, for 17-month-olds, knowledge of object-specific actions contributes to the subsequent establishment of shape bias. Furthermore, connections between objects and actions are more easily learned than those between objects and words in 12-month-old infants (Deng & Sloutsky, 2015; Eiteljoerge, Adam, Elsner, & Mani, 2019b) and toddlers (Childers & Tomasello, 2002; Hahn & Gershkoff-Stowe, 2010). Similarly, approximately 1.5-year-olds are able to relate certain objects to symbolic gestures[10] (Namy, 2001; Namy & Waxman, 1998; Sheehan, Namy, & Mills, 2007; Suanda, Walton, Broesch, Kolkin, & Namy, 2013).

Observational descriptive research has also focused on the role of actions or object functions on early word learning and categorization (Acredolo & Goodwyn, 1985, 1988; Goodwyn & Acredolo, 1993; Iverson, Capirci & Caselli, 1994; McCune, 2008; Nelson, 1974; Werner & Kaplan, 1963). Werner and Kaplan (1963) claimed that young children often apply the same label to referents according to the manner in which they handle objects (p. 117). For example, they cited the case reported by Stern and Stern (1907) in which a child used a label "nose" for all objects that are capable of being pulled, including noses, handkerchiefs, and toes of boots (Stern & Stern, 1907, p. 26). Kobayashi (1992) also documented that a Japanese child first performed a throwing action, uttered "*pōn*" (*pōn* is a Japanese onomatopoeic expression for a parabolic motion of throwing an object), and then "*pōn-ten-no*" [a thing to throw] to refer to a ball, before the child began to utter the concrete noun "ball."

Despite conflicting evidence on the central role of object-specific actions over perceptual cues in object word and object-category learning (Capone & McGregor, 2005; Kingo & Krøjgaard, 2012; Landau, Smith, & Jones, 1998; Smith, Jones, & Landau, 1996), young children are likely able to leverage such actions in word learning and object categorization, at least in some situations (e.g., actions are particularly emphasized). Children's learning strategies may also be susceptible to differences in the relative saliency between static shape and dynamic action cues (Hammer & Diesendruck, 2005). Recent studies have shown that teaching a novel word while moving a target object in a particular

[10] Interestingly, 15-month-olds (Puccini & Liszkowski, 2012) and 26-month-olds (Namy & Waxman, 1998; Suanda et al., 2013) failed to form the gesture-object connection although they could relate spoken words to certain objects, suggesting an inverse U-shape development of gesture comprehension as object words (see also Namy & Waxman, 2002).

manner (e.g., looming or shaking motions) affects how young children learn object words regardless of whether such motions are object-specific or non-object-specific (Goldstein, Elmlinger, & Schwade, 2020; Matatyaho-Bullaro, Gogate, Mason, Cadavid, & Abdel-Mottaleb, 2014). These particular motions are called the "infant-directed action" or "motionese" (Brand, Baldwin, & Ashburn, 2002), and caregivers abundantly use them as well as infant-directed speech in a word learning situation to facilitate learning (Gogate, Maganti, & Bahrick, 2015; van Schaik, Meyer, van Ham, & Hunnius, 2020). Synchronization or consistent co-occurrence of naming and such motions also contributes to early noun learning (Eiteljoerge, Adam, Elsner, & Mani, 2019a; Matatyaho & Gogate, 2008).

It may seem obvious to adult speakers or researchers that early nouns such as "shoe" mean specific object categories. However, considering the evidence that object-specific actions and simple motions affect learning object words and object categories, even though these actions and motions are conveyed by verbs, it is necessary to rethink the assumption that early nouns such as "shoe" correspond to specific object categories. The initial meanings of early nouns may not be as clear-cut and solid as previously assumed and may rather be vaguer and context-dependent.

1.4 Distinctive characteristics of children's early words

The idea that early nouns do not always correspond to specific object categories is not novel. Based on theoretical and observational research, some researchers have reported that early words have distinct label-meaning connections compared to adult words, as early words are not easily classified into ordinary parts of speech, such as nouns or verbs[11] (Church, 1961; Dewey, 1894; Lewis, 1936; McCune, 2008; Okamoto, 1982; Stevenson, 1893; Tomasello, 2003; Volterra et al., 1979; Vygotsky, 1934/1986; Werner & Kaplan, 1963). Dewey (1894) argued that an early word should be regarded as a complex, such as an "nominal-adjectival-verbal," rather than a specific noun, adjective, or verb[12]. Yoshida (2006) also pointed out

[11] The distinctiveness of early words has primarily been reported within specific categories of objects and actions. For instance, it is known that the word "doggie" for young children not only means dogs but also "four-legged animals" or even cars, but this "overextension" (Clark, 1973) has been regarded to happen only within the same parts of speech (e.g., nouns).

[12] Relatedly, Church (1961) argued that children's initial adjective "hot" may correspond to not only a single attribute but also hot things such as soup, stoves, and radiators (p. 63).

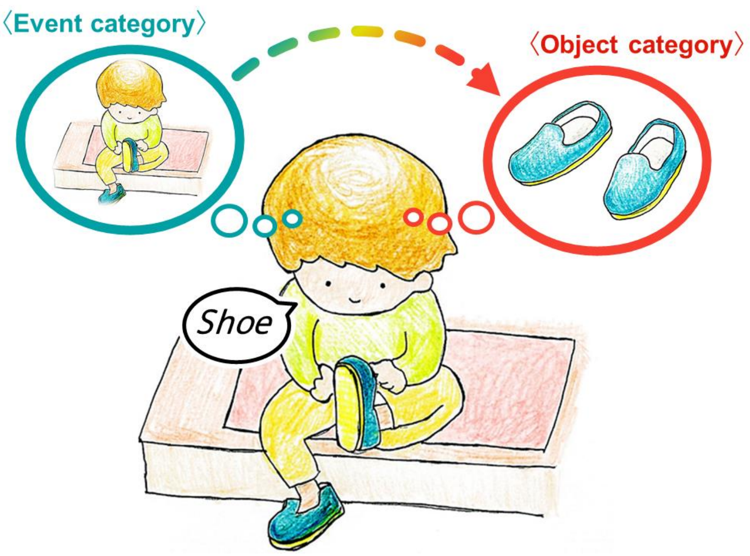

Figure 1.2 A schematic example of a developmental change in the meaning of the word "shoe" according to the hypothesis of Werner and Kaplan (1963). The initial meaning of the word "shoe" is the undifferentiated event category related to shoes, comprising of various components, such as objects and object-specific actions (e.g., "putting on"). Subsequently, the word differentiates into a specific object category, becoming a specific "object word".

the possibility that even nouns, which serve as a scaffold for subsequent word learning, have a vaguer meaning at first, such that they are entwined with correlational relationships of multiple factors around objects.

Werner and Kaplan (1963) posited in their organismic-developmental approach that children's initial words correspond to holistic, global, and undifferentiated events or activities that emerge from children's experiences within certain contexts. They stated that "the referents of early vocables remain relatively global in character, that is, total situations in which agent, action, and object are intimately fused"[13] (p. 116). For example, when an infant utters "shoe," its meaning does not only include the shoe but also the action of putting the shoe on and off, the infant as an agent, and the location related to shoes (**Figure 1.2**). Likewise, when an infant utters "boo" in a peek-a-boo play context, its meaning does not only include the disappearance or reappearance of a toy but also the toy, the caregiver, and

[13] Werner and Kaplan (1963) used the term "vocable" to refer to a phonetic aspect of words that is contrasted to a semantic aspect of words. In this thesis, we use the similar terms of "label" or "word form" for easy understanding.

the child's feelings when playing (see also McCune, 2008). Importantly, Werner and Kaplan (1963) stated that children's early words do not mean "precisely delimited components such as action per se or thing per se" (p. 137) during a child's one-word period, even when their word forms are the same as ordinary words, such as nouns or verbs. They suggested that it is only after children begin to combine two or more words (e.g., "papa boo"), or use the same word with different intonations, that adults can assume that the meanings of children's words semantically differentiate from holistic event categories into specific categories of objects, actions, etc.

It should be noted that the holistic nature of early words does not mean that young children have an implicit complete sentence in mind, although they produce only one word at a time. Strictly speaking, Werner and Kaplan (1963) claimed that early label-meaning connections were "neither true words nor true sentences" (p. 384) because children's early utterances do not have circumscribed meanings (e.g., objects or actions) or syntactic-grammatical characteristics (e.g., nouns or verbs) and because such utterances are not made of articulated and differentiated speech units (i.e., words). Hence, they avoided describing children's one-word utterances as "one-word sentences" or "holophrases" (e.g., Stevenson, 1893). Instead, they referred to these utterances as "names" or "monoremes" and argued that there were several steps in which monoremes could lead to true elaborate words.

Nelson (1983a, 1983b, 1986) claimed that children's initial concepts are events in which objects and actions are embedded as a context-bounded whole, a view compatible with that of Werner and Kaplan (1963). She emphasized that "objects are embedded in events, and they may have no privileged status therein" (Nelson, 1983b, p. 134). According to this hypothesis, object categories formed in children's minds are not one of the most fundamental blocks that build up later more complex concepts but are already "conceptual achievements" that are constructed by the abstraction or decontextualization of objects from experience. For young children, for example, "ball *implies* throwing, book *implies* reading (or chewing), and cup *implies* drinking" (Nelson, 1983b, p. 136), and objects themselves are not sufficiently individuated as part constructions. Similar to Werner and Kaplan (1963), Nelson and Lucariello (1985) pointed out that words have different meanings depending on the stage of development. Strictly speaking, words that seem to be object words do not signify objects until children can conceptually partition events as a whole into objects.

The theoretical assumption that the initial meanings of children's words are undifferentiated event categories tied to particular contexts may be supported by the observation that children's early vocabulary consists not only of object words but also of context-related words referring to places, activities, or actions such as "park," "lunch," or "kiss" (Nelson et al., 1993). In addition, Nelson (1983a) cites the case reported by Church (1966) in which a 12-month-old child went to the bathroom, took her clothes off, turned on the water, and so forth, as a response to the word "bath"[14] (Church, 1966, p. 51). The intuition that early words fuse objects and actions is further supported by the presence of "dual-category words" (i.e., words that can be used as both nouns and verbs) in children's early vocabulary (e.g., "drink," "call"; Nelson, 1995; see also Tomasello, 2002). Moreover, Roy, Frank, DeCamp, Miller, & Roy (2015) recently found that words uttered by caregivers in distinctive contexts, such as in spatial or temporal aspects, were learned earlier independently of the grammatical form of those words.

However, the observational approach alone cannot provide direct evidence as to whether the initial meanings of words are undifferentiated events comprising of both objects and actions, as posited by Werner and Kaplan (1963), or if they correspond to a specific object or action categories from the very beginning of learning. This is because even if children's words correspond to holistic event categories, uttered labels are not phonetically distinguishable from those corresponding to specific object categories. To our knowledge, no study has experimentally investigated the hypothesis of Werner and Kaplan (1963) on semantic differentiation of early words. We speculate that this is because their theoretical approach is mostly forgotten in the current developmental science due to their idiosyncrasy and to considerable variance from the contemporary consensus (Glick, 1992).

Before moving on to our experimental approach, we discuss two kinds of recent findings that provide some evidence, though not direct, for Werner and Kaplan's (1963) hypothesis, which serve as the basis for our experiment. The first is on young children's nonlinguistic ability to segment events. Some might think that forming event categories should be much more difficult for infants than forming object categories because events, which include both objects and actions, are more temporarily fluid and difficult to isolate from environments compared to objects (Gentner, 1982; Gentner & Boroditsky, 2001; Imai & Haryu, 2014). However, recent experimental studies have revealed that children have some event knowledge in toddlerhood (Friend & Pace, 2011, 2016; Göksun et al., 2011;

[14] Strictly speaking, it is unclear whether the child spontaneously responded to the word "bath" by herself and executed such a systematic series of actions because Church (1966) did not report this observation in detail and he wrote that an adult "helped" the child do these actions. However, it is possible that the child spontaneously took such actions because the author observed a similar phenomenon at a Japanese nursery school. That is, a one-year-old child responded to the word "*Gohan*" [food (or eating)] by having a seat and excitedly looking at a teacher.

Sonne, Kingo, & Krøjgaard, 2016) and in prelinguistic infancy (Baldwin, Baird, Saylor, & Clark, 2001; Hespos, Saylor, & Grossman, 2009, 2010; Saylor, Baldwin, Baird, & LaBounty, 2007; Song, Pruden, Golinkoff, & Hirsh-Pasek, 2016). In fact, 10-month-olds can detect temporal boundaries of an event from a continuous and complex flow of scenes in which an adult conducts everyday actions (Baldwin et al., 2001). Thus, even prelinguistic children likely segment events, which possibly facilitates the formation of holistic event categories. A larger body of research on young children's event segmentation would provide important insights for the basis of all word learning, including nouns and verbs. Yet, studies on nonlinguistic event knowledge and word learning have primarily been carried out in the context of verb learning (Aktan-Erciyes & Göksun, 2019; Göksun, Hirsh-Pasek, & Golinkoff, 2010; Konishi, Stahl, Golinkoff, & Hirsh-Pasek, 2016; Maguire & Dove, 2008).

The second finding describes the effect of context on object word learning. Young children form label-meaning connections differently depending on the contextual background (Chen & Yu, 2017; Goldenberg & Sandhofer, 2013; Perry, Samuelson, & Burdinie, 2014; Twomey, Ma, & Westermann, 2018; Vlach & Sandhofer, 2011), and a similar phenomenon occurs for object categorization (Goldenberg & Johnson, 2015). For instance, unlike older children, 2.5-year-olds show lower performance on novel object word learning when the contextual background varies than when it is consistent from the learning to the test phase (Vlach & Sandhofer, 2011; but see Tippenhauer & Saylor, 2019). Relatedly, toddlers can form object categories based not only on similarities in shape or function but also on contextual spatiotemporal relations such as "kitchen-related objects" (Mandler, Fivush, & Reznick, 1987; see also Roy et al., 2015). Sandhofer and Schonberg (2020) suggest that contextual cues that co-occur with target objects could be beneficial for aggregating discrete examples together in memory [15] and for forming label-object associations and object categories, especially early in the learning process.

Based on various observational descriptions about children's language use, Werner and Kaplan (1963) argued that the initial meanings of words are undifferentiated event categories tied with contexts or situations, even if they are object words. This view is

[15] Even motions, which are considered more fundamental than other developmental domains, are susceptive to contextual differences. Butler and Rovee-Collier (1989) showed that one day after learning to move a crib mobile by kicking, 3-month-old infants could retain information of causality (kicking causes a crib mobile to move) even in a different context; however, they failed to retrieve it 3 and 5 days after training despite only experiencing a change in context (appearance of a crib bumper; see also Thelen & Smith, 1994, p. 205). The effects of the context variability and delay have also been investigated in research on word learning (Werchan & Gómez, 2014; Wojcik, 2017).

increasingly supported by experimental evidence, as children can segment events in the prelinguistic stage, and children's early word and category learning are susceptible to contextual information. However, it is still unclear whether early nouns initially correspond to global event categories comprising of at least both objects and actions, as Werner and Kaplan (1963) posited, or if such word forms correspond to specific object categories from the very beginning. In order to experimentally address this question, we propose a new hypothesis on the flexibility and plasticity of early word learning: the semantic pluripotency hypothesis.

1.5 The Semantic Pluripotency hypothesis: A new perspective on early word learning

We propose the "semantic pluripotency hypothesis": a revision of Werner and Kaplan's (1963) theoretical hypothesis towards an experimentally verifiable hypothesis. This hypothesis consists of two sub-hypotheses: First, the initial meanings of children's words are holistic, global event categories that correspond to the fusion of various factors related to the word. Word meanings in turn have distinctive pluripotent characteristics. Second, the initial meanings of words subsequently differentiate into specific, discrete, and more decontextualized categories such as objects, actions, or attributes with development. In this sense, the development of early word meanings is plastic. Thus, we posit that although the surface of children's words (i.e., word forms) may appear to be the same as adults' ordinary words, the semantic contents of words dynamically change with development from holistic to specific categories, similar to stem cells that differentiate into specialized muscle, bone, or nerve cells (**Figure 1.3**).

We are aware that the term "event" has several definitions and that there is still controversy regarding this term (Pruden, Hirsh-Pasek, & Golinkoff, 2008). For instance, Kable, Lease-Spellmeyer, and Chatterjee (2002) divided fundamental categories into entities and events. Entities refer to things (e.g., people, animals, and objects) prototypically expressed by nouns, whereas events refer to "what happens to things," including actions and thematic relationships (who does what to whom) prototypically expressed by verbs and the syntactic location of nouns, respectively. Similarly, Baillargeon and Wang (2002) described events as having a more specific meaning, such as hiding objects behind an occluder (i.e., physical events), and they also contrasted event categories with object categories. On the other hand, Werner and Kaplan (1963) used the term "event" in a much broader sense, not as a component that contrasts with things, but as a whole situation that includes both objects

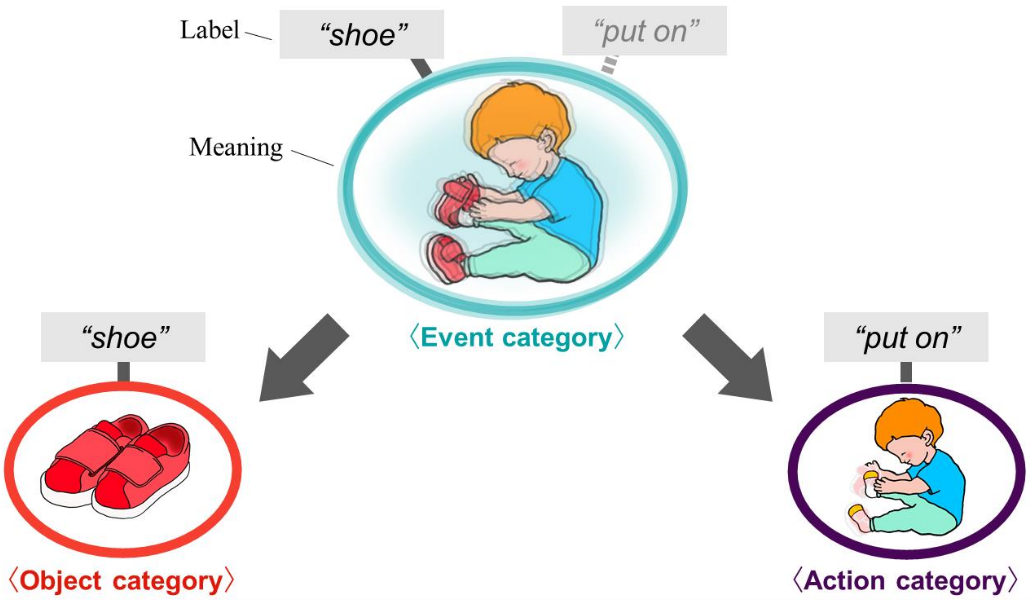

Figure 1.3 A schematic view of the “semantic pluripotency hypothesis.” The initial meanings of children’s words are undifferentiated event categories comprising of objects, actions, etc. as a context-bounded fusion emerging from children’s experiences. With development, word meanings differentiate into specific categories and become independent from each other. For the sake of simplicity, only the semantic differentiation of object and action words are described in the illustration.

and actions. They even regarded events as situations in which objects and actions as well as other components, such as children’s sensorimotor experiences and attitudes, are embedded. Nelson (1986) seemed to have a similar view by suggesting that events “incorporate objects and relations in a larger whole” (p. 3).

In the first step of verifying the semantic pluripotency hypothesis, we regard event categories as the undifferentiated fusion of objects and actions. In this thesis, we primarily focus on the meanings of early common nouns among various parts of speech because common nouns (or object words) have been considered fundamental to word learning, and because most studies on word comprehension have been conducted with nouns (Nomikou, Rohlfing, Cimiano, & Mandler, 2019; Tomasello, 2003). For decades, object words have been conventionally regarded as meaning specific object categories during early word acquisition. However, we hypothesize that the initial meanings of such words are undifferentiated event categories that subsequently differentiate into specific object categories. If our findings corroborate the semantic pluripotency hypothesis, we not only offer empirical support for Werner and Kaplan’s (1963) theory but additionally contribute to ongoing linguistic debates and set the stage for future research on children’s early word learning.

For example, the semantic pluripotency hypothesis will provide a new explanation for the long-standing controversy on the developmental order of nouns and verbs. As discussed above, children learn verbs earlier than nouns in some languages (e.g., Frank et al., in press; Tardif, 1996). If the initial meanings of children's words are undifferentiated event categories, early verbs in such languages may also correspond to undifferentiated categories during the initial stage of word learning, even when labels usually mean actions. If so, the time at which certain parts of speech (e.g., nouns or verbs) are learned can be attributed to phonetic and/or contextual consistency or invariance instead of the ease at which the referent is identified. Another potential insight provided by the semantic pluripotency hypothesis is its contribution to the integration of interrelated, but separately explored, research topics such as event categorization, contextual effects on word learning, and statistical learning. Considering that the semantic pluripotency hypothesis may enable the construction of a unified model of semantic differentiation from global to specific categories, which accounts for early word learning regardless of object or action words. Thus, the semantic pluripotency hypothesis will shed light on children's early label-meaning connections, such as semantic flexibility and plasticity with development, which differ from those for adults and older children. Note that we will revisit our hypothesis in Chapter 5 and further refine it based on our findings presented in the following chapters.

1.6 Outline of the thesis

The central purpose of this thesis is to experimentally investigate the semantic pluripotency hypothesis with a focus on common nouns. We will present the initial meanings of object words and their developmental changes. This thesis consists of five chapters (**Figure 1.4**), and the following chapters are outlined below.

Chapter 2 introduces the initial meanings of object words and whether they change with development. Using a two-alternative forced-choice pointing task, we investigated toddlers' semantic comprehension of familiar object words (e.g., "shoe") at two levels: first, as global event categories that included both objects and actions and second, as specific object categories that are independent from object-specific actions. We selected a wide age range, from 19 to 35-months, to observe potential changes in object word meanings throughout development. As we will see later, our results support the semantic pluripotency hypothesis, although they are preliminary. Since we found that semantic differentiation of object words occurred around the 21-month-old mark, we wanted to replicate our findings

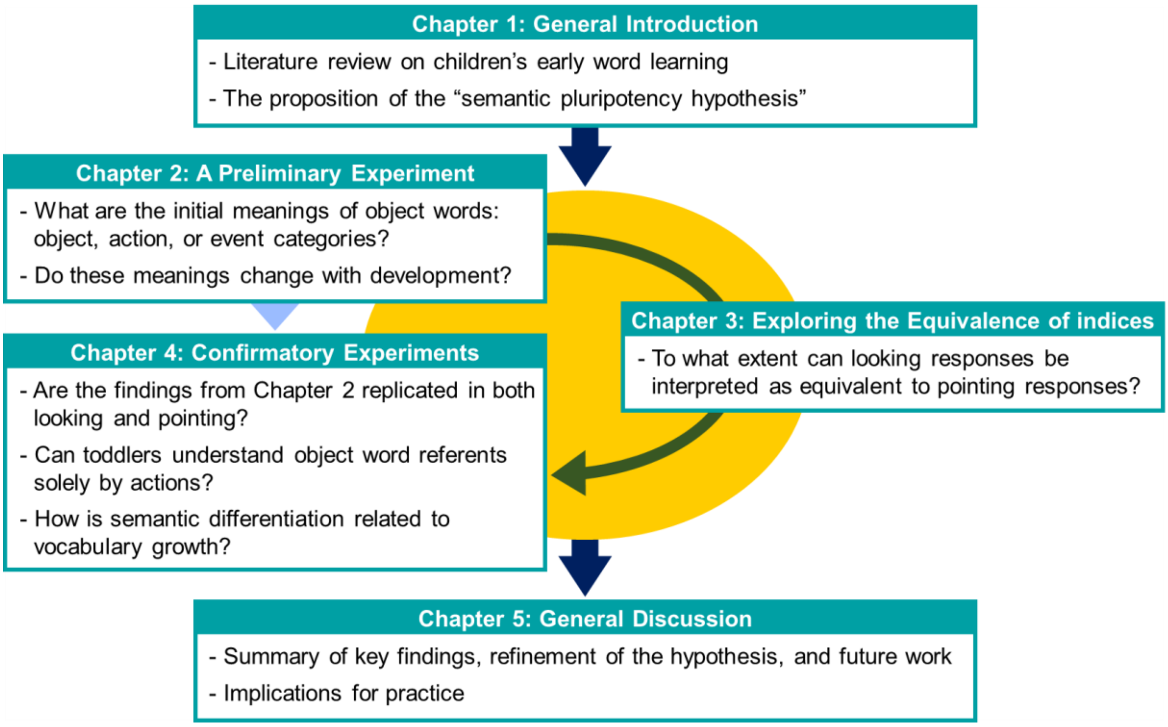

Figure 1.4 The thesis structure and research questions addressed in each chapter.

in more children around this critical period. However, since toddlers younger than two years old often did not provide clear pointing responses, they did appear to spend more time looking at one of the stimuli presented simultaneously. Therefore, it was necessary to measure both the pointing and looking responses in the following study to reduce data loss.

Chapter 3 describes whether looking and pointing responses are equivalent in forced-choice tasks. We investigated the extent to which pointing responses (i.e., left or right) were predicted from concurrent preferential looking using data from toddlers at 18–23 months with both pointing and looking indices. Through this chapter, we find that looking responses are reasonable alternative indices to pointing, even in light of some qualitative differences. Although Chapter 3 does not directly address the developmental change in object word meanings, this chapter plays a critical role in bridging Chapters 2 and 4 by confirming the equivalence of two different measurements of task performance.

Chapter 4 describes and discusses the experiments that test the semantic pluripotency hypothesis with the methods developed in Chapter 3. Using both cross-sectional and longitudinal data of 18–23 month-old toddlers, we replicated the results obtained in Chapter 2. In addition, we investigated whether toddlers could understand object word referents solely based on object-specific actions. We further looked into how developmental changes in object word meanings are related to concurrent and later vocabulary growth of common nouns and verbs.

Finally, in Chapter 5, we summarize the key findings, refine the semantic pluripotency hypothesis with the use of a dynamic systems approach (Thelen & Smith 1994), and discuss future directions. As we will later see, the dynamic systems approach contributes to our understanding of the dynamic process of the semantic differentiation of words from global to specific categories, enabling the construction of a word learning model. Finally, we discuss the implications of the semantic pluripotency hypothesis for caregivers, educators, and clinicians.

Chapter 2

Initial Noun Meanings Do Not Differentiate into Object Categories: A Preliminary Study

Based on:

Hagihara, H., & Sakagami, M. (2020). Initial noun meanings do not differentiate into object categories: An experimental approach to Werner and Kaplan's hypothesis. *Journal of Experimental Child Psychology*, *190*, 104710. https://doi.org/10.1016/j.jecp.2019.104710

In this chapter, we preliminarily explore the initial meanings of object words and their developmental changes using a two-alternative forced-choice pointing task. After briefly summarizing previous findings on children's early vocabulary and early word learning, we look into the first experiment, investigating the semantic pluripotency hypothesis.

2.1 Introduction

Observational studies of most languages have revealed that nouns are acquired earlier than other parts of speech (e.g., verbs) during initial development (Bornstein et al., 2004; Caselli, Casadio, & Bates, 1999; Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, in press; Gentner & Boroditsky, 2001). Conversely, for some verb-friendly languages such as Mandarin, verbs are relatively equivalent to nouns (or even dominant) in early vocabulary (Brown, 1998; Choi & Gopnik, 1995; Tardif, 1996; Tardif et al., 2008; Tardif, Gelman, & Xu, 1999). Research on the developmental order in which nouns and verbs are learned has been conducted over many decades, and has elucidated the language-general and language-specific features of early word learning (e.g., Gentner, 1982; Gentner & Boroditsky, 2001; Gleitman & Gleitman, 1992; Gogate & Hollich, 2016; Goldin-Meadow, Seligman, &

Gelman, 1976; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011; Twomey & Hilton, 2020; Waxman et al., 2013). Further, especially in English, most words are mainly divided into two classes based on their grammatical and logical properties: nouns and verbs (Whorf, 1940/1956). This has resulted in comparisons of word learning between nouns and verbs. Early learned nouns generally correspond to "object categories," while early verbs correspond to "action categories" (Gogate & Hollich, 2016). Researchers have thus investigated cognitive mechanisms in infancy and toddlerhood, which would explain why object words are advantageous for learning compared to verbs in most languages.

For example, Gentner (1982) argued that children learn object words before action words since objects are more easily segregated from environmental surroundings than other relational categories (such as actions); this is known as the "natural partitions hypothesis" (see also Gentner & Boroditsky, 2001). Markman (1989, 1992, 1994) proposed word-learning constraints or biases by which children can narrow down the possible referents of the words they hear. Although it is controversial whether such biases are innate or learned, one of the most well-studied biases is the "shape bias," whereby children tend to generalize the same word for similarly shaped objects (Gershkoff-Stowe & Smith, 2004; Graham & Poulin-Dubois, 1999; Imai, Gentner, & Uchida, 1994; Kucker et al., 2019; Landau, Smith, & Jones, 1988, 1998; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Yee, Jones, Smith, 2012). Maguire, Hirsh-Pasek, and Golinkoff (2006) proposed the SICI continuum, whereby the difficulty in word learning is prescribed by the degree to which the word references four factors (i.e., shape, individuality, concreteness, and imageability). According to this hypothesis, early nouns are easier to learn than early verbs because, relatively speaking, all four factors in the former are usually easier compared to the latter.

These theoretical hypotheses implicitly assume that early noun meanings correspond to specific object categories characterized by static and perceptual features (e.g., shape). However, information on dynamic and conventional actions specific to objects (i.e., object functions) is also essential for early noun learning (Booth & Waxman, 2002; Kemler Nelson, 1995; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson, Russell, Duke, & Jones, 2000; Kobayashi, 1997, 1998; Nelson, 1973a; Ross, Nelson, Wetstone, & Tanouye, 1986) and even object category learning (Booth, 2006; Booth, Schuler, & Zajicek, 2010; Hernik & Csibra, 2009; Horst, Oakes, & Madole, 2005; Madole, Oakes, & Cohen, 1993; Perone & Oakes, 2006; Träuble & Pauen, 2007, 2011) despite that such actions are normally conveyed by verbs (e.g., the function of "putting on" creates the object category or object word of "shoes"). Ware and Booth (2010) demonstrated that knowledge about object-specific actions facilitates the later formation of shape bias for children aged 17 months. These actions also guide children to focus on particular properties when learning the names of unfamiliar objects (Kobayashi, 1998). This critical role of object-specific actions in object word and object category learning implies that initial noun meanings might not be restricted to static object categories alone.

Werner and Kaplan (1963) proposed a unique theoretical hypothesis regarding early language development. Their organismic-developmental approach explicitly argues that words' semantic content is not the same between adult speakers and young children who have recently begun to utter word-like vocables. Based on various observational studies of children's language development, they posited that initial word meanings are characteristically holistic and global event categories, and are not sufficiently differentiated into specific categories of objects, actions, and so on. For example, when a child utters "boo" when playing peek-a-boo, its meaning represents an undifferentiated event of the playful activity as a whole, rather than a strictly delimited movement of disappearance or reappearance, unless the child has uttered two words (e.g., "papa boo") in order to refer to the object and the action separately (see also McCune, 2008). Therefore, at least during the single-word period, children speak only one word at a time not because they have to isolate extremely partial aspects of what they want to express (due to the limitation of their productive vocabulary size), but rather because such a single word is holophrastic and encompasses their understanding of a global event. This view is also supported by Nelson (1983a, 1983b, 1986), who suggested that events as a whole are the first concepts learned by children. Nelson (1995) also claimed that children's early vocabulary includes several words that could be used both as nouns and verbs, such as "drink" (see also Tomasello, 2002). The theoretical view of Werner and Kaplan (1963) is extremely innovative in that it explains the emergence and differentiation of early words without assuming correspondences such as noun-object or verb-action. However, their approach has generally been forgotten in current developmental science because of its idiosyncrasy and departure from more conventional approaches (Glick, 1992).

Unfortunately, no experimental studies have directly examined Werner and Kaplan's (1963) hypothesis that object words' initial meanings are undifferentiated events. The observational approach alone cannot directly address the semantic content of early words. Even if early word meanings are undifferentiated event categories, object words uttered by young children are indistinguishable from ordinary words because their phonetic features are the same. Hence, in this chapter, we preliminarily explore the hypothesis of Werner and Kaplan (1963) about the initial meanings of object words from an experimental angle. To achieve this aim, in Chapter 1, we proposed the "semantic pluripotency hypothesis" as a refinement of their theoretical conjecture into an experimentally verifiable one. The semantic pluripotency hypothesis is distinguished by two features: (1) initial word meanings are distinctively pluripotent, and (2) they are plastic, as they subsequently differentiate into specific categories along with development. As the first step in verifying this hypothesis, we assert that object words' initial meanings are undifferentiated global event categories comprising both objects and actions as a fusion (e.g., "putting shoes on"), and they subsequently differentiate into specific rigid object categories independent of actions (e.g., "shoes" alone) during a certain phase of language development (**Figure 2.1**).

In order to determine how young children comprehend object words, we adopted a two-alternative forced-choice pointing task; this method is less demanding for children as it does not require them to produce words or act something out to respond to the experimenter's utterance (Ambridge & Rowland, 2013). Using this technique, children watch two juxtaposed pictures or video stimuli and are asked to choose one of them that matches the experimenter's instruction by simply pointing. In addition to its less demanding nature, this approach has various advantages such that it can be easily administered without expensive specialized equipment, pointing responses can be easily coded, analyzed, and interpreted, and it is applicable to a broader age range of children and adults (see Noble, Rowland, & Pine, 2011).

Using a two-alternative forced-choice pointing task, we examined whether object words' meanings essentially differentiate into specific object categories throughout early language development, or are deeply influenced by object-specific actions at the beginning of this development. Most previous studies that have verified the effect of object-specific actions on word learning have focused on the latter half of the one- to two-year-old age range (e.g., Kobayashi, 1997; Kemler Nelson, Russell et al., 2000; Ware & Booth, 2010). In this preliminary study, we employed the same—but comparatively wide—age range so that we could detect developmental shifts in word meanings if such changes existed.

2.2 Method

2.2.1 Participants

We included 36 Japanese monolingual children aged 19 to 35 months in this study's final analysis (19 boys and 17 girls; M_{age} = 27.4 months, *SD* = 5.1; see **Figure 2.2** for demographic information). We recruited the participants from five nursery schools. We asked the parents of all participants to complete the Japanese MacArthur-Bates Communicative Development Inventory (J-MCDI; Ogura & Watamaki, 2004) to estimate

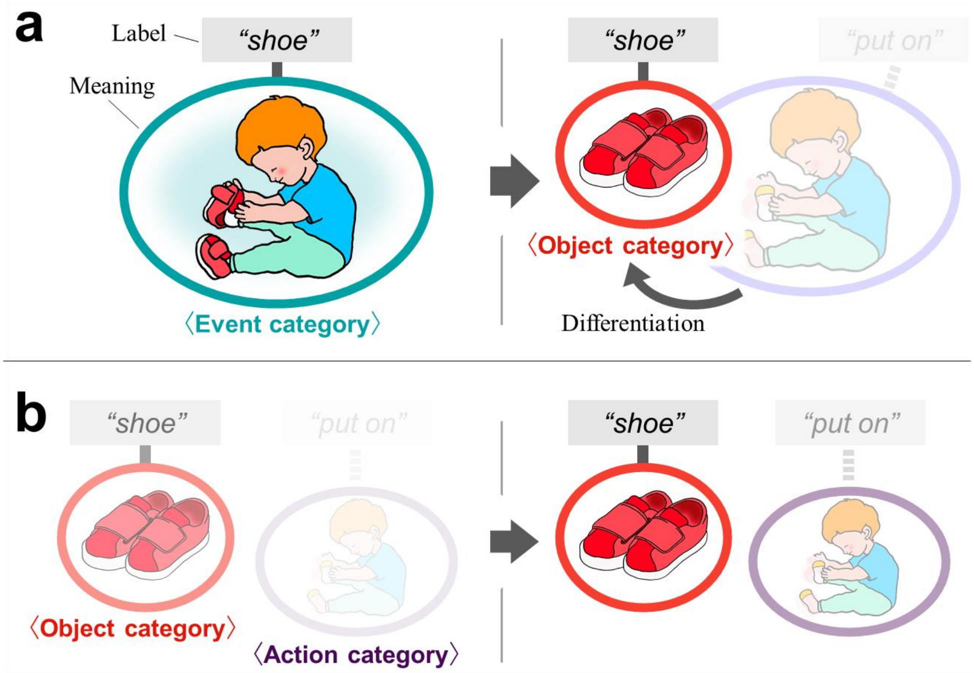

Figure 2.1 Schematic view of the hypothesis explored in this study. (a) The semantic pluripotency hypothesis. Object words' initial meanings are holistic event categories comprising both objects and actions, and subsequently differentiate into rigid object categories. (b) The conventional view of early object word learning. Object word meanings are essentially object categories from the very beginning of learning. This label-meaning connection becomes more precise and robust with development. The word–action connection (the acquisition of which is delayed compared with nouns in many languages) is described for comparison.

the participants' productive vocabulary size. We excluded another 16 children from data analysis because of difficulty in coding pointing responses ($n = 9$), being raised in a bilingual environment ($n = 2$), falling outside the age requirements ($n = 2$), experimental error ($n = 1$), difficulty completing all trials of the experiment ($n = 1$), and incomplete answers to the J-MCDI ($n = 1$). We obtained written consent from the parents of all participants in advance. All participants were thought to be developing typically, as none of them were receiving developmental support, such as medical treatment. This research was in accordance with the guidelines laid down in the Declaration of Helsinki and approved by the ethics committee for human and animal research at Kyoto University's Graduate School of Human and Environmental Studies.

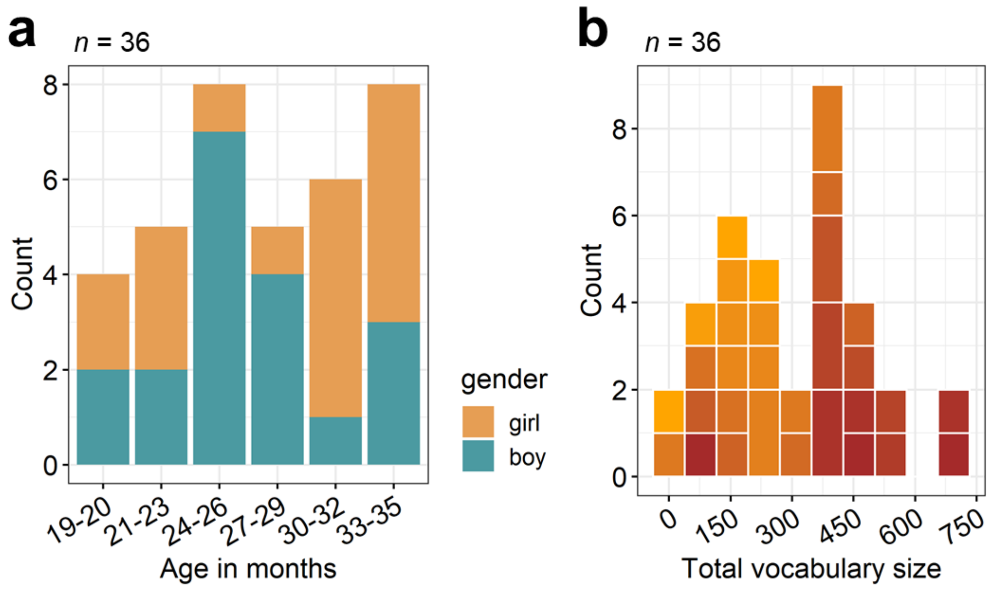

Figure 2.2 Demographic information on participants included in the final analysis. The histograms are presented for (a) age in months and (b) total vocabulary size.

2.2.2 Stimulus materials

For the two-alternative forced-choice task, we created short video stimuli in which a girl[16] performed specific actions with particular objects based on previous research using similar materials and methods (Imai, Haryu, & Okada, 2005). Each stimulus lasted for about seven seconds. We selected shoes and a cup as the target objects because the names of these objects are typically acquired during early language development (Ogura, Watamaki, & Inaba, 2016); these objects are familiar to toddlers in their everyday lives and can be easily used to assume specific, conventional actions. We thus assigned "putting on" and "drinking with" to the shoes and cup, respectively (We refer to these actions as "target-specific actions" hereafter). We then prepared filler objects with which target-specific actions could potentially be performed; these contained different shape features from the target objects (i.e., a pair of baskets and a toy shovel corresponded to the shoes and cup, respectively). We emphasized the differences in the shape features of filler objects by the following points:

[16] Instead of an adult, we chose an older child to be the person in the video stimuli based on the finding that young children prefer watching actions with objects performed by older children compared to actions carried out by peers and adults (Zmyj, Daum, Prinz, & Aschersleben, 2012).

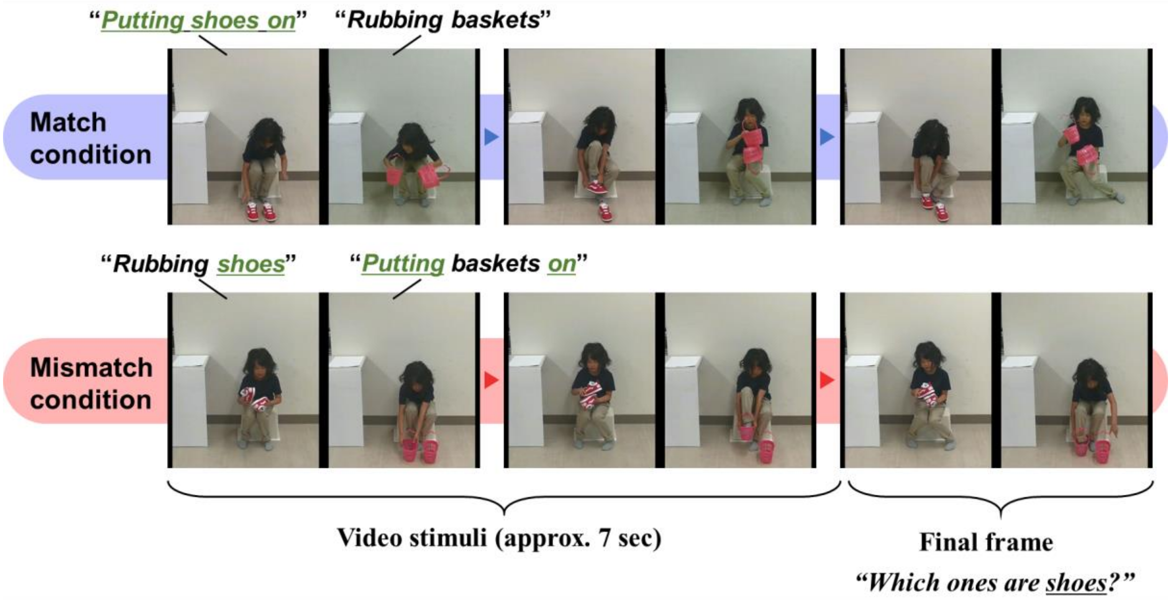

Figure 2.3 Example of the two-alternative forced-choice task. Each pair of stimuli was juxtaposed according to the condition (match or mismatch). The video stopped on the final frame and the experimenter asked which stimulus corresponded to the target object. The target objects and target-specific actions are shown in green (also underlined).

unlike the shoes, the baskets were symmetrical, with a vertically oriented handle; unlike the cup, the toy shovel had a flat, asymmetrical shape, while its handle was linear. We also assigned filler actions that were meaningless and different from the target-specific actions (e.g., "rubbing in front of self" and "making a circular motion on a table" for the shoes and cup, respectively).

We conducted the two-alternative forced-choice pointing task using a within-subjects design, employing the video stimuli according to two conditions (i.e., the match and mismatch conditions; **Figure 2.3**). For the match condition, a video displaying the target-specific actions utilizing the target objects was juxtaposed with a video of the filler actions using the filler objects (e.g., "putting shoes on" vs. "rubbing two baskets in front of her"). For the mismatch condition, a video portraying the filler actions using the target objects was juxtaposed with the target-specific actions using the filler objects (i.e., the target objects and target-specific actions were separated). We also prepared video stimuli based on these conditions, in which the presented sides were switched. These stimuli ended just before the target-specific actions were completed. There were eight test stimuli for each participant: two target objects (shoes and cup) used for two conditions (match and mismatch) and two presented sides.

2.2.3 Apparatus and procedure

All participants individually completed the tasks in vacant rooms at each nursery school. Each child was seated on a small chair or in the lap of a nursery school teacher. For the three children seated on a teacher's lap, the experimenter asked the teachers to refrain from performing task cues (e.g., any utterances related to the video stimuli) and to drop their eyes to avoid looking directly at the monitor. The experimenter first played with each child for approximately five minutes to build rapport. The experimenter than asked the child to look at the 23-inch screen (532 mm × 322 mm) displaying the video stimuli, with a viewing distance of approximately 60 cm. Each experiment required approximately 15 minutes to complete and was videotaped for later coding.

Warm-up trials

Prior to the test trials, the participants engaged in three warm-up trials to confirm that they could point to what they considered the correct side of paired stimuli. The experimenter showed the participants a juxtaposed picture or video pairs on the monitor, and encouraged them to point to either side according to the experimenter's question, for three trials. The first juxtaposition showed pictures of a stuffed dog and a toy car. The experimenter asked the following: "*Wanwan/Bubu wa docchi*?" [Which one is a dog/car?] (*Wanwan* is Japanese baby-talk for dogs, and *Bubu* is for cars). The experimenter said "*Sou dane*" [Yes, it is] when the participants pointed to the correct side, and corrected them by saying "*Wanwan/Bubu wa kocchi dane*" [The dog/car is on this side] when they pointed to the wrong side or failed to point at all. The second trial consisted of a picture of a toy car and a short video in which someone moved a stuffed dog. The third trial entailed short videos in which people moved a toy car and a stuffed dog. Here, the questions were the same as in the first trial. The experimenter repeated the warm-up trials if the children did not understand the task rules.

Test trials

The participants engaged in eight test trials each. The experimenter presented video stimuli after a presentation, showing a center gaze-point for 1.5 seconds accompanied by a chime. Immediately after the stimuli stopped (just before the target-specific action was completed), the experimenter asked "*O wa docchi?*" [Which one is the *O*?], where *O* either represented "*Kutsu*" [shoes] or "*Koppu*" [cup]. Note that the Japanese do not distinguish between singular and plural forms. The experimenter moved to the next trial when the participants pointed to their preferred side. The experimenter also moved on if the participants exhibited no pointing reactions after seven seconds had passed since the time the question was posed.

The experimenter left the final video frame on the monitor until the next trial.

The video stimuli were blocked by each target object, and the presentation order was pseudo-randomized within blocks. The presentation order between blocks was counterbalanced. After the first block (consisting of four trials) was finished, pictures unrelated to the task were presented for approximately 30 seconds to prevent the children from becoming bored and losing interest in the task. The experimenter then conducted the second block (comprising four trials)[17].

2.2.4 Data analysis

Indices and coding

We adopted the following indices for analysis.

Target–object choice responses. For each trial, when the participants pointed to the side of the video stimulus with the target objects, regardless of the actions displayed, we regarded this to be a target-object choice (i.e., correct) response. The participants appeared to have some concepts of object words when providing this response for the match condition. If the participants consistently gave this response for the mismatch condition, this suggested that object words' meanings are restricted to specific object categories. Two raters independently coded the toddlers' responses for all children during each trial ($36 \times 8 = 288$ trials). The interrater reliability was "almost perfect," based on the criteria of Landis and Koch (1977) ($\kappa = 0.97$, 99.3% agreement). Two trials were coded differently between raters. Here, a third rater newly coded the responses, and majority coding was adopted. Note that no participants pointed to only one side for all trials (i.e., side bias).

Development indices. We employed two indices to reflect the degree of participants' development: chronological age (months) and total vocabulary size, calculated by summing up the checked items from the J-MCDI.

[17] To confirm the task's appropriateness, we conducted it for 46 adults, specifically undergraduate and graduate students (28 males, 17 females, and 1 of another gender; M_{age} = 23.5 years, SD = 4.6). The results showed that regardless of the conditions and target objects, all participants succeeded in reliably choosing the stimulus where the target object was presented, with a probability of 100%.

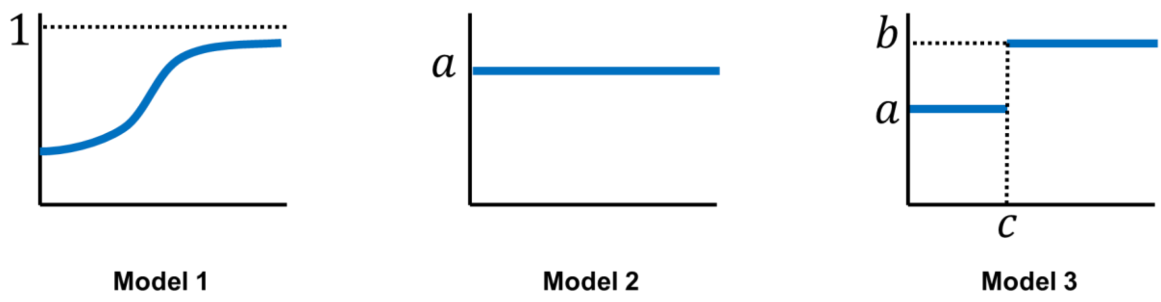

Figure 2.4 Schematic diagrams of candidate statistical models. The logistical model (Model 1), constant model (Model 2), and discontinuously constant model (Model 3) were proposed. The y axis indicates the probability of target-object choice responses, while the x axis signals developmental indices (i.e., age in months or vocabulary size).

Statistical modeling

We performed statistical modeling to predict developmental tendencies toward target-object choice responses. Assuming that pointing responses were independent for each trial, we first proposed three types of simple model candidates, using f as the probability of the target-object choice response and x as the development index (**Figure 2.4**). Model 1 was logistical; it expressed changes in probability from 0 to 1 according to two parameters:

$$f(x; a, b) = \frac{1}{1 + \exp[-a(x - b)]} \ . \tag{2.1}$$

Model 2 was constant. This meant that the tendency of the target-object choice was unchanged, regardless of development:

$$f(x; a) = a \ . \tag{2.2}$$

Model 3 was discontinuously constant. This implies that the tendency of the target-object choice changed drastically at a particular developmental age (given in months) or vocabulary size, c:

$$f(x;\ a, b, c) = \begin{cases} a, & x < c \\ b, & x \geq c \ . \end{cases} \tag{2.3}$$

We included Model 3 in the model candidates to represent a serious developmental shift in a short period, as seen in productive vocabulary size (i.e., vocabulary spurt).

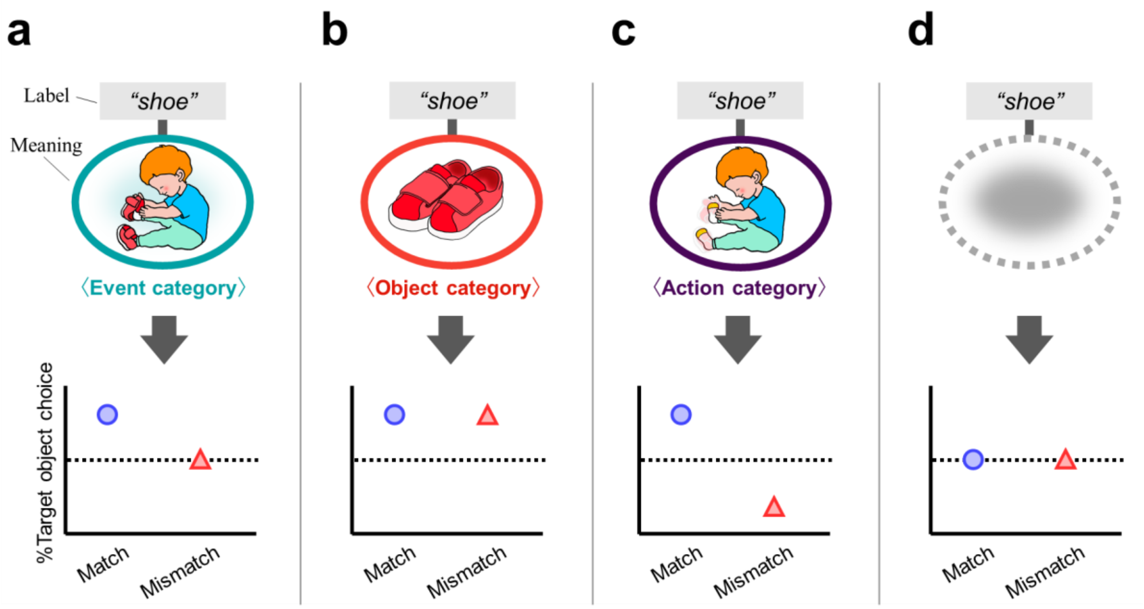

Figure 2.5 A schematic view of the possible results. The upper figures depict possible label-meaning connections. The lower graphs denote the probability of target-object choice responses for each condition. The dotted horizontal line refers to the level of chance. Each panel corresponds to the prediction when object words' meanings are (a) undifferentiated event categories, (b) specific object categories, (c) object-specific action categories, and (d) not yet constructed.

We estimated the parameters of each model for each condition that best fit the collected data using the maximum likelihood method. We then selected the optimal models based on Akaike's information criterion (AIC; Akaike, 1973). The AIC was used to evaluate the balance between the model's prediction accuracy (as expressed by the maximum log likelihood) and the cost of estimation (as expressed by the number of parameters). The smaller the AIC, the better the model.

Predictions

There were four possible results. If object words' meanings were undifferentiated global event categories comprising both objects and actions, as we hypothesized, then the probability of target-object choice responses in the match condition would exceed the level of chance, while that in the mismatch condition would remain at the level of chance. If object words' meanings were specific object categories independent of object-specific actions, then the probability of target-object choice responses would be above the level of chance in both conditions. By contrast, if such words' meanings were essentially formed by object-specific actions (versus objects themselves), then the probability of task performance in the match condition would exceed the level of chance, whereas that in the mismatch condition would be below the level of chance. People who assume that object words' "core meanings" must be object categories themselves may feel this possibility to be somewhat peculiar. However, it cannot be easily ruled out, as some researchers have argued for the importance of object-specific actions in object word learning (e.g., Kobayashi, 1992; McCune, 2008; Nelson, 1973a, 1974). Even Werner and Kaplan (1963) stated that young children often use the same label for phenomena based on how the objects are used in a given context (p. 117). Lastly, if object words' meanings had not yet emerged, then the probability of target-object choice responses would remain at the level of chance in both conditions. **Figure 2.5** portrays a schematic view of these four possibilities.

We predicted that younger children (for whom object words' meanings are undifferentiated event categories) would respond to the target-object choice with a high probability in the match condition, but would respond at the level of chance in the mismatch condition. On the other hand, we predicted that older children (for whom object words' meanings are differentiated into specific object categories) would consistently respond to the target-object choice with a high probability for both the match and mismatch conditions. Models 1 or 2 would thus be selected in the match condition, while Models 1 or 3 would be selected in the mismatch condition.

2.3 Results

For the age in months developmental index, Model 1 was selected as the best fit for the match condition, while Model 3 was selected for the mismatch condition (**Figure 2.6-a** and **Table 2.1**). The selected model in the match condition indicated that even toddlers aged 19 months in our study could show the target-object choice response with a high probability (72%). This increased along with development. However, the probability of the target-object choice response for children aged 21 months and younger remained at the chance level (50%) for the mismatch condition. This probability rose rapidly (90%) around 21 months of age.

A similar outcome was found using vocabulary size as the developmental index (**Figure 2.6-b** and **Table 2.1**). By employing the J-MCDI, the total productive vocabulary size was calculated from 11 to 706 words. The selected models were the same as those used for age. A vocabulary size of approximately 140 words was the boundary where the probability of the target-object choice response rapidly increased (93%). This corresponded

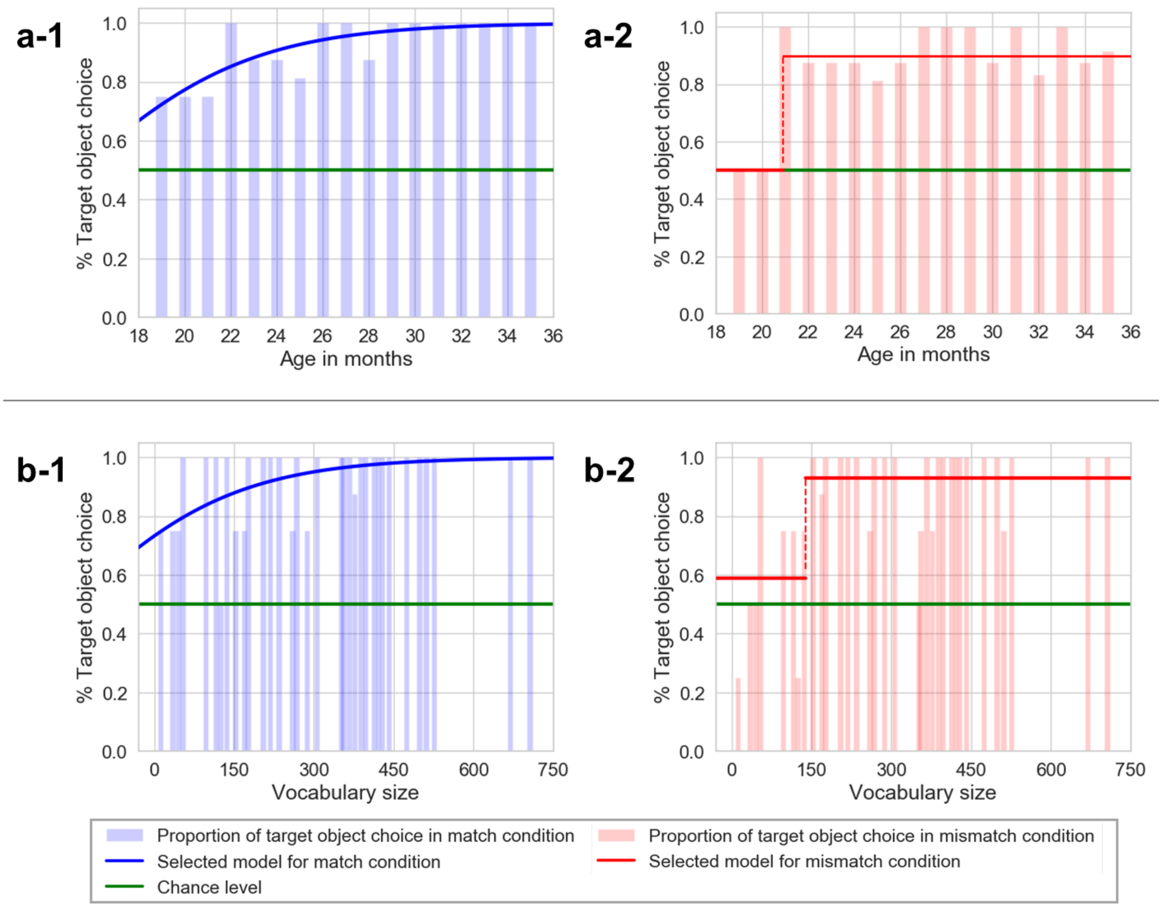

Figure 2.6 The developmental tendency for the probability of the target-object choice responses and the statistical modeling results. We divided each trial into classes for every month of age (a) or vocabulary size (b). We then calculated and displayed (as bar graphs) the proportions of all trials in each class, showing the target-object choice reaction. We superimposed each selected model (Model 1 for the match condition and Model 3 for the mismatch condition) onto each graph. A: The results of using age in months as the developmental index; the match (a-1) and mismatch (a-2) conditions. B: The results of using total productive vocabulary size as the developmental index; the match (b-1) and mismatch (b-2) conditions.

to 22–23 months of age for boys and 21–22 months for girls when compared with the 50th percentile estimate from the J-MCDI (Ogura & Watamaki, 2004). The initial probability phase of the target-object choice response in the mismatch condition was slightly higher (59%) than when age was used as the developmental index. However, both analyses were consistent in the sense that the probability of the target-object choice only changed drastically in the mismatch condition at a certain point in development.

We employed these developmental boundaries to divide the participants into two

Table 2.1 Estimated parameter values and the AICs of the candidate models.

	Age in months				Vocabulary size			
	Parameter			AIC	Parameter			AIC
	a	b	c		a	b	c	
Match condition								
Model 1	0.26	15.36		**69.35**	0.0065	-155.01		**72.78**
Model 2	0.92			79.72	0.92			79.72
Model 3	0.75	0.95	21.00	76.54	0.82	0.97	170.00	74.67
Mismatch condition								
Model 1	0.13	12.44		116.82	0.0061	-48.16		109.21
Model 2	0.85			121.64	0.85			121.64
Model 3	0.50	0.90	20.97	**112.28**	0.59	0.93	139.25	**106.87**

Note. The AICs for the selected models are shown in **bold**.

groups for further preliminary analysis. We counted the number of target-object choice responses for the four trials for each condition and participant. The mean proportions were then calculated. The mean proportion in the group younger than 21 months of age did not differ significantly from the chance level for either the match (75%, $t(3) = 2.44,\ p = 0.092,\ r = 0.82$) or mismatch (50%, $t(3) \approx 0,\ p = 1.0,\ r \approx 0.00$) conditions. The mean for the group 21 months of age and older was significantly above the chance level for the match (95%, $t(31) = 23.99,\ p < 0.001,\ r = 0.97$) and mismatch (90%, $t(31) = 14.66,\ p < 0.001,\ r = 0.93$) conditions. For vocabulary size groups, the mean proportion did not differ significantly from the level of chance only in the group with less than a 140-word vocabulary for the mismatch condition (59%, $t(7) = 1.00,\ p = 0.35,\ r = 0.35$). Others were above the chance level, including the group with less than a 140-word vocabulary for the match condition (84%, $t(7) = 5.23,\ p = 0.0012,\ r = 0.89$) and the group with a 140-word or greater vocabulary for the match (95%, $t(27) = 22.31,\ p < 0.001,\ r = 0.97$) and mismatch (93%, $t(27) = 16.97,\ p < 0.001,\ r = 0.96$) conditions. Although preliminary, **Appendix A** describes the outcomes of a two-way analysis of variance (ANOVA) for the different ages and vocabulary size groups; this analysis generally reproduced the statistical modeling results.

2.4 Discussion

This chapter preliminarily addresses the semantic pluripotency hypothesis, whereby for young children, object words' initial meanings do not differentiate into specific categories of objects as they do for adult speakers, but rather correspond to holistic, global, and undifferentiated event categories. The statistical modeling results predicted that even younger participants could make the target-object choice responses with a high probability; this increased with further development in the match condition (Model 1 was selected). Conversely, in the mismatch condition, this probability remained near the level of chance for participants younger than 21 months of age or those with vocabulary sizes smaller than 140 words; this rapidly increased (Model 3 was selected). These findings imply that at least two developmental phases are related to object words' semantic content, even when their labels are the same. In one phase, object words' meanings were undifferentiated event categories comprising both objects and actions. This was explained as follows: target-object choice responses appropriately corresponded to object words only when the target objects and target-specific actions were presented in the same stimulus. In the other phase, object words' meanings were differentiated into rigid object categories (becoming "true object words"). This was explained as follows: target-object choice responses occurred even when the stimulus presented strange actions with target objects. These findings support the semantic pluripotency hypothesis, and provide the first experimental evidence—although provisional due to our study's preliminary nature—for the theoretical and observational assumptions that early words consist of unique label-meaning connections that differ from those of adults, such that early words are not easily classified as specific parts of speech (e.g., nouns and verbs) (Church, 1961; Dewey, 1894; Lewis, 1936; McCune, 2008; Okamoto, 1982; Stevenson, 1893; Tomasello, 2003; Volterra, Bates, Benigni, Bretherton, & Camaioni, 1979; Vygotsky, 1934/1986; Werner & Kaplan, 1963). Further, our findings pave the way for comprehensive verification of these assumptions, related to both word production and word comprehension.

Regarding the early phase of this process, Werner and Kaplan (1963) strictly designated labels for very young children whereby semantic content encompasses undifferentiated events as "names" or "monoremes" (rather than "words") because (1) children's utterances in this phase do not fit into specific, differentiated semantic categories (e.g., objects or actions), and (2) these utterances lack specific syntactic and grammatical functions (e.g., subject or predicate, noun or verb). Names, or monoremes, are neither sufficiently decontextualized nor abstracted enough to carry syntactic and grammatical functions (see also McCune, 2008). This theoretical view suggests that object words' initial meanings are strongly connected to children's experiences of holistic events in an everyday context, rather than object categories themselves. In fact, words experienced in more distinctive contexts are produced earlier, regardless of the classification of parts of speech (nouns, predicates, and closed classes); moreover, these contextual features are better predictors of word learning than input frequency (Roy, Frank, DeCamp, Miller, & Roy, 2015).

The semantic differentiation from events into object categories (i.e., when "names" transform into "true object words") was predicted to emerge drastically around 21 months of age or when vocabulary size reached 140 words among the Japanese participants, which corresponded to 22–23 months of age for boys and 21–22 months for girls. Interestingly, the statistical modeling results suggest that this change does not take place gradually over a long period of toddlerhood, but rather occurs steeply in a short time. Referencing developmental theories such as the organismic-developmental approach (Werner & Kaplan, 1963) and the dynamic systems approach (Thelen & Smith, 1994), McCune (2008) marshaled the four essential cognitive and motor abilities as contributors to language learning: (1) the mental representation needed to form the meaning (i.e., the signified); (2) the phonetic skill necessary to form the vocal patterns (i.e., the signifier); (3) recognition of label-meaning correspondence, and (4) social-communicative capacity with others through the use of preverbal gestures and/or vocalization. The most important point of McCune's proposal is that referential language production begins when all of these subordinate abilities reach threshold levels (p. 213). This proposal offers a reasonable explanation for the developmental mechanisms of the semantic differentiation of object words. That is, once all the abilities contributing to this transition reach a certain level, such a change should occur very quickly. However, due to the preliminary study, we cannot definitively conclude that the semantic differentiation of object words occurs in a short burst of development, nor that the critical period of such a transition corresponds to exactly 21 months or 140 words (in terms of vocabulary size). Before looking more into the mechanisms of such a semantic transition, we next have to confirm how well the current results are robust and replicable. Indeed, some well-known cognitive abilities related to early word learning—such as shape bias (Gershkoff-Stowe & Smith, 2004; Graham & Poulin-Dubois, 1999; Imai et al., 1994; Landau et al., 1988, 1998; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith et al., 2002; Yee et al., 2012), fast mapping (Ellis, Borovsky, Elman, & Evans, 2015; Schafer & Plunkett, 1998; Torkildsen et al., 2008; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; Woodward, Markman, & Fitzsimmons, 1994), and the meta-linguistic knowledge of "things have their own names" (i.e., "naming insight") (Dore, 1978; Kamhi, 1986; McShane, 1979; Nelson, 1983a)—generally emerge a few months earlier, or several dozens of vocabulary sizes fewer, compared to the developmental markers predicted in our study. Such discrepancies in the occurrence period of cognitive development may be explained by the small sample size simply affecting the results, or that these abilities, according to McCune (2008), might serve as standby states until the other abilities required for the transition are sufficiently developed[18].

There may be technical criticisms concerning the decreased proportion of target-object choice responses in the mismatch condition. For example, some might think that this happened because participants younger than 21 months or with a vocabulary size of less than 140 words thought that the filler objects (i.e., the baskets and toy shovel) were the exact target objects (i.e., the shoes and cup). However, the filler and target objects differed in shape, and the participants were always able to compare the appearance of the filler and target objects since they were presented simultaneously. Thus, if the participants thought that the baskets were the shoes because they were available for “putting on,” this would suggest that children judge object words’ meanings not based on objects’ appearance, but rather on object-specific actions (or object functions or affordances). However, we did not derive such outcomes.

Werner and Kaplan’s (1963) hypothesis on initial word meanings cannot be addressed directly by observational studies alone, and unfortunately, their hypothesis has not been investigated experimentally. We believe that the aim of this preliminary study has been fulfilled, despite several limitations, in showing the possible experimental verification of the semantic pluripotency hypothesis about object words, and in identifying the sensitive period where the semantic differentiation of such words occurs. Among several points in which the research design can be improved, we raise three major concerns that we will deal with in subsequent chapters.

First and foremost, it is necessary to conduct a similar experiment using a larger sample size, with a narrower target age range around the critical period when the semantic differentiation of object words occurs, not only to confirm the replicability of our findings, but also to determine whether the semantic pluripotency hypothesis is a universal developmental principle, or merely one possible strategy for early word learning. Particularly, narrowing down the target age range to 18–23 months would be preferable. However, since the method we employed requires volitional arm movement as an index reflecting toddlers’ choices, we often did not observe clear pointing responses, despite looking at either side of the monitor (i.e., left or right). In fact, among the 52 potential participants, we excluded 9 toddlers (17%) due to the difficulty we encountered in coding

[18] Relatedly, Werner and Kaplan (1963) noted that children remain at the undifferentiated single-word period for a quite long time.

pointing responses. As long as we strictly adopt manual measures such as pointing, absent responses are treated as missing values. Less robust observability of volitional pointing responses is frequently seen, especially in toddlerhood (Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Hendrickson, Poulin-Dubois, Zesiger, & Friend, 2017). To deal with this problem, alternative measures such as looking responses (e.g., preferential looking) are available. However, it is still unclear how, and to what extent, these two different measurements of looking and pointing responses can be treated as equivalent to each other. Therefore, we directly address this concern in the next chapter, before investigating the robustness of the preliminary findings described in this chapter.

Second, although object words' initial meanings were likely undifferentiated event categories comprising both objects and actions, it is still unclear whether toddlers can comprehend object word referents solely based on object-specific actions. It would be worthwhile to examine this possibility because, as discussed in the section on predictions, some researchers stress the importance of object-specific actions in object word learning (e.g., Kobayashi, 1992; McCune, 2008; Nelson, 1973a, 1974; Werner and Kaplan, 1963). The following task represents an improvement: toddlers would be simultaneously shown video stimuli juxtaposing "putting baskets on" and "performing meaningless actions with baskets." They would then be asked to indicate the shoes. This would enable the scrutiny of object word meanings in a way that depends mostly on object-specific actions. Relatedly, two possible interpretations should be identified for the finding that the probability of target-object choice responses remained at the level of chance in younger toddlers in the mismatch condition, since the two options competed with each other[19] (e.g., displaying "shoe" on one side and "put on" on the other side). One possibility is that the toddlers failed to properly recognize the object word referents under the situation in which objects and object-specific actions were not matched, as the initial meanings of such words were global events comprising both objects and actions. Alternatively, the toddlers may be more likely to choose the stimulus that includes object-specific actions, since they placed relatively great

[19] The competing stimuli in the mismatch condition may have demanded a greater cognitive load or answer confidence compared to the match condition. Indeed, some participants appeared to ponder their choice during the experiment, and their production of pointing responses was occasionally delayed in such cases. Exploratorily, we conducted a reaction time analysis, measured from when the experimenter's question was complete to when the participant made pointing responses in each trial, on the assumption that this index could reflect the degree of cognitive load or answer confidence during the task. The reaction times barely differed between conditions, suggesting that the cognitive load between conditions could be equivalent, although not completely (see **Appendix B**).

weight on how objects are used in object word comprehension. We will look at these remaining issues by adding two other conditions in the experimental task in Chapter 4.

Third, future research should explore how the semantic differentiation of object words is related to vocabulary growth. According to Werner and Kaplan (1963), toddlers only utter one word at a time, not because they have no choice but to refer only to very partial aspects of phenomena among what they want to express (due to an extremely limited vocabulary), but rather because one-word utterances encompass their understanding of global and undifferentiated events themselves. This hypothesis is partially supported by our study, in which object words' initial meanings included both objects and object-specific actions as a fusion. However, if their hypothesis is true, we would expect to observe a more in-depth phenomenon regarding the relationship between semantic differentiation and vocabulary growth. That is, once object word meanings are differentiated into specific object categories, toddlers would not be able to describe actions by undifferentiated object words, which would lead them to develop new labels capable of depicting actions in particular (i.e., action words). We will also investigate this possibility in Chapter 4 using both cross-sectional and longitudinal data.

2.5 Summary of this chapter

During the initial stages of language development, nouns generally corresponding to "object categories" are learned earlier than other parts of speech (e.g., verbs). However, it is unclear whether object words' initial meanings are the same for adults and toddlers, the latter of whom have recently begun to utter words. In this chapter, we have preliminary and experimentally explored the semantic pluripotency hypothesis, which is a refinement of the theoretical conjecture of Werner and Kaplan (1963). We posited that object words' initial meanings would not sufficiently differentiate into specific categories of objects themselves, but rather correspond to holistic event categories comprising both objects and object-specific actions. The experimenter prompted toddlers aged 19–35 months (n = 36) to select one of two juxtaposed video stimuli (e.g., "putting shoes on" versus "rubbing two baskets" in the match condition; "putting baskets on as if they were shoes" versus "rubbing shoes" in the mismatch condition) using questions about object words (e.g., "Which ones are shoes?"). Statistical modeling demonstrated that for toddlers younger than 21 months of age or for those who have a vocabulary size smaller than 140 words, object words' meanings were deeply influenced by object-specific actions (e.g., "putting on"), while they subsequently differentiated the meanings of object words into specific object categories independent of such actions (e.g., "shoes" alone), thus causing them to become "true object words.” These findings support the semantic pluripotency hypothesis and provide the first experimental evidence for the theoretical and observational assumptions that early words are not easily classified into parts of speech by their forms alone (e.g., nouns and verbs). We have therefore discussed the flexibility and uniqueness of label-meaning connections during early language development.

Chapter 3

Exploring the Equivalence between Looking and Pointing Responses in Forced-Choice Tasks

Based on:

Hagihara, H., Ienaga, N., Terayama, K., Moriguchi, Y., & & Sakagami, M. (2021). Looking represents choosing in toddlers: Exploring the equivalence between multimodal measures in forced-choice tasks. *Infancy*, *26*(1), 148–167. https://doi.org/10.1111/infa.12377

This chapter explores to what extent, and how, looking responses can be interpreted as equivalent to pointing ones. Through this investigation, we aim to apply the experimental task used in Chapter 2 to the study in Chapter 4. We intend to provide the foundation for utilizing both pointing and looking responses as indices reflecting task performance on semantic differentiation of object words to prevent missing data from increasing due to the frequent absence of pointing responses and to analyze data of different indices in the same way.

3.1 Introduction

The two-alternative forced-choice paradigm has long been leveraged in various studies on cognitive development, such as numeric skills, false-belief understanding, and prosocial characters (e.g., Fantz, 1964; Hamlin, Wynn, & Bloom, 2007, 2010; Onishi & Baillargeon, 2005; Southgate, Senju, & Chibra, 2007; Starkey, Spelke, & Gelman, 1983; Wagner & Johnson, 2011). As measures to estimate cognitive abilities, manual responses, such as pointing, touching, or reaching, have been widely utilized.

Especially in the field of language development, well-known examples of the two-alternative forced-choice tasks using children's arm responses are the forced-choice pointing paradigm (Fernandes, Marcus, Di Nubila, & Vouloumanos, 2006; Fisher, 1996; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Noble, Rowland, & Pine, 2011) and the Computerized Comprehension Task (Friend & Keplinger, 2003, 2008). In these methods, children watch two juxtaposed pictures or video clips and are asked to choose, by manual responses, the one that matches auditory instructions. These methods have various advantages (Noble et al., 2011). First, methods using manual measures do not require expensive specialized equipment and can be administered easily. Second, such measures can be easily coded, even when manual (not automated) coding is adopted, since children produce an overt, volitional response. Third, these methods provide direct, unambiguous, and less noisy indices that can be analyzed and interpreted easily. This characteristic of the indices is also advantageous since there is less room for analytical arbitrariness, such as conducting statistical analysis in a haphazard manner or summarizing data in a self-serving manner. Fourth, manual measures are applicable to a broader age group of children and adults; hence, long-term developmental differences can be investigated.

Methods using manual responses are generally suitable for children older than two years old (Ambridge & Rowland, 2013), yet some studies have adapted this method to toddlers around 18 months old (Friend & Keplinger, 2003, 2008; Gurteen, Horne, & Erjavec, 2011). However, since methods using manual measures require a volitional response of arm movement, how to address absent responses if observed is often uncertain, particularly when applying these methods to toddlers (Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015). For example, Hendrickson, Poulin-Dubois, Zesiger, & Friend (2017) reported that, in a familiar word comprehension task, absent touching responses were observed for roughly one-third of trials with 16-month-olds and 10% with 22-month-olds. Also, we observed in Chapter 2 that from 52 toddlers aged 19–35 months who participated in the preliminary study, 9 toddlers (17%) were excluded due to the difficulty in coding pointing responses. In strictly adopting manual measures, absent responses are treated as missing values. However, are there any possibilities to avoid wasting the collected data?

In such cases, looking responses, such as preferential looking, may be available as alternative measures, if they are extractable and codable (e.g., from video recordings). In fact, there are many infant studies that use looking responses as measures of two-alternative forced-choice tasks, such as the intermodal preferential looking paradigm (Bailey & Plunkett, 2002; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Golinkoff, Ma, Song, &

Hirsh-Pasek, 2013; Yuan & Fisher, 2009) or the looking-while-listening paradigm (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). For example, the preferential looking paradigm leverages young children's tendency to look significantly longer at a stimulus that matches linguistic input than a distracter presented side-by-side (Tafreshi, Thompson, & Racine, 2014), and this method has been used to verify whether children can identify the correct referent of novel words (Chan, Meints, Lieven, & Tomasello, 2010; Gurteen, et al., 2011; Horváth, Myers, Foster, & Plunkett, 2015) or familiar words (Durrant, Delle Luche, Cattani, & Floccia, 2015; Mani & Plunkett, 2011; Valleau, Konishi, Golinkoff, Hirsh-Pasek, & Arunachalam, 2018). These methods using looking responses are advantageous as they are applicable even for infants under 18 months old (Imai et al., 2015; Mani & Plunkett, 2010; Smith & Yu, 2008), since they do not need children's volitional manual responses but only spontaneous looking ones.[20] However, if looking responses are used as a dependent variable instead of manual ones to reduce the exclusion rate or missing data, how well can we treat the results of looking measurements as equivalent to those of manual ones?

Few studies have investigated children's looking and concurrent manual responses within the same two-alternative forced-choice task, because most infant and toddler studies used either of these measures selectively (Hendrickson et al., 2015). Moreover, some researchers remain skeptical about looking time as an appropriate index reflecting higher-order cognitive abilities (e.g., Haith, 1998), and there were dissociations of results between modalities in research, not only on language but also on other cognitive development[21] (Abbot-Smith, Chang, Rowland, Ferguson, & Pine, 2017; Ahmed & Ruffman, 1998;

---

[20] Recently, a new method using looking responses called the "visual array task" was proposed (Hauschild, Pomales-Ramos, & Strauss, 2020). In this task, eight items are arrayed and presented simultaneously and prompt children to look at one of them that matches the instruction. This method is likely beneficial for obtaining more detailed data on object word and object category learning. However, we are not sure it can be applied to research on action word and action category learning, because more than two video stimuli presented simultaneously may be hard for young children to process.

[21] Even within the same modality of looking responses, contradictory results are often obtained in infant research using preferential looking measures. For example, after exposure to visual or auditory stimuli, infants are known to show either familiarity or novelty preferences to the test when two stimuli are presented simultaneously (Roder, Bushnell, & Sasseville, 2000). Since the preference that an infant shows is influenced by several factors, such as lexical knowledge (DePaolis, Keren-Portnoy, & Vihman, 2016) and past experience with the experimental paradigm (Santolin, Garcia, Zettersten, Sebastian-Galles, & Saffran, 2021), often, different results are reported for similar tasks.

Charles & Rivera, 2009; Gurteen et al., 2011; Ruffman, Garnham, Import, & Connolly, 2001; Winters, Dubuc, & Higham, 2015). Hence, it is unknown to what extent looking and manual measurements can be interpreted analogously.

To our knowledge, Hendrickson and colleagues are the only researchers who conducted the forced-choice pointing and the intermodal preferential looking paradigms simultaneously with toddlers (Hendrickson & Friend, 2013; Hendrickson et al., 2015, 2017). In Hendrickson and Friend (2013), 16–18-month-olds participated in the familiar word comprehension task. Toddlers' looking and manual responses were recorded via HD video cameras and were coded manually. Looking responses were categorized frame-by-frame into three (i.e., left fixation, right fixation, or away look), whereas manual responses were categorized into three for each trial (i.e., target touch, distractor touch, no touch). The results showed that toddlers looked significantly longer at the stimulus to which they subsequently reached regardless of stimuli type (i.e., target or distractor). Furthermore, in trials where reaching responses were not observed, toddlers looked at the target stimulus significantly longer than chance. Referring to recent connectionist studies (Munakata, 2001; Munakata & McClelland, 2003), Hendrickson and colleagues interpreted these findings to indicate that looking and manual responses reflect different levels of understanding for the word-referent association. That is, preferential looking can be observed even when representations of words are fragile, whereas reaching can be demonstrated for only robust representations. Their subsequent studies enhanced this view and further explored the way to detect children's knowledge status on a certain word by leveraging both looking and manual measures (Hendrickson et al., 2015, 2017). However, since they have mostly focused on the different interpretabilities between two modalities, the question of how well each measure can be interpreted as equivalent remains unexplored.

As the first step to directly address this question, we investigated how accurately toddlers' volitional pointing (i.e., left or right) could be predicted from preferential looking. If pointing were accurately predictable from preferential looking, then these two measurements are related and looking responses can, to some extent, be reasonably used as alternative measures to manual ones. If the prediction accuracy were low, this indicates that these two indices reflect different cognitive processes or have different robustness to noise, irrelevant of choice. For this study, we utilized video data of the experiments regarding the semantic differentiation where both looking and pointing measures could be coded (see Chapter 4 for more details). In this data, 18–23-month-old toddlers participated in a two-alternative forced-choice task evaluating whether the meanings of object words were affected by object-specific actions. Here, we particularly focused on exploring the temporal features of frame-by-frame coded preferential looking that would predict binary pointing

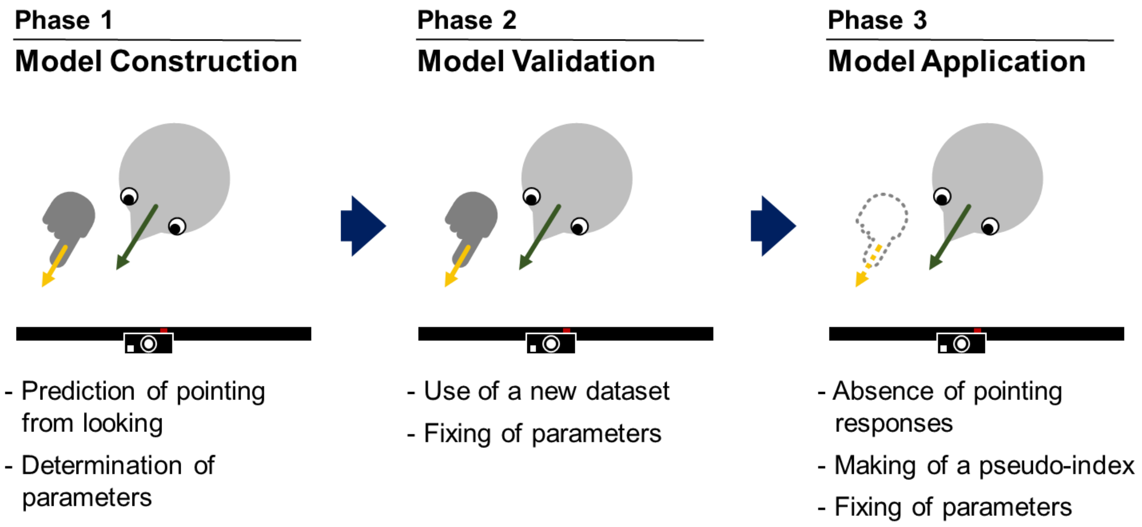

Figure 3.1 A schematic view of the flow of the study in this chapter.

responses most adequately.

This study consisted of three phases (**Figure 3.1**). In Phase 1, we created two types of models for the prediction of pointing from preferential looking: Simple Majority Vote (SMV) and Machine Learning-Based (MLB) models. In SMV, the proportion of total looks to juxtaposed stimuli (left or right) for each trial was calculated while changing the target time window, and the dominant side was used as a prediction index. In MLB, we adopted the decision-tree-based algorithm LightGBM (Ke et al., 2017). We chose this algorithm for several reasons, including the fact that it is a state-of-the-art option for a relatively small number of input variables. We used this machine learning method because it was expected that certain particular time ranges and/or their combinations would yield higher prediction accuracy than merely using the proportion of looks to either stimulus. In Phase 2, we conducted the validation test of the created models by using a dataset not used in Phase 1 in each model. If it turned out that pointing and preferential looking were closely related, are features of looking with clear pointing were the same as those without pointing? In Phase 3, we exploratorily applied the created models to data of toddlers who did not produce clear pointing to preliminarily investigate this question. Since there were no absolute correct answers (i.e., absent pointing responses), we adopted manual estimations of pointing responses from looking behavior as a pseudo-correct index to evaluate the applicability of the created models. If the agreement of pointing estimations between human coders and the models were equivalent to the prediction accuracy calculated in Phases 1 and 2, it would indicate that looking responses were similar regardless of executing manual responses. If the agreement dropped, it would suggest that looking responses without manual ones were qualitatively different from those with manual ones.

3.2 Phase 1: Model construction

3.2.1 Methods

Video recordings

For this study, we utilized videos in which toddlers' familiar word comprehension was investigated using the forced-choice pointing paradigm (see Chapter 4 for details). Inter-rater reliability of pointing responses (left or right) was confirmed with 97.8% of agreement (κ = 0.96). Of the participants who showed explicit pointing responses for 75% or more in all 16 experiment trials, 36 toddlers were selected and allocated to model construction (n = 24) and validation (n = 12) in this study so that age and gender were not biased. For Phase 1, we used videos of 369 trials with 24 monolingual Japanese toddlers aged 18–23 months (12 boys and 12 girls; M_{age} = 21.1 months, SD =1.7). Fifteen trials (8 toddlers) were excluded for lack of explicit pointing responses. Each remaining participant responded clearly in 12–16 trials (M = 15.4, SD =1.1).

In the experimental task, toddlers sat on a small chair or the lap of a nursery school teacher and looked at the 21.5-in touch screen (490 × 243 mm) with a viewing distance of approximately 30 cm. Toddlers' preferential looking and pointing responses were recorded via a webcam at the center of the top of the screen (30 frames/second). Toddlers completed the forced-choice task modified from the one in Chapter 2. This task aimed to investigate whether the meanings of object words were affected by object-specific actions. This task consisted of 16 trials in which four different conditions were included, which varied in terms of how much participants had to depend on object-specific actions to make judgments about referents of familiar object words. For example, in one condition, immediately after watching two juxtaposed videos of "putting on shoes" (the target stimulus) and "rubbing two baskets in front of one's chest" (the foil stimulus), participants were prompted to choose one of them by pointing to answer "*Kutsu wa docchi?*" [Which ones are the shoes?]. In another condition, they watched video stimuli located side-by-side of "rubbing shoes in front of one's chest" (the target stimulus) and "putting on two baskets as if they were shoes" (the foil stimulus) and then were asked to choose the one that matched the question of which ones are the shoes. Before performing the tasks, participants were prompted to look at icons on the screen's center and each corner for later calculation of angle compensation; they also engaged in warm-up trials to understand the task rule. This research was conducted according to guidelines outlined in the Declaration of Helsinki, with written informed consent obtained from the parents of all participants before data collection. All procedures involving human subjects in this research were approved by the ethics committee for human and animal research of the Graduate School of Human and Environmental Studies at Kyoto University.

Face/gaze detection and pre-processing

The video for each trial was cropped with a time window from when the question ended to 2,000 ms thereafter since, at a later time, looking behavior is no longer considered to be related to the stimulus (Delle Luche, Durrant, Poltrock, & Floccia, 2015; Swingley & Aslin, 2000). We set the starting point of the potential time window not at the onset or offset of the target word but at the end of the question sentence because, in Japanese, a listener cannot determine if a sentence is a question unless they hear the sentence through to the end due to the grammatical difference in word order from English. For example, in the sentence "*Kutsu wa docchi?*" [Which ones are the shoes?], "*kutsu*" refers to the target word "shoes," "*wa*" refers to a postpositional particle indicating that the preceding word is the subject, and "*docchi*" refers to the interrogative marker "which ones." Note that Japanese allows for a dropping of the verb "are" in the sentence, does not distinguish between singular and plural forms of nouns, and does not use articles like "the."

In line with recent vigorous studies on automatic gaze estimation using webcam-based data (Chouinard, Scott, & Cusack, 2019; Papoutsaki et al., 2016), we used an open-source library, OpenFace 2.2.0 (Baltrušaitis, Zadeh, Lim, & Morency, 2018) for automatic face and gaze estimation. We adopted this toolkit because it provides rich information such as facial landmarks, gaze directions, and three-dimensional head positions and angles, because it is freely available for research purposes, and works in the local environments, which reduces ethical concerns. Although this toolkit requires only an RGB camera to estimate gaze direction using machine learning techniques, the precision is far less than eye-trackers (Higuchi et al., 2018). However, facial detection is more robust than gaze detection. Hence, we used horizontal face angles as well as gaze angles as measurements indicating whether a participant looked to the right or left side of the screen. Additionally, although complete preferential looking coding requires annotation of not only whether a child is looking left or right but also whether a child is looking at or away from the screen, we began with the simplified coding to distinguish whether a child was looking right or left relative to the center of the monitor.

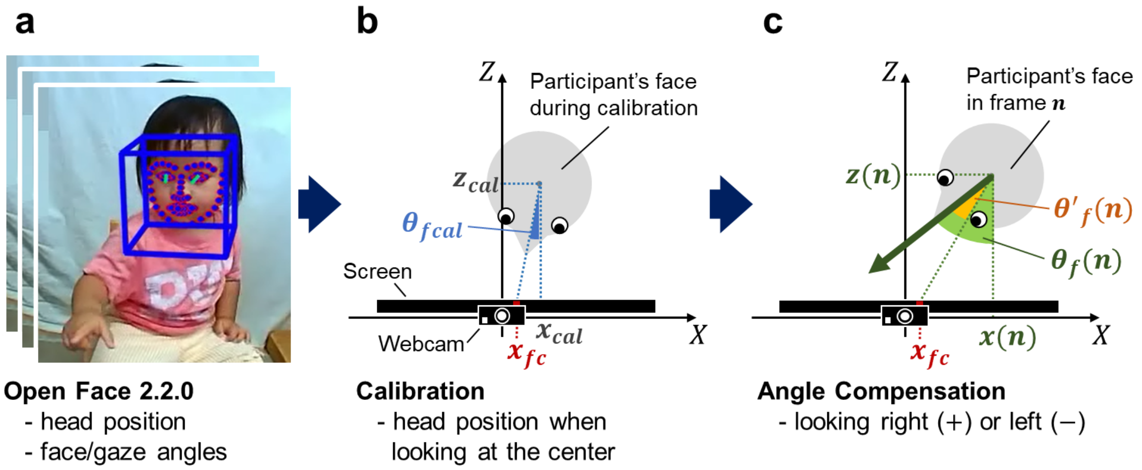

Figure 3.2 Pre-processing flow for estimating preferential looking. (a) Using video data recorded by a webcam located at the center of the top of the screen, head positions, and raw face and gaze angles were estimated by OpenFace 2.2.0. (b) The head position when a participant looked at the center icon of the screen was calculated to compensate face and gaze angles so that these indices were always zero when a participant looked at the center of the screen regardless of changing head position. (c) Compensated face and gaze angles were calculated so that the indices were positive when looking right relative to the center of the screen.

After estimating face and gaze angles using OpenFace 2.2.0, we calculated angle compensation so that the indices were always zero when a participant looked at the center of the monitor regardless of changing their head position (**Figure 3.2**). First, we calculated x_{fc} which was the head position when a child looked at the center icon on the screen as follows:

$$x_{fc} = x_{cal} - z_{cal} \times \tan(\theta_{fcal}), \tag{3.1}$$

where x_{cal}, z_{cal}, and θ_{fcal} represent the average of 30 frames of each variable when a child looked at the center icon of the screen. The variable x stands for the horizontal distance between the webcam and participants' head along with the screen, z for the distance between the webcam and participants' head, and θ_f for the raw horizontal face angle, respectively. The compensated face angle $\theta'_f(n)$ in frame n was then calculated as follows:

$$\theta'_f(n) = \theta_f(n) - arctan\left(\frac{x(n) - x_{fc}}{z(n)}\right). \tag{3.2}$$

Thus, when $\theta'_f(n)$ is greater than 0, it indicates that the toddler is looking right. Note that the expression calculating the compensated gaze angle $\theta'_g(n)$ needs slight changes because the plus and minus of its angle are inverted from the face angle in the use of OpenFace 2.2.0 as follows:

$$\theta'_g(n) = -\theta_g(n) - arctan\left(\frac{x(n) - x_{gc}}{z(n)}\right),$$
$$x_{gc} = x_{cal} + z_{cal} \times tan(\theta_{gcal}). \quad (3.3)$$

As in the face direction, $\theta_g(n)$ stands for the raw horizontal gaze angle and θ_{gcal} for the averaged angle of looking at the center of the screen. For each frame, OpenFace 2.2.0 provides the "Confidence" value, which indicates how precisely a face can be detected, ranging from 0 to 1 (a higher value indicates successful face detection). When the face detection failed (i.e., the Confidence value was low), the temporal linear interpolation was conducted for estimated variables.

To confirm how reliable face and gaze angle estimations in OpenFace 2.2.0 were, we verified the agreement between the estimation and frame-by-frame manual coding of preferential looking (left or right). A trained coder manually annotated the toddlers' preferential looking for approximately 25% of the data in Phase 1, in which four of each participant's trials were pseudo-randomly extracted (4 trials × 24 participants = 96 trials).

Creating models for predicting pointing responses from preferential looking

Using time-series data of face and gaze horizontal angles as an input, we created two types of classifiers, which predict pointing responses from preferential looking. We regarded the pre-existing human annotation of pointing responses as the correct answer.

Simple Majority Vote (SMV) model. In the intermodal preferential looking paradigm, researchers have conventionally compared the proportion of looks to the target stimulus to the proportion of looks to the foil stimulus within a certain time window (Ambridge & Rowland, 2013). In line with this procedure, we converted the compensated face and gaze angles into a binary index (left or right), and regarded the dominant side of looks as a prediction for pointing. One major difference of the SMV model from the conventional approach was that the pointing prediction was calculated while changing the target time window to reflect the temporal features of looking responses in the optimization of the prediction accuracy of pointing responses (**Figure 3.3**). We calculated a variable $P_{i,j}$ modeled for this method, where $NR_{i,j}$ and $NL_{i,j}$ stand for the number of frames in

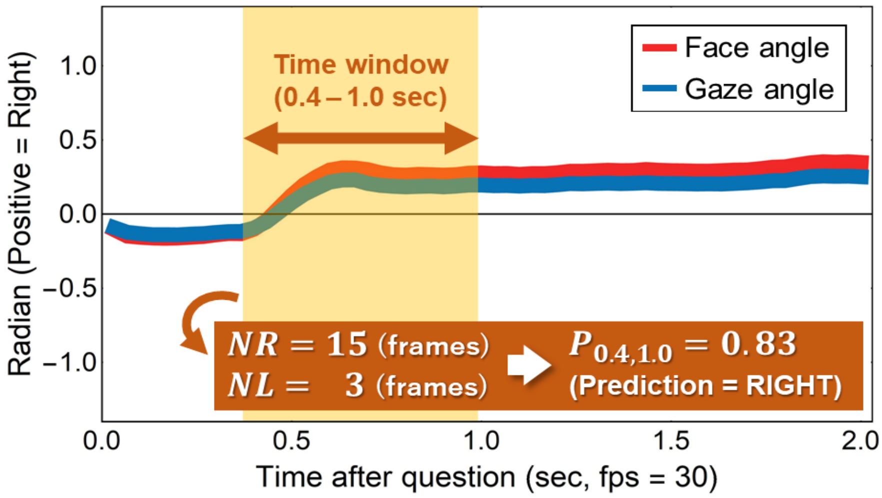

Figure 3.3 A schematic view of the SMV model. The number of frames where a participant looked right or left was calculated from the compensated face and gaze angles within a certain time window (from 0.4 to 1.0 sec in this case). The dominant side was regarded as a prediction of pointing responses. The target time window was moved between 0.0 to 2.0 sec after the completion of the question to optimize prediction accuracy.

which a participant looked to the right and left side, respectively. The variable $P_{i,j}$ was computed while changing the starting time point i and the ending time point j of the target time window as follows:

$$P_{i,j} = 0.5 \times \frac{NR_{i,j} - NL_{i,j}}{NR_{i,j} + NL_{i,j}} + 0.5\ . \tag{3.4}$$

$P_{i,j}$ ranged between 0 and 1 (indicated a participant completely looked left or right, respectively) and when $P_{i,j}$ was greater than 0.5, it was predicted that the participant pointed right. If $P_{i,j}$ was equal to 0.5, we defaulted the prediction to the right side.

Machine Learning-Based (MLB) model. In addition to the SMV model, we used the decision-tree-based algorithm LightGBM (Ke et al., 2017) as a classifier since we expected that certain time ranges and/or their combinations would yield a higher prediction accuracy than simply using the proportion of looks to the target stimulus. We adopted this algorithm because it is a state-of-the-art option for a relatively small number of input variables, it is one of the most popular methods in recent machine learning competitions, and the contribution of each input variable to the prediction, called “importance,” can be easily quantified and visualized using LightGBM. First, we standardized the compensated face and gaze angles, respectively. We then determined the hyperparameters of the MLB model using the grid search optimization procedure, where all combinations of parameters were used to explore the optimal values of each parameter. LightGBM has several hyperparameters that are related to prediction accuracy such as the number of trees and the learning rate. For example, the number of trees refers to how many decision trees are combined; these are the sub-elements that make up the model. The learning rate refers to how much information about the learning error at a certain step is propagated to the next step. The prediction accuracy of each parameter was calculated with 10-fold cross-validation to avoid overfitting to the specific data. In 10-fold cross-validation, the given data was split into training (90%) and validation (10%), repeated training 10 times for each partition, and the model performance was evaluated by averaging the obtained results. We adopted the parameters that produced the highest accuracy and the final prediction model was created using all the data. The prediction variable produced by LightGBM was continuous ranged from 0 to 1, where a value greater than 0.5 indicated the prediction that a participant pointed right. As in the SMV model, a value that equaled to 0.5 was defaulted to pointing right.

Evaluation of created models

To evaluate the prediction accuracy of the created SMV and MLB models, the predicted side of pointing responses (right or left) produced by each model was compared to previously obtained manual coding for all trials. As indices, we used the area under the curve (AUC) of the receiver operating characteristic (ROC) curve, accuracy rate, and kappa coefficient. AUC is a measurement reflecting how much the model has the discriminative ability, where 1.0 indicates the model can perfectly predict the correct results whereas 0.5 indicates chance level. According to the rule of thumb (Akobeng, 2007; Swets, 1988), an AUC greater than 0.9 has high, 0.7–0.9 has moderate, and 0.5–0.7 has low accuracy. The kappa coefficient is also a well-used measurement, where a higher value indicates higher accuracy. Based on the criterion by Landis & Koch (1977), kappa of 0.81–1.00 has almost perfect, 0.61–0.80 has substantial, and 0.41–0.60 has moderate strength of accuracy.

3.2.2 Results

The agreement of preferential looking between the estimation in OpenFace 2.2.0 and manual coding

For 96 trials randomly extracted from Phase 1 data, preferential looking estimated using

Table 3.1 Evaluation of the created models of predicting pointing responses with highest accuracy and their robustness.

	Phase 1. Model construction (369 trials)		Phase 2. Model validation (183 trials)		Phase 3. Model application to no-pointing trials (176 trials)	
	SMV	MLB	SMV	MLB	SMV	MLB
Face angles as input						
AUC	0.945	0.941	0.934	0.938	0.908	0.877
Accuracy rate	0.892	0.897	0.858	0.869	0.818	0.795
κ	0.781	0.793	0.703	0.727	0.637	0.590
Gaze angles as input						
AUC	0.941	0.942	0.933	0.952	0.897	0.864
Accuracy rate	0.892	0.892	0.858	0.869	0.818	0.778
κ	0.781	0.781	0.703	0.726	0.637	0.557

OpenFace 2.2.0 demonstrated almost perfect reliability with frame-by-frame human coding both in the face (κ = 0.85, 92.7% agreement) and gaze (κ = 0.82, 91.2% agreement) directions. Therefore, we continued to use the automated face and gaze angle estimations produced by OpenFace 2.2.0.

To what extent did preferential looking reflect pointing responses?

The indices reflecting prediction accuracy for the best models of the SMV and the MLB are shown in **Table 3.1**. In the SMV model, we explored the optimal time window that predicted toddlers' pointing responses using all the data in Phase 1. The best time window was found to be from 0.0 to 2.0 s immediately after the question ended in both models using face and gaze angle trajectories, while the narrow time windows including the very beginning or the end of all the potential time range showed lower prediction accuracy (**Figure 3.4**). The SMV model with the best time window demonstrated the highest accuracy in both models using face and gaze direction (89.2%), with high AUC and substantial kappa. The MLB model also showed high prediction rates in AUC, accuracy rate, and kappa coefficient, which were equivalent to the best SMV models. The temporal feature reflecting the prediction of pointing responses were visualized using "importance," the contribution probability of each input frame to the prediction (**Figure 3.5**). Approximately, among all 2.0 s of the potential time range, the first 1.0 s and the last 0.4 s were relatively critical to the classification for both face and gaze angle trajectories.

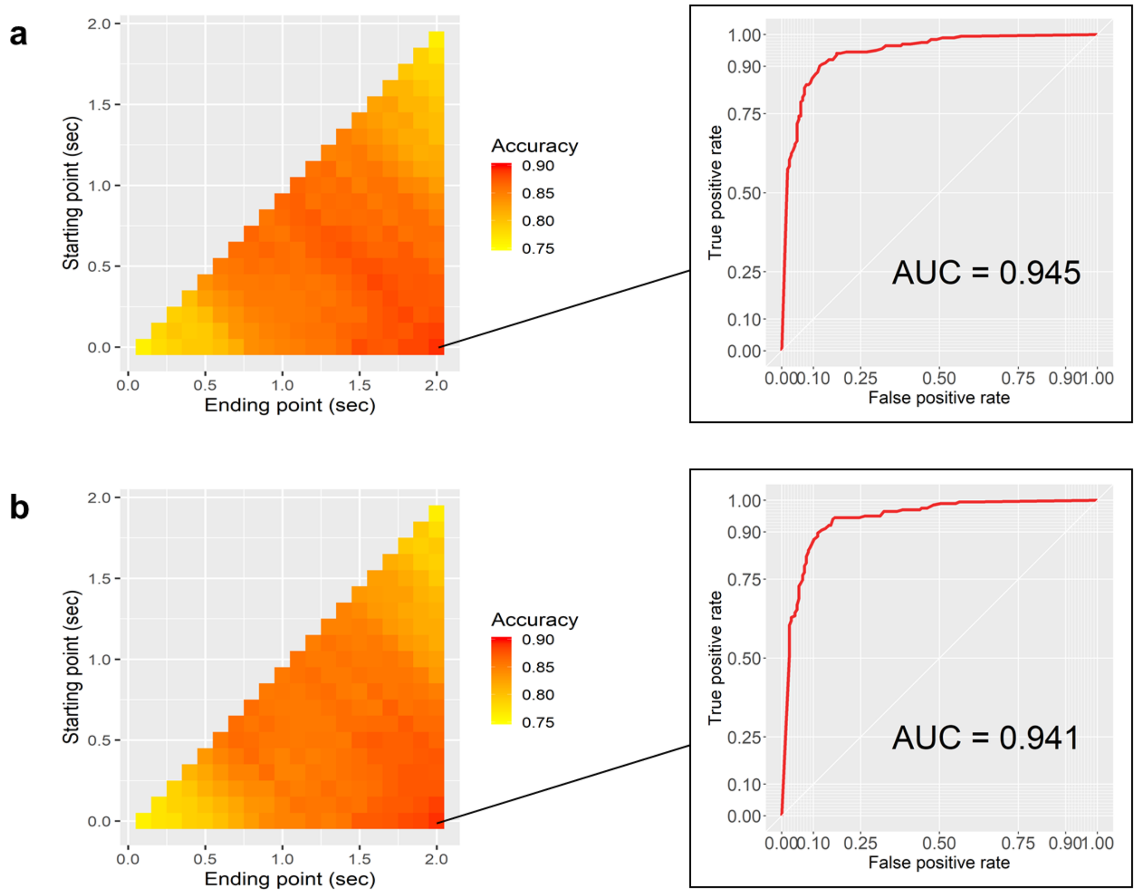

Figure 3.4 Prediction accuracy of the SMV model in each time window and ROC curves of models in the best time window. For the heatmaps, the y-axis indicates the starting point of the target time window, and the x-axis indicates the ending point. The darker red the color, the more accurate the prediction. The ROC curves with the best time window are shown on the right. (a) The heatmap of the accuracy rate in models using face angle trajectories. The highest accuracy was seen when the time window was set from 0.0 to 2.0 s. (b) The heatmap of the accuracy rate in models using gaze angle trajectories. The best time window was the same as in the model using face direction.

3.3 Phase 2: Model validation

3.3.1 Methods

Video recordings

To evaluate the robustness of the created best models, another dataset that was not used in Phase 1 was utilized to conduct the model validation. As described previously, we used webcam-based data in which toddlers' familiar word comprehension was investigated (see

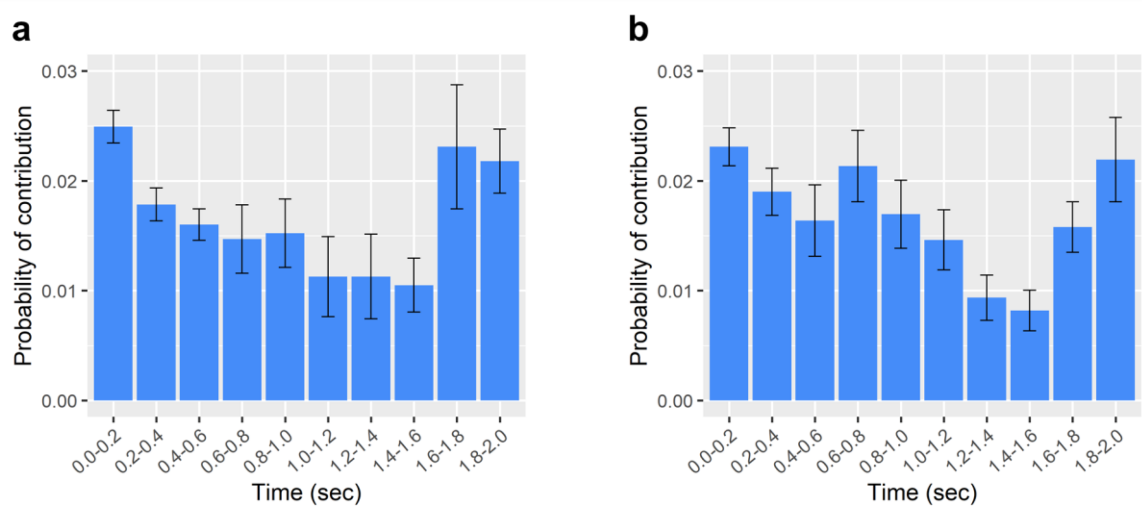

Figure 3.5 Contribution value of each time window to the prediction in the MLB model. The y-axis indicates the probability of contribution of each input to the classification prediction, called "importance," produced by LightGBM. The x-axis indicates the time after the completion of the question summarized every 0.2 s (i.e., 6 frames). The error bar indicates standard error. "Importance" using (a) face angles and (b) gaze angles, respectively.

Chapter 4 for details). For Phase 2, we used videos of 183 trials with 12 monolingual Japanese toddlers aged 18–23 months (6 boys and 6 girls; M_{age} = 20.7, *SD* = 1.8). Nine trials (5 toddlers) were excluded for lack of explicit pointing responses. Each remaining participant responded clearly in 13–16 trials (*M* = 15.3, *SD* = 1.1). The experimental task was the same as in Phase 1.

Face/gaze detection, pre-processing, and model validation

All video recordings were pre-processed as in Phase 1. Namely, we estimated face and gaze angle using OpenFace 2.2.0, identified the head position when a participant looked at the center of the screen, and compensated angles so that the values were always zero when a participant looked at the center of the screen regardless of changing their head position. To predict pointing responses, the pre-processed data were then put into the SMV and the MLB models, which showed the highest accuracy in Phase 1 with fixed parameters. We evaluated these models' prediction accuracy using AUC, accuracy rate, and kappa coefficient.

3.3.2 Results

How robust were the created models?

Using a new dataset that was not used in Phase 1 as input, both the SMV and the MLB models still demonstrated high performance (**Table 1**). In all models, the prediction accuracy of pointing responses was around 86%. AUC showed high accuracy and kappa was substantial.

3.4 Phase 3: Model application to data with no-pointing responses

In this phase, we exploratorily applied the created models to data lacking overt pointing responses to preliminarily investigate whether features of preferential looking with clear pointing were equivalent to those without pointing. Indeed, there were no absolute correct answers; however, by calculating the agreement of pointing estimations from looking behavior between human coders and the created models, and comparing this agreement to the prediction accuracy obtained in Phases 1 and 2, we tried to see if there were qualitative differences between looking responses with and without manual ones. If the agreement in Phase 3 was equivalently high to the prediction accuracy in Phases 1 and 2, it would indicate that looking responses were similar to some extent regardless of pointing execution. However, if the agreement dropped, it would suggest that looking responses without arm movement had some different qualitative features compared to those with clear arm movement.

3.4.1 Methods

Video recordings

For Phase 3, we used videos of 176 trials with 12 monolingual Japanese toddlers aged 18–22 months (6 boys and 6 girls; M_{age} = 19.3, *SD* =1.5) from the same data used in Phases 1 and 2 (see Chapter 4). This sample was extracted from the participants who lacked clear pointing responses for two-thirds or more in all 16 trials so that age and gender were not biased. Another 16 trials with five participants were excluded because they showed explicit pointing behavior. The final data used for analysis included between 11 and 16 trials with each participant (*M* = 14.7, *SD* = 1.8). The experimental task was the same as in Phase 1.

Since there were no absolute correct answers in the videos in Phase 3 (i.e., absent pointing responses), we preliminarily adopted manual annotation for estimating the toddlers' volitional choice from looking behavior as a pseudo-correct index. A trained and a naïve rater independently evaluated which side the participant seemed to choose for all videos based on toddlers' preferential looking behavior. For discrepancies in raters' coding, a third rater annotated participants' choosing; the annotations that were agreed upon by two of the three raters were used. To verify how accurately the human raters can estimate pointing responses based only on preferential looking without seeing the exact pointing, the raters also manually annotated approximately 25% of the data in Phase 1 while participants' pointing responses were masked in the video. Four of each participant's trials from the data in Phase 1 were pseudo-randomly extracted (4 trials × 24 participants = 96 trials) and annotated as in Phase 3. For all manual annotations in Phases 1 and 3, raters also evaluated their degree of confidence in their estimation of pointing using a five-point Likert scale, where 5 indicated highest confidence. These confidence values were calculated by averaging their ratings.

Face/gaze detection, pre-processing, and model application

All video recordings were pre-processed as in Phase 1 using OpenFace 2.2.0. The pre-processed data was then put into the SMV and the MLB models to predict pointing responses. Note that the parameters in these models were the same as those in Phase 2, which means that the models used in Phase 3 were identical to the best models constructed in Phase 1. The agreement of pointing predictions (left or right) between the created models and human rater was evaluated using AUC, accuracy rate, and kappa coefficient. In addition, we defined the confidence value of pointing predictions for each model. For the SMV and the MLB models, the final variable for the prediction ranged from 0 to 1, where 0 indicated left while 1 indicated right, the distance of the value to 0.5 (e.g., $|P_{i,j} - 0.5|$ for the SMV model) was treated as the confidence value. The correlation of confidence values between each model and manual annotation was analyzed using Spearman's rank correlation coefficient.

3.4.2 Results

Manual estimation of pointing responses based only on preferential looking

For 96 trials randomly extracted from Phase 1 data, inter-rater reliability was substantial (κ

= 0.77, 88.5% agreement). Eleven of the trials were coded differently between raters, requiring the third rater to produce a majority opinion. The adopted manual annotation demonstrated almost perfect reliability of correct responses (κ = 0.92, 95.8% agreement). Therefore, it indicated that human raters could reliably estimate pointing responses based only on preferential looking behavior, not on the exact pointing responses. For all 176 trials in Phase 3, inter-rater reliability was also substantial (κ = 0.62, 81.2% agreement).

Agreement between the created models and manual annotation for no-pointing trials

The SMV and the MLB models were applied to the data lacking clear pointing responses (176 trials in Phase 3). When focusing on AUC, the agreement between model and manual based estimation demonstrated high accuracy only in the SMV model using face directions as input, whereas other models showed moderate accuracy despite close values (**Table 1**). The reliability between the SMV model and manual annotation was substantial when using both face and gaze directions, while that with the MLB model was moderate (κ < 0.6).

Correlation of confidence between created models and manual annotation

For 96 trials with clear pointing responses randomly extracted from Phase 1 data, the confidence values in both the SMV and the MLB models showed a significant positive correlation with the manually annotated one, and the correlation in the SMV model was slightly higher than that in the MLB model (**Table 3.2**). For 176 trials with no-pointing responses, the confidence values in all created models were still significantly positive despite being relatively weaker than the values from Phase 1, which ranged from 0.202 to 0.261.

3.5 Discussion

To address the question of how well looking measurements can be interpreted as equivalent to manual ones in two-alternative forced-choice tasks, we investigated how accurately pointing responses (i.e., left or right) could be predicted from concurrent preferential looking in this chapter. Using pre-existing webcam-based data, we created the SVM and the MLB models and tested their prediction abilities. Results showed that toddlers' preferential looking substantially predicted pointing responses, even though looking was only roughly quantified by face or gaze using a webcam.

Table 3.2 Correlation of confidence value between created models and manual annotation

	For trials of clear pointing responses (96 trials randomly extracted from data in Phase 1)		For trials of no-pointing responses (All 176 trials from data in Phase 3)	
	SMV	MLB	SMV	MLB
Face angles as input				
ρ	0.445	0.348	0.202	0.228
p-value	< 0.0001	0.0005	0.0072	0.0023
Gaze angles as input				
ρ	0.433	0.361	0.261	0.213
p-value	< 0.0001	0.0003	0.0005	0.0045

Note. Spearman's rank correlation coefficient was used.

From Phases 1 and 2 using data with clear pointing responses, we found that both the SMV and MLB models showed equivalently high prediction accuracy. This indicates that preferential looking can be interpreted as equivalent to concurrent pointing responses at least to some extent, which is compatible with previous findings that toddlers looked longer at the stimulus that matched their subsequent reaching (Hendrickson & Friend, 2013). Regarding the temporal features of looking responses, we explored the most appropriate time window for the prediction and found that it was from 0.0 to 2.0 s after the completion of the question through the SMV model construction. The result that it was necessary to investigate at least 2 s as a potential time range is compatible with the conventional and empirical method where, in the preferential looking paradigm, the time window of approximately 2-s duration has been utilized for analysis (Delle Luche et al., 2015; Swingley & Aslin, 2000). The fact that the best time window of our results and the conventional one were consistent provides supporting evidence that the traditional naïve setting of time window was a reasonable way of delimiting and summarizing time-series data to reflect toddlers' volitional response to a question.

According to the "importance" visualization produced through the MLB construction, the contribution rate of each input for the prediction showed roughly an inversed U shape in accordance with elapsed time, which indicates there was no critical narrow time window reflecting toddlers' volitional choice in general. This important visualization showed that the best time-bin that contributed to the prediction was 0–200 ms in the MLB model, especially when using face angles. This might be attributed to some methodological errors (e.g., overfitting), or the adjusted starting point of the potential time window to match Japanese grammar. We set the starting point of the potential time window at the end of the question because, as mentioned above, in Japanese, a listener cannot interpret if a sentence is a question until the sentence is heard through the end. The grammatical characteristics of Japanese, in which the target word is positioned at the beginning of the sentence and the interrogative marker comes at the end of the sentence, might have resulted in the highest probability of the contribution in the 0–200-ms time-bin, such that participants were able to start looking at the stimulus during the time it took to produce the utterance from the target word to the interrogative marker. Besides, in this study using LightGBM (Ke et al., 2017), a similar prediction accuracy was obtained from both the SMV and the MLB models. Paradoxically, this fact showed the effectiveness of the traditional method for analysis in which the proportion of total looks to the target stimuli within a certain time window was regarded as a dependent variable. If devising input data, the accuracy of the MLB model might be improved by adding angular velocity or facial expression variables or converting raw time-series data to different ones.

The results from no-pointing trials (Phase 3) revealed that only the SMV model showed high accuracy and substantial agreement to manual estimations of volitional choices although each prediction index of the SMV model was close to the one of the MLB model. Note that in Phase 3, manual estimation was regarded as a pseudo-correct index of pointing; however, some may be skeptical of the validity of the index itself. Indeed, there were no absolute correct answers, but considering human raters could estimate toddlers' choices from data in Phase 1 with almost perfect accuracy without seeing the exact arm movement, manual annotation seemed practically reasonable to use as a pseudo-correct index at a certain level. Overall, the substantial agreement of pointing between manual and model-based estimations indicated that toddlers possibly demonstrated their volitional choices by looking responses even when manual responses were absent. Hence, it would be practically reasonable, to some extent, to use preferential looking as an alternative measure of pointing to avoid wasting collected data due to missing manual responses, at least for children aged 18–23 months. However, considering the prediction indices in Phase 3 dropped compared to ones in Phases 1 and 2, preferential looking without pointing would be qualitatively different from that with overt pointing. Hendrickson and Friend (2013) claimed that looking and manual responses reflect different levels of word understanding, based on connectionist studies (Munakata, 2001; Munakata & McClelland, 2003). Hendrickson et al. (2017) further revealed that, in the familiar word comprehension task, words for which toddlers did not execute reaching at 16 months were still unknown at 22 months, whereas words for which 16-month-olds reached to the distracter rather than the target stimulus were more likely to be known six months later. These findings indicate that the absence of manual responses itself provides insightful information on early word knowledge (e.g., fragility or uncertainty).

For both the SMV and the MLB models, the correlation of confidence value between the created models and human raters remained moderate or weak, which indicated that human raters likely conducted confidence evaluation based not only on mere time-series data of face and gaze angles but also other richer information that could be obtained from videos. Extracting other variables such as facial expression might be beneficial to a more precise prediction of confidence for pointing estimation. Reliable confidence prediction can be utilized practically when estimating children's choices based only on preferential looking behavior automatically. For example, data with high confidence values may be used for subsequent analysis, whereas data with low confidence values may need manual inspection or elimination. Such techniques might be useful in situations where arm responses are accidentally lost in webcam-based videos during the forced-choice pointing paradigm due to the narrow viewing angle of a camera.

Although this was not the main objective of this study, we could confirm the reliability and the usefulness of OpenFace 2.2.0 (Baltrušaitis et al., 2018), which was quite important and necessary to judge if we could rely on time-series data of face and gaze angles produced by this tool. As long as the webcam is positioned at the center of the top of the screen, the screen size is relatively large, and very rough data of preferential looking is enough for analysis, OpenFace 2.2.0 may be a useful and powerful tool for the automatic coding. At this point, it is not clear whether this remains powerful when the webcam is located in a different position, but generally, it seems robust in most challenging conditions (Higuchi et al., 2018). To confirm the availability of these automated coding algorithms is quite beneficial for researchers who have limited resources or are in restricted situations to reduce the burden of data collection. Although eye-trackers can collect and annotate looking responses easily (Ambridge & Rowland, 2013; Delle Luche, et al., 2015) with high temporal and spatial resolution, situations where eye-trackers are available are still restricted because these devices are still expensive and many of them are not handy enough to conduct experiments outside a laboratory setting (e.g., nursery schools or online experiments). In contrast, frame-by-frame manual coding is labor-intensive and time-consuming (Friend & Keplinger, 2008). High-cost data collection regarding time and money can be an obstacle to the promotion of open science, such that researchers who have fewer resources cannot collect each data sample or participate in larger international projects (Frank, 2019). The usefulness of OpenFace 2.2.0 shown in this study would be helpful to overcome such obstacles as evidenced by recent vigorous studies on webcam-based data collection (Chouinard et al., 2019; Papoutsaki et al., 2016; Scott, Chu, & Schulz, 2017; Scott & Schulz, 2017; Semmelmann, Hönekopp, & Weigelt, 2017; Tran, Cabral, Patel, & Cusack, 2017).

Taken together, this study revealed that looking and manual measurements could be interpreted analogously. Hence, it would be practically reasonable to use preferential looking in the forced-choice pointing tasks as a dependent variable instead of manual ones when the unpredicted or unignorable amount of trials lacking pointing responses is observed. However, looking measurements with and without manual ones may have different features that reflect different language abilities. Therefore, it may be recommended and beneficial that both results from pointing and looking measures, and the proportion of absent pointing responses should be reported to interpret obtained data in more detail. In fact, some researchers conducted a similar task using both modalities as dependent variables and compared differences between them (Abbot-Smith et al., 2017; Childers, Porter, Dolan, Whitehead, & McIntyre, 2020; Gurteen et al., 2011; Hendrickson & Friend, 2013; Hendrickson et al., 2015, 2017; Suanda, Walton, Broesch, Kolkin, & Namy, 2013). Another practical strategy would be to estimate and interpolate missing pointing responses (i.e., left or right) from preferential looking, leveraging the findings that both manual and looking measurements have overlapping underlying information on toddlers' volitional choice and the former measures are substantially predicted from the latter ones regardless of the execution of pointing or reaching. This strategy will help researchers prevent from wasting or excluding the collected data. However, researchers adopting this option should confirm that the results from manual measures with and without their interpolation from looking measures are compatible with each other. Researchers may have to consider what is implied by absent manual responses since such absences themselves may reflect toddlers' knowledge states about word comprehension.

Overall, we succeeded in accomplishing the central purpose of this chapter, namely, to lay the groundwork for applying the experimental task used in Chapter 2 to experiments involving 18–23-month-olds toddlers without missing data due to the lack of volitional arm responses. In the next chapter, we will adopt both preferential looking and pointing responses as indices reflecting task performance on the semantic differentiation of object words. Based on the findings of the current study, we will use the SMV model to discretize looking data so that we can analyze both looking and pointing data in the exact same way and compare the results from them directly.

For future research, we raise several limitations and suggestions on how to address them. First, it is not obvious whether the created models in this study can be applied to other studies using the intermodal preferential looking or the forced-choice pointing paradigm due to a single experiment used for analysis. To construct more general models, it will be necessary to explore if there is a change in prediction accuracy or visualized importance tendency as a result of extending the range of the potential time window. For instance, a more appropriate time window may be found outside of the 2-s duration, since the best time window was explored only within this time interval in this study. Future studies should apply the models to other experimental tasks. In addition, comparing the differences between setting a starting point at the onset of the target word and at the end of the question would be beneficial to determine which is more important in choosing action: listening to the target word while understanding that the sentence is a question or only listening to the target word. By developing software that can easily create models from other datasets using the method proposed in this study as in a related previous work (Kominsky, 2019), other researchers can examine these questions.

Second, although we found that preferential looking while executing pointing responses were closely related to pointing itself, it remains unclear whether this relationship can be generalized to preferential looking irrelevant to the execution of arm movements. To directly address this, an experimental design that allows separating looking and manual responses is needed, such as asking toddlers to only look at the target stimulus (e.g., “Look! There are shoes!”) then asking them to point at it at least two seconds later (e.g., “Which ones are the shoes? Point to them!”). Relatedly, researchers should investigate a more nuanced relationship between looking and manual measurements. For adults or older children, this relationship has been scrutinized using more sophisticated models such as drift-diffusion models (Ratcliff, Love, Thompson, & Opfer, 2012; Thomas, Molter, Krajbich, Heekeren, & Mohr, 2019). However, since model fitting using drift-diffusion models requires numerous trials per participant (Lerche, Voss, & Nagler, 2017), there is only one study, to our knowledge, that applied them to toddlers (Leckey et al., 2020). A methodological improvement is needed to apply these models to young children.

Third, there were some prediction errors despite the substantial accuracy of created models. Roughly speaking, the trials in which the predictions were incorrect had any of the following features: face and gaze direction mostly remained around the center of the monitor for the target time window; pointing responses occurred within first 1 s and then the participant looked at the other stimulus; or, right up to pointing, participants looked at the side of a stimulus opposite the one chosen (some examples were described in **Appendix C**). Since some of these features can be dealt with by predicting the timing of pointing responses, it may be beneficial to explore the possibility of estimating reaction times for each trial. Additionally, more precise confidence quantification may contribute to improving the prediction accuracy of pointing responses from preferential looking. In relation to the creation of more reliably predicting models, methodological research on estimation and interpolation of missing pointing responses from looking ones may be another interesting direction for research. For example, future studies could verify which interpolation method reflects the true result between the created models in this study and usual statistic techniques that estimate missing data by creating artificial missing data to evaluate how the models proposed here are useful for missing data interpolation.

Further investigation of the relationships between manual and looking responses is outside the scope of the main purpose of this thesis, which is to examine the semantic pluripotency hypothesis. Therefore, we will not step more into the limitations discussed here in the following chapters. Yet, the study addressed in this chapter has presented fruitful future directions of research from various points of view. Despite several limitations, we believe that this study bridges two different measurements (i.e., looking and manual) in two-alternative forced-choice tasks practically and theoretically. Since two-alternative forced-choice tasks have been widely used in research, not only related to early language development but also to other cognitive domains, the findings of this study will help researchers avoid wasting collected data and allow them to interpret results from different modalities more comprehensively. Additionally, in terms of webcam-based data collection, this study will contribute to conducting online experiments as in related studies (Chouinard et al., 2019; Papoutsaki et al., 2016; Scott et al., 2017; Scott & Schulz, 2017; Semmelmann et al., 2017; Tran et al., 2017) and offer further promotion of the open science movement (Frank, 2019).

3.6 Summary of this chapter

In the two-alternative forced-choice paradigm, manual responses, such as pointing, have been widely used as measures to estimate cognitive abilities. While pointing measurements can be easily collected, coded, analyzed, and interpreted, absent responses are often observed, particularly when adopting these measures for toddler studies, which leads to a growth in missing data. Although looking responses, such as preferential looking, can be available as alternative measures in such cases, it is unknown how well looking measurements can be interpreted as equivalent to manual ones. In this chapter, we aimed to answer this question by investigating how accurately pointing responses (i.e., left or right) could be predicted from concurrent preferential looking. Using pre-existing videos of toddlers aged 18–23 months engaged in an intermodal word comprehension task, we developed models predicting manual responses from looking ones. Results showed substantial prediction accuracy for both the Simple Majority Vote and the Machine Learning-Based classifiers, which indicates that looking responses can be reasonable alternative measures of manual ones. However, the further exploratory analysis revealed that when applying the created models on the data of toddlers who did not produce clear pointing responses, the estimation agreement of missing pointing between the models and the human coders dropped slightly. This indicates that looking responses without pointing were qualitatively different from those with pointing. Bridging the two measurements in forced-choice tasks would help researchers avoid wasting collected data due to the absence of manual responses and allow them to interpret results from different modalities comprehensively.

Chapter 4

When "Shoe" Becomes Free From "Putting On"

Based on:

Hagihara, H., Yamamoto, H., Moriguchi, Y., & Sakagami, M. (2020). When "shoe" becomes free from "putting on": Initial meanings of object words are intertwined with object-specific actions. Manuscript in progress.

Based on the findings obtained in Chapters 2 and 3, this chapter addresses the semantic pluripotency hypothesis more directly and deeply. After quickly summarizing backgrounds on early word learning again, we examine the robustness of the results of our preliminary study (Chapter 2) using both pointing and looking responses. Furthermore, we investigate how tightly object-specific actions are entwined with object word meanings and how developmental changes in object word meanings are related to vocabulary growth.

4.1 Introduction

When categorizing children's early vocabulary into parts of speech, nouns hold the dominant position in many languages, compared to verbs (Bornstein et al., 2004; Caselli, Casadio, & Bates, 1999; Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, in press; Gentner & Boroditsky, 2001), except for some verb-friendly languages (Brown, 1998; Choi & Gopnik, 1995; Tardif, 1996; Tardif et al., 2008; Tardif, Gelman, & Xu, 1999). Traditionally, noun learning has been contrasted with verb learning to elucidate the universal characteristics of early language development (Gentner, 1982; Gentner & Boroditsky, 2001; Gleitman & Gleitman, 1992; Gogate & Hollich, 2016; Goldin-Meadow, Seligman, & Gelman, 1976; McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011;

Twomey & Hilton, 2020; Waxman et al., 2013). Since early nouns semantically correspond to object categories (e.g., “shoe”) and early verbs correspond to action categories (e.g., “put on”) (Gogate & Hollich, 2016), researchers have been motivated to explain, both theoretically and empirically, why such “object words” are advantageous for learning over action words. Gentner (1982), proposed the “natural partitions hypothesis” in which concrete objects can be more easily individuated from the surrounding environment than actions or other features because concrete objects are temporally and spatially invariant. This invariant, static feature of objects facilitates children’s object word learning. The effects of static features that objects have on label mapping onto object categories are well documented. One of the most well-known principles is called the “shape bias,” according to which children are likely to generalize the same label to objects with similar shapes (Gershkoff-Stowe & Smith, 2004; Graham & Poulin-Dubois, 1999; Imai, Gentner, & Uchida, 1994; Kucker et al., 2019; Landau, Smith, & Jones, 1988, 1998; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Yee, Jones, Smith, 2012).

However, both static and dynamic information relevant to objects, such as object-specific actions, are important for learning object categories (Booth, 2006; Booth, Schuler, & Zajicek, 2010; Hernik & Csibra, 2009; Horst, Oakes, & Madole, 2005; Madole, Oakes, & Cohen, 1993; Perone & Oakes, 2006; Träuble & Pauen, 2007, 2011) and object words (Booth & Waxman, 2002; Kemler Nelson, 1995; Kemler Nelson, Frankenfield, Morris, & Blair, 2000; Kemler Nelson, Russell, Duke, & Jones, 2000; Kobayashi, 1997, 1998; Nelson, 1973a; Ross, Nelson, Wetstone, & Tanouye, 1986). Like shape bias, toddlers can generalize the referent of novel labels based on object-specific actions or object functions. For example, teaching a new label while emphasizing a particular action enabled two-year-old toddlers to generalize the same name to objects with the same function, despite different shapes (Kobayashi, 1997). Knowledge about object-specific actions even scaffolds the later establishment of shape bias for 17-month-olds (Ware & Booth, 2010). Moreover, object-action connections are easily learned, compared to object-word connections for 12-month infants (Deng & Sloutsky, 2015; Eiteljoerge, Adam, Elsner, & Mani, 2019b) and toddlers (Childers & Tomasello, 2002; Hahn & Gershkoff-Stowe, 2010). Although actions are basically referred to by verbs, they play a significant role even in early noun learning. This implies that the initial meanings of object words might not only comprise objects with particular static information (e.g., shape, appearance) but also actions.

Some researchers, based on theoretical and observational assumptions, argued that early words have distinctive label-meaning connections, such that they cannot be easily classified into specific parts of speech, such as nouns or verbs (Church, 1961; Dewey, 1894;

Lewis, 1936; McCune, 2008; Okamoto, 1982; Stevenson, 1893; Tomasello, 2003; Volterra, Bates, Benigni, Bretherton, & Camaioni, 1979; Vygotsky, 1934/1986; Werner & Kaplan, 1963). Werner and Kaplan (1963) claimed that early labels encompassed holistic, global, and undifferentiated events or activities that were strongly related to children’s experiences in everyday contexts, rather than specific object or action categories themselves. They, therefore, designated these label-meaning connections as “names” or “monoremes,” instead of “words.” For example, young children’s utterance “ball” would initially represent an entire ball-play activity, rather than a strict and specific object “ball” or action “rolling,” at least during the holophrastic stage, and it is only when the child begins to utter two words (e.g., “ball rolling”), that the meaning of the word “ball” can be interpreted as a specific object category. Nelson (1983a, 1983b, 1986) also suggested that events in which objects and actions were embedded were children’s initial concepts. The assumption that the formation of event categories preceded that of specific object categories was supported by observational findings, as early vocabulary consisted not only of object words but also of other words referring to places, events, or actions that were tied with particular contexts such as “park,” “lunch,” or “kiss” (Nelson, Hampson, & Shaw, 1993). Moreover, words uttered in distinctive contexts in terms of spatial, temporal, or linguistic aspects were learned earlier, regardless of grammatical types (Roy, Frank, DeCamp, Miller, & Roy, 2015). However, the observational approach alone cannot directly investigate whether the initial meanings of words are entire undifferentiated events, as Werner and Kaplan (1963) hypothesized, or specific categories, such as objects or actions, because even if a word corresponds to an undifferentiated meaning, its phonetic features are indistinguishable from those of an ordinary word that corresponds to specific meanings.

To experimentally address this question, we proposed the “semantic pluripotency hypothesis” (see Chapter 1), in which we posited that the initial meanings of object words were undifferentiated global event categories including both objects and actions as a fusion and that they subsequently differentiate into specific object categories independent of actions. In Chapter2, we preliminarily investigated this hypothesis, using a forced-choice pointing task for Japanese monolingual toddlers aged 19–35 months. Toddlers watched two juxtaposed video stimuli, where a girl was doing an action with certain objects and chose one of them according to questions about familiar object words (e.g., “Which ones are shoes?”). We found that even young participants could choose the appropriate stimulus when objects and actions were compatible on one side (e.g., “putting shoes on” as the target vs. “rubbing two baskets” as the filler in the match condition). In contrast, when objects and object-specific actions were incompatible (e.g., “rubbing shoes in front of a girl’s chest” vs. “putting on two baskets as if they were shoes” in the mismatch condition), the probability of choosing a stimulus with target objects remained at chance level for toddlers younger than 21 months or those with a vocabulary size of fewer than 140 words, calculated using the Japanese MacArthur-Bates Communicative Development Inventory (J-MCDI; Ogura & Watamaki, 2004), whereas the probability drastically increased for older participants or those with a larger vocabulary size. These findings supported the semantic pluripotency hypothesis, suggesting that the initial meanings of object words were tied with object-specific actions, being global and undifferentiated event categories in this sense (e.g., "putting shoes on"), while they subsequently differentiated into specific object categories independent of actions (e.g., "shoes" alone).

Our previous study provided the first experimental evidence for the semantic pluripotency hypothesis; however, due to the preliminary study, these findings were not only uncertain in the reproducibility but also had two major limitations that required further investigation. First, although both object and action categories are likely intertwined in the initial meanings of object words, it is still unclear whether toddlers can comprehend object word referents solely based on object-specific actions. Researchers who held that children's early concepts were more holistic than those of adults, argued the importance of object-specific actions in object word learning (Kobayashi, 1992; McCune, 2008; Nelson, 1973a, 1974; Werner and Kaplan, 1963). Even Werner and Kaplan (1963) stated that young children often assigned the same label to phenomena, based on how objects are used in a context (p. 117). To clarify the role of such actions in object word comprehension, it would be beneficial to add other conditions to the forced-choice task, such as presenting two stimuli of doing the object-specific or the filler action using the filler object (e.g., "putting baskets on" vs. "rubbing baskets" in the different-object condition). The addition of these new conditions is also helpful for interpreting the results of the previous study more precisely. Although the probability of choosing the target stimulus remained at chance level in younger toddlers, when target objects and object-specific actions were shown separately (i.e., the mismatch condition), there were two possible interpretations, since the two options in this condition competed with each other (e.g., displaying "shoe" on one side and "put on" on the other). One possibility, called the *holistic-fusion-based* explanation, was that toddlers failed to properly recognize the object word referents under the situation where an object and an object-specific action were not matched because the initial meanings of such words were the undifferentiated fusion of both objects and actions. The other possibility, called the *action-weight-based* explanation, was that toddlers were more likely to choose the stimulus, including object-specific actions, because they placed relatively great weight on actions in object word comprehension. If younger toddlers can judge object word referents solely by object-specific actions, the action-weight-based explanation would be appropriate for interpreting the results in the mismatch condition. If the proportion of choosing the stimulus with the object-specific action remains at chance level in a new condition, the holistic-fusion-based explanation would be the appropriate one for the interpretation of the previous results.

Second, a more in-depth investigation should be conducted to determine how the semantic differentiation of object words relates to vocabulary growth. Some researchers have attributed slower production of verbs than of nouns to the difference in the difficulty in determining accurate referents, such that action categories are more difficult to form than object categories (e.g., Gentner, 1982; Imai, Haryu, & Okada, 2005; Waxman et al., 2013). However, based on the semantic pluripotency hypothesis that the initial meanings of object words correspond to undifferentiated event categories, another explanation would be possible. When object word meanings encompass both object and action categories, verbs might not be necessary because children can express both objects and actions by object words alone. Subsequently, once object words are differentiated into specific categories, children would be encouraged to develop new labels that can depict specific actions, which results in verb vocabulary growth. Observational studies have attempted to identify the occurrence of such semantic differentiation by focusing on children’s two-word utterances, because when two words are used in combination (e.g., “doggie gone”), it is obvious that each word refers to different specific categories (e.g., McCune, 2008; Werner & Kaplan, 1963). In contrast, the experimental task used in Chapter 2 will enable us to investigate the relationship between semantic differentiation of object words and verb vocabulary growth. Specifically, we posited that task performance in the mismatch condition, which evaluates toddlers’ ability to recognize an object as a referent of object words regardless of how it is used, would relate to concurrent or later action words vocabulary size, compared to other conditions.

Using both cross-sectional (Study 1) and longitudinal (Study 2) data for 18–23 month old toddlers, we examined (1) the robustness of the findings that early noun meanings developed from undifferentiated events to specific object categories, (2) whether toddlers could comprehend object word referents based solely on object-specific actions, and (3) how semantic differentiation of object words related to vocabulary size (**Figure 4.1**).

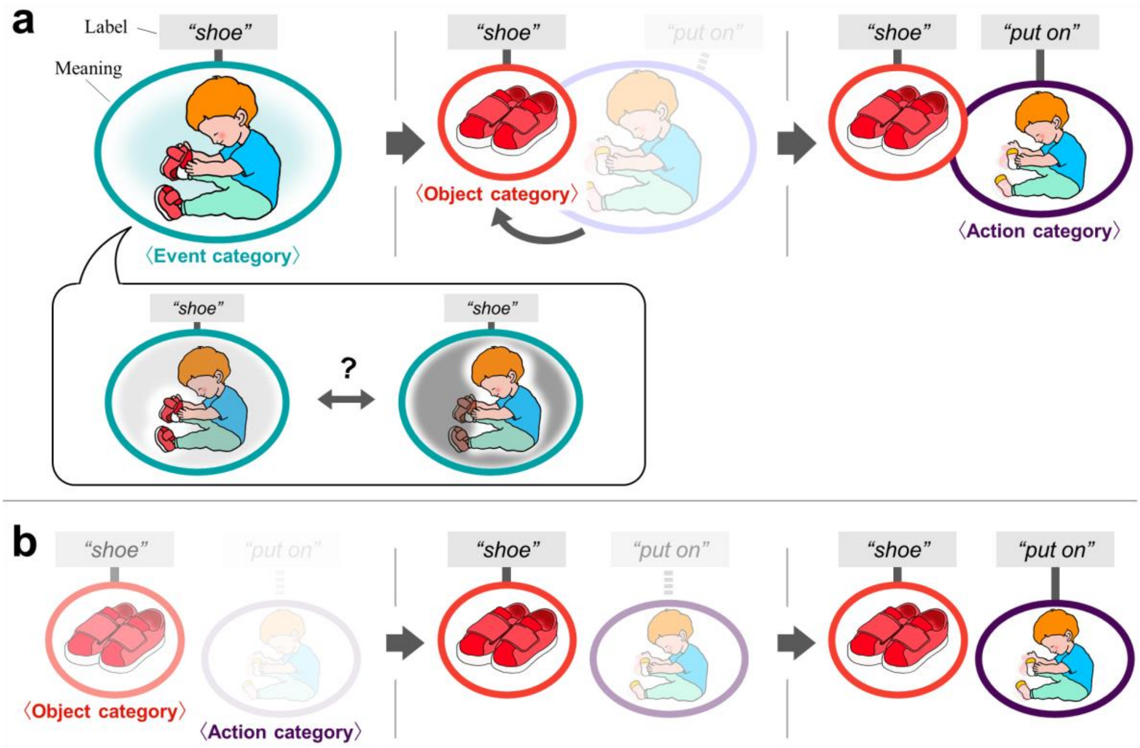

Figure 4.1 A schematic view of the hypothesis examined in Chapter 4. (a) The initial meanings of object words are global event categories that consist of both object and object-specific actions, and they subsequently differentiate into specific object categories. This semantic differentiation encourages children to develop new label-meaning connections to express specific actions, which results in the growth of the action words vocabulary. The balloon illustrates two possible hypotheses to be examined in this study (which we called the *holistic-fusion-based* explanation and the *action-weight-based* explanation). (b) The conventional view of early word learning. Object words inherently correspond to specific object categories from when first learned, and they become more precise with development. Action words are learned more slowly than object words due to difficulty in determining accurate referents.

4.2 Study 1: A cross-sectional study

4.2.1 Method

Participants

Sixty-nine Japanese monolingual toddlers aged 18–23 months were included in the final analyses, such that each age group included at least ten participants and the gender ratios were approximately equal (34 boys and 35 girls; M_{age} = 20.5 months, SD = 1.8; **Figure 4.2**).

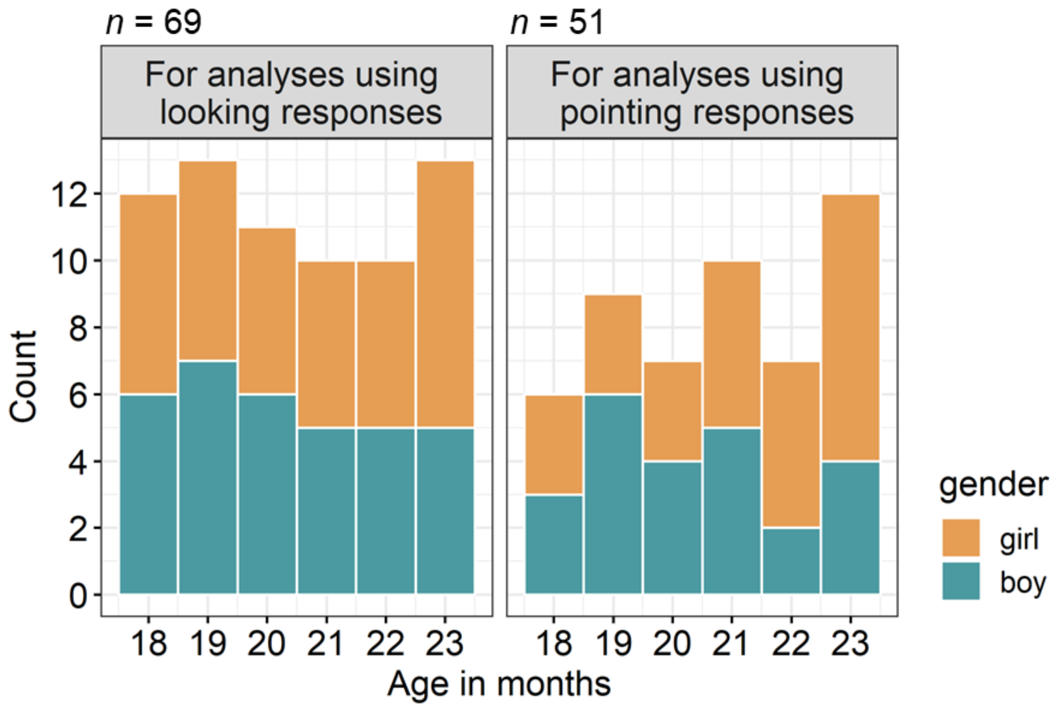

Figure 4.2 Demographic information on participants included in the final analyses in Study 1. Eighteen participants were excluded from analyses using volitional pointing responses as an index for task performance because of pointing to only one side for all trials (i.e., side bias; *n* = 8) or of complete absence of pointing responses (*n* = 10).

All participants were recruited from 14 nursery schools. An additional 13 toddlers were excluded from data analysis due to difficulty in engaging in experimental tasks due to inattention (*n* = 8) or crying (i.e., fussiness; *n* = 3), falling outside the target age range (*n* = 1), or being raised in a bilingual environment (*n* = 1). We asked parents to complete the J-MCDI (Ogura & Watamaki, 2004) to estimate toddlers’ expressive vocabulary size. All participants were thought to be typically developed given none of them had received developmental support, such as medical habilitation. The sample size was determined based on recent previous studies using similar experimental tasks and measurements (Ferguson, Graf, & Waxman, 2014; Gogate & Maganti, 2017; Imai et al., 2015), considering the previous findings that there are roughly two developmental phases on the semantic contents of early object words, from global to specific categories (see Chapter 2). Written consent was obtained from the parents of all participants in advance. This research was conducted according to guidelines laid down in the Declaration of Helsinki and approved by the ethics committee for human and animal research of the Graduate School of Human and Environmental Studies at Kyoto University.

Stimulus materials

We adopted the two-alternative forced-choice task used in Chapter 2, during which children were prompted to choose one of the side-by-side presented video stimuli following the question about familiar object words. For the target object words, “shoes” and “cup” were selected in the task because these words are learned early in language development (Ogura, Watamiki, & Inaba, 2016), are familiar to children in everyday lives, and these objects have particular object-specific actions such as “putting on” and “drinking with.” We call these actions target-specific actions. A pair of baskets and a toy shovel were assigned to filler objects corresponding to the shoes and cup, respectively. These filler objects differed in shape from the target objects: unlike the shoes, the baskets had a vertically oriented handle and were symmetrical, and, unlike the cup, the toy shovel had a linear handle and was flat and asymmetrical. For filler actions that were meaningless and distinctively different from the target-specific actions, “rubbing in front of self” and “making a circular motion on a table” were assigned to the shoes and cup, respectively.

In our study in Chapter 2, there were two conditions: match and mismatch. For the match condition, a video stimulus where a girl was doing the target-specific action with the target object was juxtaposed with one of the filler action performed using the filler object (e.g., “putting shoes on” vs. “rubbing two baskets in front of her”). For the mismatch condition, a video stimulus where a girl was doing the filler action with the target object was juxtaposed with one of the target-specific actions performed using the filler object as if it were the target object (e.g., “rubbing shoes in front of her” vs. “putting on two baskets”).

In addition to these two conditions, we added the different-object and absent-object conditions to examine whether toddlers could comprehend object word referents based solely on object-specific actions at two different levels. For the different-object condition, video stimuli where a girl was doing the target-specific or the filler action with the filler object were juxtaposed (e.g., “putting on two baskets” vs. “rubbing baskets in front of her”). For the absent-object condition, video stimuli of the pantomimes of the target-specific or the filler actions without any object were juxtaposed.

Each video stimulus had a duration of approximately 7 s and ended just before the actions were completed (**Figure 4.3**). There were four test stimuli per condition: two target objects (shoes and cup) and two presented sides where the correct stimulus was displayed.

Apparatus and Procedure

All participants engaged in the task individually in a vacant space at each nursery school, sitting on a small chair or on the lap of a nursery school teacher. Teachers were refrained

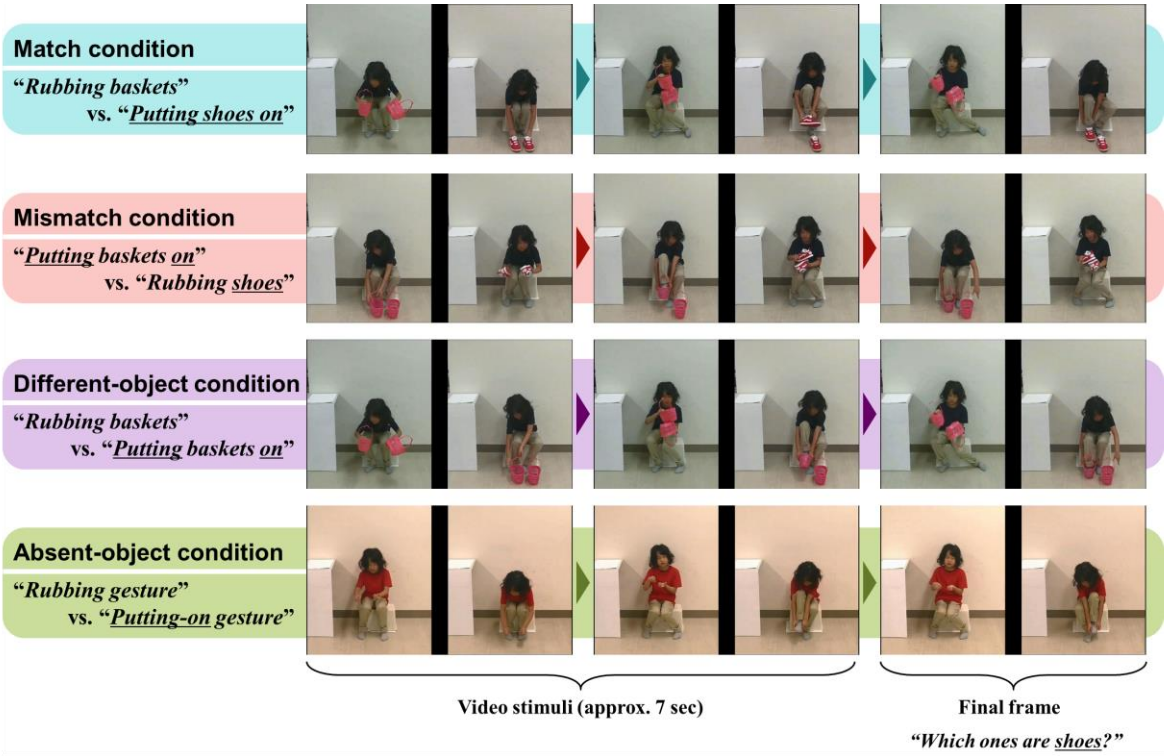

Figure 4.3 An example of the forced-choice task used in this study. A pair of stimuli was presented side-by-side according to the condition. On the final frame, participants were asked to choose the stimulus that matched the target object words (e.g., “Which ones are shoes?”). In this example, the stimulus on the right side was regarded as the correct answer. Underbars indicate the target objects and target-specific actions.

from providing any task cues such as utterances, gestures, eye contact, or facial expressions relating to the video stimuli. Toddlers were prompted to look at the 21.5 in touch screen (490 × 243 mm) with a viewing distance of approximately 30 cm. Two cameras were set to record toddlers’ responses for later coding (30 frames/second): One was placed in an overhead position so that participants’ entire body could be seen; the other (a webcam) was set at the center of the top of the screen so that participants’ preferential looking could be recorded. The forced-choice task was conducted so that both pointing and looking responses could be obtained concurrently, as seen in previous studies (Childers, Porter, Dolan, Whitehead, & McIntyre, 2020; Hendrickson & Friend, 2013; Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Hendrickson, Poulin-Dubois, Zesiger, & Friend, 2017) because toddlers in our target age range often showed an absence of volitional pointing, which possibly increased missing data. The forced-choice task required approximately 10 minutes to complete. All participants completed sixteen test trials where all four conditions were included (i.e., within-participants design).

Warm-up trials. Before the test trials, participants were prompted to look at small icons, such as a dog or a car at the center and each corner on the screen, which was for later calculation of gaze angles to the monitor. Subsequently, participants engaged in three warm-up trials to understand the task rule. In the first trial, the juxtaposition of a stuffed dog and a toy car was displayed on the screen. The experimenter asked "*Wanwan/Bubu wa docchi?*" [Which one is a dog/car?] (*Wanwan* or *Bubu* is Japanese baby talk for dogs or cars, respectively) to participants. The experimenter encouraged participants by saying "*Sou dane*" [Yes, it is] when participants pointed to or explicitly looked at the correct side. The experimenter corrected participants by pointing the correct side and saying "*Wanwan/Bubu wa kocchi dane*" [The dog/car is this side] when they pointed to or explicitly looked at the incorrect side, or gave unclear responses. In the second trial, a pair of short videos in which a woman was eating or changing clothes were displayed while asking a question of "*Gohan-pakupaku/Okigae wa docchi?*" [Which one is eating/changing clothes?] (*Gohan* is a Japanese conventional noun and *Pakupaku* is Japanese baby talk for eating or food, while *Okigae* is a Japanese conventional noun for changing clothes). This trial was inserted to encourage children to pay attention not only to objects but also to actions. The third trial consisted of short videos in which a man was moving a stuffed dog or a toy car. The question was the same as in the first trial.

Test trials. Each trial started after a presentation showing a center gaze point for one second with a chime sound. Immediately after the video stimuli stopped, which was the moment just before the actions were completed, the experimenter asked "*O wa docchi?*" [Which one is the *O*?] where *O* represents the target object word, either "*Kutsu*" [shoes] or "*Koppu*" [a cup]. Note that the Japanese do not distinguish between singular and plural forms. There are two different expressions for shoes in Japanese depending on regions or environments: one is the conventional word "*Kutsu*" and the other is baby talk "*Kukku*." Hence, the experimenter asked nursery school teachers in advance how they usually called the target objects when talking to children, and adopted that expression during the experiment. Among the 69 participants, "*Kutsu*" was used for 49 (71%) and "*Kukku*" was used for 20 (29%) toddlers. The experimenter moved to the next trial when participants showed pointing responses or at least two seconds after the completion of the question for participants who did not show pointing responses. If a participant pointed within two seconds after the question, the experimenter waited for at least two seconds then moved to the next trial. The potential time window of 2 s was chosen because at a later time, looking behavior was no longer considered to be related to the stimulus (Delle Luche, Durrant, Poltrock, & Floccia, 2015; Swingley & Aslin, 2000) and there was a time constraint for collecting data at nursery schools. The final video frame was left on the screen until the next trial started.

Each participant engaged in 16 test trials. The first two trials were the match condition, one of them was with shoes, and the other was with a cup. The order and presented side were pseudo-randomized. We showed two trials on the match condition first, in which objects and actions were compatible, to avoid toddlers thinking of filler objects as the exact target objects (i.e., baskets as shoes or a toy shovel as a cup). In the first eight trials, participants engaged in a series of trials on all conditions and target objects. The presentation side was pseudo-randomized. Except for the first two trials of the match condition, the order of trials in the remaining six trials was also pseudo-randomized. The latter eight trials were conducted in the same order as the first eight trials, with the presentation sides switched. After the 4th, 8th, and thereafter every two trials, short videos unrelated to the experiment were presented for about 10 seconds to prevent toddlers from becoming bored or losing interest in the task.

After the forced-choice task, toddlers engaged in the other two cognitive tasks evaluating children’s abilities in working memory and pretend actions; however, we do not report them here because these tasks were out of the scope of the current research.

Data analyses

Indices and coding. Based on Chapter 3’s recommendation that both results from pointing and looking measures should be reported to interpret obtained data in more detail, we used these two responses as indices reflecting task performance. Using part of the data obtained in this study, we investigated how well preferential looking could be interpreted as equivalent to pointing responses and explored the most appropriate time window to optimize the prediction accuracy of pointing from looking responses (see Chapter 3). We revealed that the stimulus at which toddlers looked longer in a certain time window for each trial was almost identical to the one they pointed to, and this prediction accuracy was highest when the target time window was set between 0.0 and 2.0 s immediately after the question ended (89.2% agreement; the Simple Majority Vote model). Therefore, we adopted this majority vote method in which the stimulus with the larger proportion of looking during the two-second period was regarded as participants’ choice. This method allowed us not only to prevent the missing data from increasing, but also to compare the results obtained from looking responses with those from pointing responses directly.

Following the procedure explained in Chapter 3, we pre-processed the webcam data to automatically estimate toddlers' horizontal gaze angles using an open-source library, OpenFace 2.2.0 (Baltrušaitis et al., 2018), and compensating gaze angles so that the values were always zero when children looked at the center of the screen, regardless of head position changes. These processes enabled us to automatically annotate whether toddlers looked right, left, or away. A trained coder manually annotated looking responses frame-by-frame (i.e., looking left or right, and looking at or away from the screen) for approximately 20% of the data (222 trials) to confirm the reliability of the automated estimation of preferential looking. For left or right annotation, "almost perfect" reliability based on Landis and Koch (1977) criterion was obtained (κ = 0.82, 91.0% agreement). Hence, we continued using automated estimation. However, for looking-away annotation, the automated annotations were 4.5 times more numerous than the manual annotations, which indicated that the range of estimated gaze angles when looking at the screen was too narrow in OpenFace 2.2.0. Thus, a trained coder manually checked all frames that were annotated as looking away in the automated method and corrected the annotation if necessary. A total of 73.5 % of the frames were corrected from looking away to looking at the screen and manual annotation was adopted. Finally, coding for all frames was integrated and classified into three categories: looking right, left, or away.

We calculated the proportion of looking at the correct stimulus for each trial. For the match and mismatch conditions, we regarded the stimulus that contained the exact target objects (i.e., shoes or a cup) as the correct answer. For the different-object and absent-object conditions, we regarded the stimulus that contained target-specific actions (i.e., putting on or drinking) as the correct answer. The proportion of looking at the correct stimulus was calculated by dividing the total frames when looking at the correct stimulus by the total frames when looking at the screen within a two-second time window immediately after the question ended (i.e., correct stimulus / (correct + filler stimuli)). Based on the Simple Majority Vote method in Chapter 3, a trial in which the proportion of looking at the correct stimulus was greater than 0.5 was regarded as the correct response. If the proportion was equal to 0.5, we defaulted the correct response. This way, we calculated the number of correct responses from preferential looking.

Pointing responses were also evaluated by a trained coder to later investigate the equivalence of the results of looking and pointing measures. A naïve coder independently evaluated the pointed side for approximately 25% of the data (272 trials) with clear pointing responses, which showed almost perfect inter-rater reliability (κ = 0.96, 97.8% agreement). The number of cases of pointing to the correct stimulus was calculated.

As a developmental index, we used chronological age (months). We also used the total vocabulary size, calculated by summing check items from the J-MCDI as a supplemental measurement reflecting children’s linguistic ability. Furthermore, we calculated the number of common nouns (Caselli et al., 1999; Ogura et al., 2016) that consisted of subgroups of animals, vehicles, toys, food and drink, clothing, body parts, furniture and rooms, and small household objects on the J-MCDI (maximum of 281 items) as well as the number of verbs that consisted of action words (maximum of 103 items). The numbers of common nouns and verbs represented the toddlers’ vocabulary size of object and action words, respectively.

Data rejection. None of the participants looked only at one of the two sides (left or right) of the monitor for all 16 trials. Three individual trials from three participants were excluded because they did not look at the monitor at all within the target time window. Regarding pointing responses, eight participants pointed to only one side for all trials in which they showed pointing, and an additional ten participants did not show clear pointing responses at all. Thus, they were excluded from later analysis using pointing responses. The final sample size was 51 (24 boys and 27 girls; M_{age} = 20.8 months, SD = 1.7; see **Figure 4.2** for demographic information).

Bayesian statistical modeling. We conducted three analyses using Bayesian hierarchical generalized linear models to estimate the effects of developmental indices and conditions on task performance (Analyses 1 and 2) and the effect of age on vocabulary growth (Analysis 3). For Analysis 1, we focused on the match and mismatch conditions to investigate toddlers’ understanding of object words as global event categories or as differentiated object categories. The *preferential looking model* used the number of correct responses calculated from looking responses for all trials within the same participants and conditions as the dependent variables, while the *pointing model* used the number of pointing to the correct stimulus per the number of trials with pointing within the same participants and conditions. Both models were fitted using a binomial distribution and a logit link function. Developmental indices (age in months as the primary and total vocabulary size as the supplementary measure), conditions, and interaction between developmental indices and conditions were used as fixed effects. We chose a hierarchical model using a random intercept to consider individual differences, which accounted for the within-participants effect. Weakly informative Student-*t* priors ($\nu = 4,\ \mu = 0,\ \sigma = 0.5$) were used for the random intercepts to stabilize the parameter estimates. For Analysis 2, we focused on the different-object and absent-object conditions to explore to what extent object-specific actions related to object word meanings. The variables in model candidates and priors were the same as in Analysis 1. For Analysis 3, we focused on developmental changes in the numbers of common nouns and verbs to explore the relationships between the semantic change of object words and vocabulary size. The number of words was used as the dependent variable and age in months was used as the independent variable. To detect the change point where vocabulary size increased drastically, we regarded age as a categorical variable. The model was fitted using a Poisson distribution with overdispersion and a log link function. We looked for differences in vocabulary size between adjacent ages. The correlations between task performance in each condition and the number of words were further explored while moving through the target two-month age range.

For Bayesian statistical modeling, the posterior median (MED) and 95% Bayesian credible intervals (CIs) of parameter estimates or expected values were used to interpret the results. We adopted a model comparison approach for Analyses 1 and 2 using the widely applicable information criterion (WAIC; Watanabe, 2010a, 2010b) to determine the best model among candidate model sets. Using posterior distributions, the WAIC was calculated from the difference between the mean log-likelihood as a prediction ability of the model and the sum variance in log-likelihood as a penalty. A smaller WAIC value indicates a better model. The models were fitted using RStan 2.19.3 (Stan Development Team, 2019). We set four chains and iterations of 20,000 with burn-in samples of 2,000. The convergence of parameter estimates was confirmed by checking if the Rhat values were below 1.1 (Gelman et al., 2013).

Predictions. For Analysis 1, we expected to replicate the results obtained in Chapter 2 for both looking and pointing measures. If object word meanings were undifferentiated global event categories comprising both objects and actions, the task performance in the match condition would be above chance level, whereas that in the mismatch condition would remain at chance level. If such word meanings were specific object categories independent of actions, the task performance in both the match and the mismatch conditions would exceed chance level. We predicted that younger participants would be in the former case, and older participants would be in the latter case.

For Analysis 2, at least for older participants, we expected that the task performance in both the different-object and the absent-object conditions would remain at chance level because object words for them semantically differentiated and became independent of actions. For younger participants, if initial meanings of object words were the undifferentiated fusion of both objects and actions (i.e., the holistic-fusion-based explanation), the same predictions would apply for older participants, because toddlers could not detect object word referents solely by object-specific actions. In contrast, if a relatively great weight was given to object-specific actions in object word comprehension (i.e., the action-weight-based explanation), the task performance, at least in the different-object condition, would be above chance level. Moreover, if actions dominated the initial meanings of object words and objects played only subsidiary roles, the task performance in both the different-object and the absent-object conditions would exceed chance level.

For Analysis 3, we predicted that the number of common nouns increased across the whole range of target ages, whereas the number of verbs started to increase after the period in which the task performance in the mismatch condition exceeded chance level (i.e., the object word meanings differentiated into specific object categories). Since we posited that toddlers whose object word meanings were sufficiently confined to object categories experienced the necessity of constructing new label-meaning connections to express actions specifically, task performance, especially in the mismatch condition among the four conditions, was expected to positively relate to the number of verbs.

4.2.2 Results

Equivalent interpretability of preferential looking and pointing measures

We first confirmed the practical validity of using the majority-vote-based index calculated from preferential looking, which is different from the original study using only pointing responses (see Chapter 2). For toddlers who showed clear pointing responses in at least one trial for each condition, the proportion of correct responses calculated from looking and pointing measures was positively correlated regardless of conditions using Spearman’s rank correlation coefficient (the lowest was $\rho = 0.49$, $p = 0.0003$, n = 51 in the different-object condition; the highest was $\rho = 0.72$, $p < 0.0001$, n = 51 in the mismatch condition). The correlation became even stronger when limited to toddlers who showed pointing responses in all four trials in each condition (ρ ranging from 0.69 to 0.83; see **Appendix D** for more details).

Analysis 1: Are the meanings of object words undifferentiated event categories or specific object categories?

The best model for both the preferential looking and the pointing models included the fixed effects of age and condition, but not their interaction (see **Appendix E**, **Table E.1** for model comparison). The developmental trend shown in **Figure 4.4-a** demonstrated that the proportion of correct responses increased with age in months, and it was lower in the mismatch condition than in the match condition (see **Table 4.1** for parameter estimates). For the preferential looking model, the proportion of correct responses was above chance level in the match condition even for 18 months (MED = 0.687, CI [0.593, 0.769]). In the

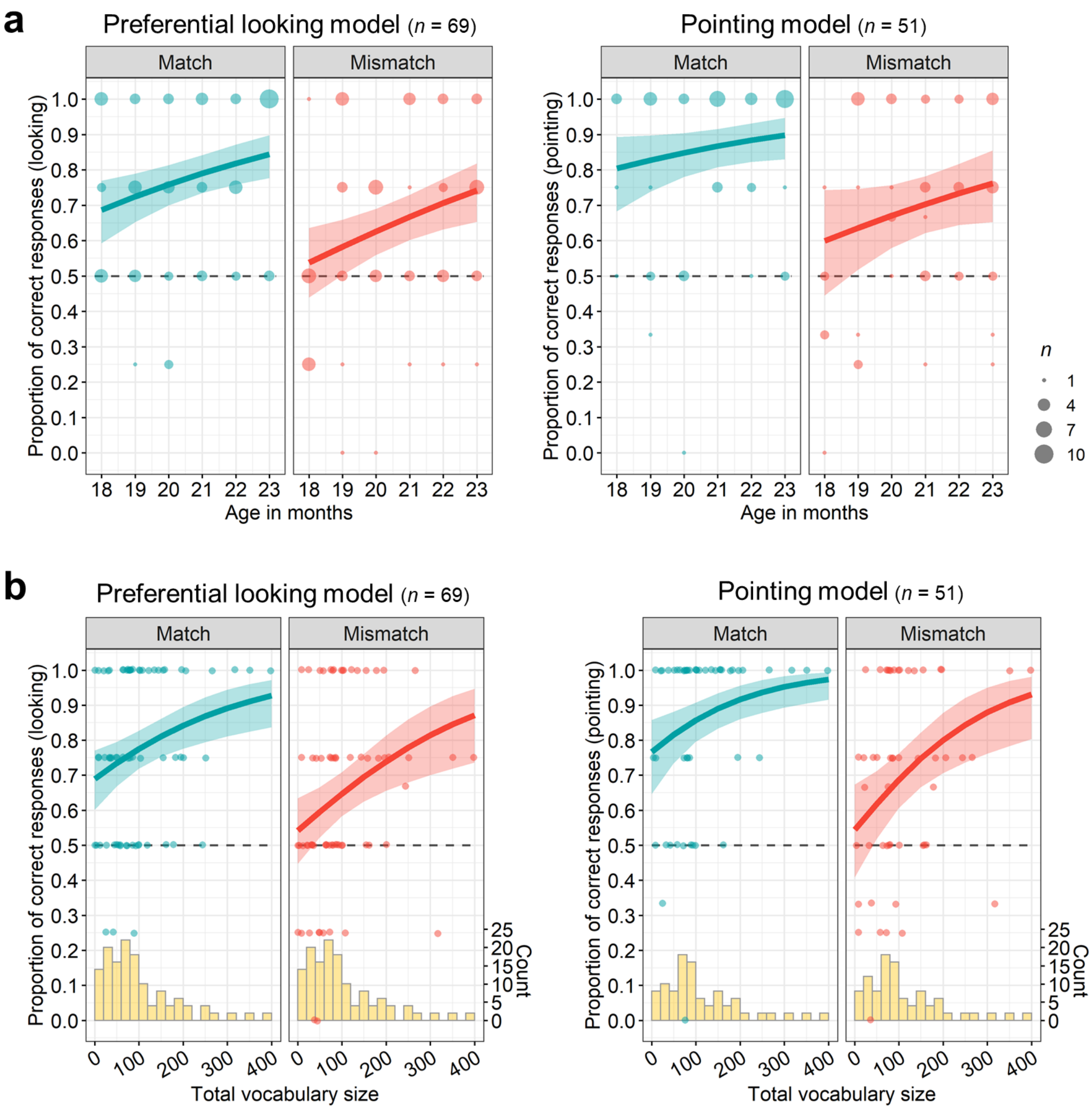

Figure 4.4 Developmental changes in the proportion of correct responses between the match and mismatch conditions in Study 1. The proportion of correct responses was calculated from the majority-vote-based measurement using preferential looking (left panels) and the number of responses pointing to the correct stimulus (right panels). Thick lines and shaded areas represent the posterior median and 95% Bayesian credible intervals of expected values estimated from the selected best models, respectively. The dashed horizontal line indicates the chance level. (a) Age in months is used as a developmental index. The size of data points indicates the number of participants located at the same coordinates. (b) Total vocabulary size is used as a developmental index. Dots represent observed data. The histogram shows the distribution of the number of participants regarding total vocabulary size.

Table 4.1 Posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates, and widely applicable information criterion (WAIC) for the best models with the match and the mismatch conditions in Study 1.

Developmental index	Parameter	Posterior median [95% CI]		WAIC (SE)
age (months)	***The preferential looking model (n = 69)***			
	age	0.181	[0.059, 0.308]	**369.94** (13.03)
	condition (mismatch)	-0.634	[-1.016, -0.258]	
	intercept	-2.472	[-5.049, 0.022]	
	σ	0.402	[0.055, 0.753]	
	The pointing model (n = 51)			
	age	0.152	[-0.030, 0.343]	**229.50** (12.76)
	condition (mismatch)	-1.014	[-1.564, -0.490]	
	intercept	-1.316	[-5.245, 2.522]	
	σ	0.568	[0.087, 1.041]	
vocabulary size	***The preferential looking model (n = 69)***			
	vocabulary size	0.004	[0.002, 0.007]	**369.32** (13.41)
	condition (mismatch)	-0.632	[-1.018, -0.244]	
	intercept	0.797	[0.410, 1.211]	
	σ	0.384	[0.063, 0.734]	
	The pointing model (n = 51)			
	vocabulary size	0.006	[0.002, 0.010]	**224.11** (13.28)
	condition (mismatch)	-1.009	[-1.559, -0.471]	
	intercept	1.188	[0.606, 1.795]	
	σ	0.411	[0.056, 0.898]	

Note. The parameter σ represents the standard deviation of the random intercept.

mismatch condition, it exceeded chance level for 19 months (MED = 0.583, CI [0.503, 0.659]) and older toddlers, whereas it remained near chance level for 18 months (MED = 0.538, CI [0.439, 0.636]). The pointing model showed a similar pattern as the proportion of correct responses was above chance level in the match condition for 18 months (MED = 0.805, CI [0.683, 0.893]) as well as in the mismatch condition for 19 months (MED = 0.635, CI [0.518, 0.746]) and older ages, but its interval did not exceed chance level in the mismatch condition for 18 months (MED = 0.599, CI [0.446, 0.743]). Roughly, the proportion of correct responses tended to be slightly higher in the pointing model than in the preferential looking model.

Similar results were obtained when using total vocabulary size as a developmental index (**Figure 4.4-b** and **Table 4.1**; see also **Appendix E**, **Table E.2** for model comparison). The vocabulary size when the interval of proportion of correct responses in the mismatch condition exceeded chance level was approximately 40 and 50 words in the preferential looking and the pointing models, respectively. This vocabulary size corresponded to approximately 19–20 months for boys and 18–19 months for girls based on the 50th percentile estimate from the J-MCDI (Ogura & Watamaki, 2004).

Analysis 2: Can toddlers understand object word referents based solely on object-specific actions?

The null model (i.e., including intercept only) was selected as the best model for both the preferential looking and the pointing models (see **Appendix E**, **Table E.3** for model comparison). The proportion of correct responses remained near chance level in the former model (MED = 0.495, CI [0.451, 0.538]), whereas it was slightly above chance level in the latter model (MED = 0.563, CI [0.504, 0.621]). These results are shown in **Figure 4.5-a**, and the parameter estimates are shown in **Table 4.2**. Besides, model candidates that included total vocabulary size as a developmental index were not selected based on the comparison of WAIC values (see **Figure 4.5-b** and **Appendix E**, **Table E.4**).

Analysis 3: Developmental changes in vocabulary size

Expressive vocabulary size ranged between 1 and 398 (Median = 79, *M* = 100, *SD* = 84.06). Common nouns ranged from 0 to 181 (Median = 23, *M* = 39.14, *SD* = 43.12), and verbs ranged from 0 to 59 (Median = 2, *M* = 6.42, *SD* = 11.29). The number of common nouns gradually increased during the target age range from 18 months (MED = 5.64, CI [2.49, 12.28]) to 23 months (MED = 48.29, CI [23.77, 97.40]), as shown in **Figure 4.6**. The CIs of the difference in common nouns between all adjacent ages straddled zero, indicating that there was no steep developmental change in our sample (**Table 4.3**). However, the number of verbs abruptly increased from 20 months (MED = 0.52, CI [0.14, 1.66]) to 21 months (MED = 4.62, CI [1.58, 12.77]). The CIs of the difference in verbs between adjacent ages exceeded zero for only 20–21 months (MED = 2.20, CI [0.61, 3.84]).

The results of the correlation analysis between task performance and the number of words are shown in **Figure 4.7** (see also **Appendix F** for details). Roughly speaking, the task performance in the mismatch condition showed relatively stable high values of Spearman's rank correlation for both common nouns (the lowest was $\rho = 0.18$, $n = 25$ for 18–19 months using the preferential-looking-based index; the highest was $\rho = 0.62$, $n = 16$ for 19–20 months using the pointing-based index) and verbs (the lowest was $\rho = -0.068$, $n = 25$ for 18–19 months using the preferential-looking-based index; the highest

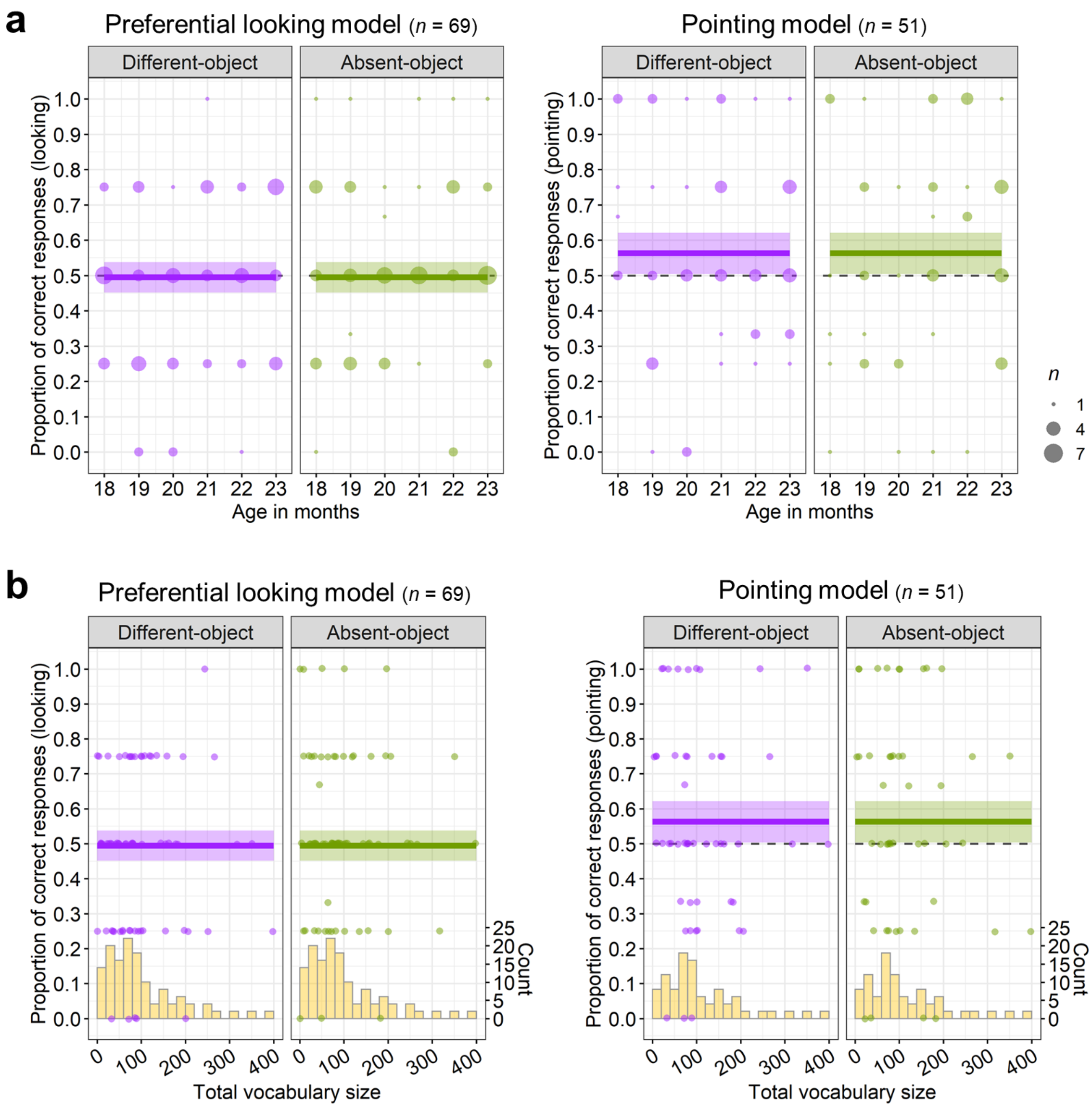

Figure 4.5 Developmental changes in the proportion of correct responses between the different-object and absent-object conditions in Study 1. (a) Age in months and (b) total vocabulary size are used as developmental indices. The legends are the same as in Figure 4.4.

was $\rho = 0.57$, $n = 17$ for 21–22 months using the pointing-based index), compared with other conditions. Conversely, the different-object and absent-object conditions showed relatively negative or near-zero values of correlations, although they depended on age range and on the task performance measures.

Table 4.2 Posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates, and widely applicable information criterion (WAIC) for the best models with the different-object and the absent-object conditions in Study 1.

Parameter	Posterior median [95% CI]		WAIC (SE)
The preferential looking model (n = 69)			
intercept	-0.022	[-0.195, 0.152]	**376.18** (11.75)
σ	0.141	[0.019, 0.412]	
The pointing model (n = 51)			
intercept	0.252	[0.016, 0.495]	**258.12** (9.30)
σ	0.281	[0.044, 0.663]	

Note. The parameter σ represents the standard deviation of the random intercept. There were no best models among model candidates that included age in months or vocabulary size.

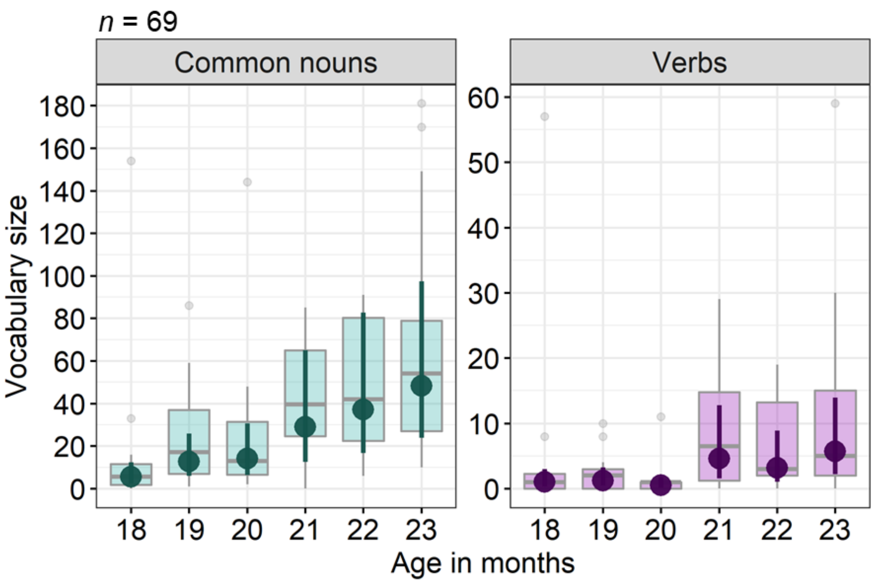

Figure 4.6 Developmental changes in vocabulary size of common nouns and verbs in Study 1. Box plots are observed data. Large dots and thick lines represent the posterior median and 95% Bayesian credible intervals of expected values, respectively.

4.2.3 Discussion

The results from Analysis 1 indicated that initial meanings of object words were undifferentiated event categories, whereas they subsequently differentiated into specific object categories independent of object-specific actions. Roughly speaking, we succeeded

Table 4.3 Posterior median and 95% Bayesian credible intervals (CIs) of expected values of differences in vocabulary size between adjacent ages.

Adjacent ages	Posterior median [95% CI]			
	Common nouns		Verbs	
19 minus 18 months	0.81	[-0.27, 1.90]	0.15	[-1.30, 1.65]
20 minus 19 months	0.10	[-0.95, 1.17]	-0.87	[-2.47, 0.68]
21 minus 20 months	0.72	[-0.41, 1.86]	**2.20**	**[0.61, 3.84]**
22 minus 21 months	0.25	[-0.89, 1.43]	0.38	[-1.90, 1.10]
23 minus 22 months	0.26	[-0.81, 1.33]	0.59	[-0.81, 2.01]

Note. The CI that exceeded zero is shown in bold.

in replicating the preliminary findings in Chapter 2. The results from Analysis 2 suggested that toddlers over the entire age range in this study could not judge object word referents based solely on object-specific actions. Taken together, the results in the mismatch condition would be observed because younger toddlers failed to properly recognize object word referents when an object and an object-specific action were not matched. Finally, Analysis 3 revealed that, as expected, the vocabulary size for verbs increased sharply a few months after object word meanings became free from object-specific actions. Exploratory correlation analysis also provided supporting evidence on the positive relationship between task performance in the mismatch condition and the number of verbs, especially in older participants. However, due to cross-sectional data, these results provided only indirect evidence of the relationship between semantic differentiation and vocabulary growth. We directly investigated whether individual differences in comprehension of object words predicted later vocabulary increase using longitudinal data in Study 2.

4.3 Study 2: A longitudinal study

4.3.1 Method

Participants

Of the younger participants in Study 1, 16 toddlers who were able to participate in data collection approximately two months later were included in the final analysis (9 boys and 7 girls; M_{age} = 18.6 months, SD = 0.5, 18–19 months range for the first time; M_{age} = 21.1 months, SD = 0.7, 20–22 months range for the second time). The mean interval between the first and the second time was 69.6 days (SD = 2.7; 65–77 days range). The parents again

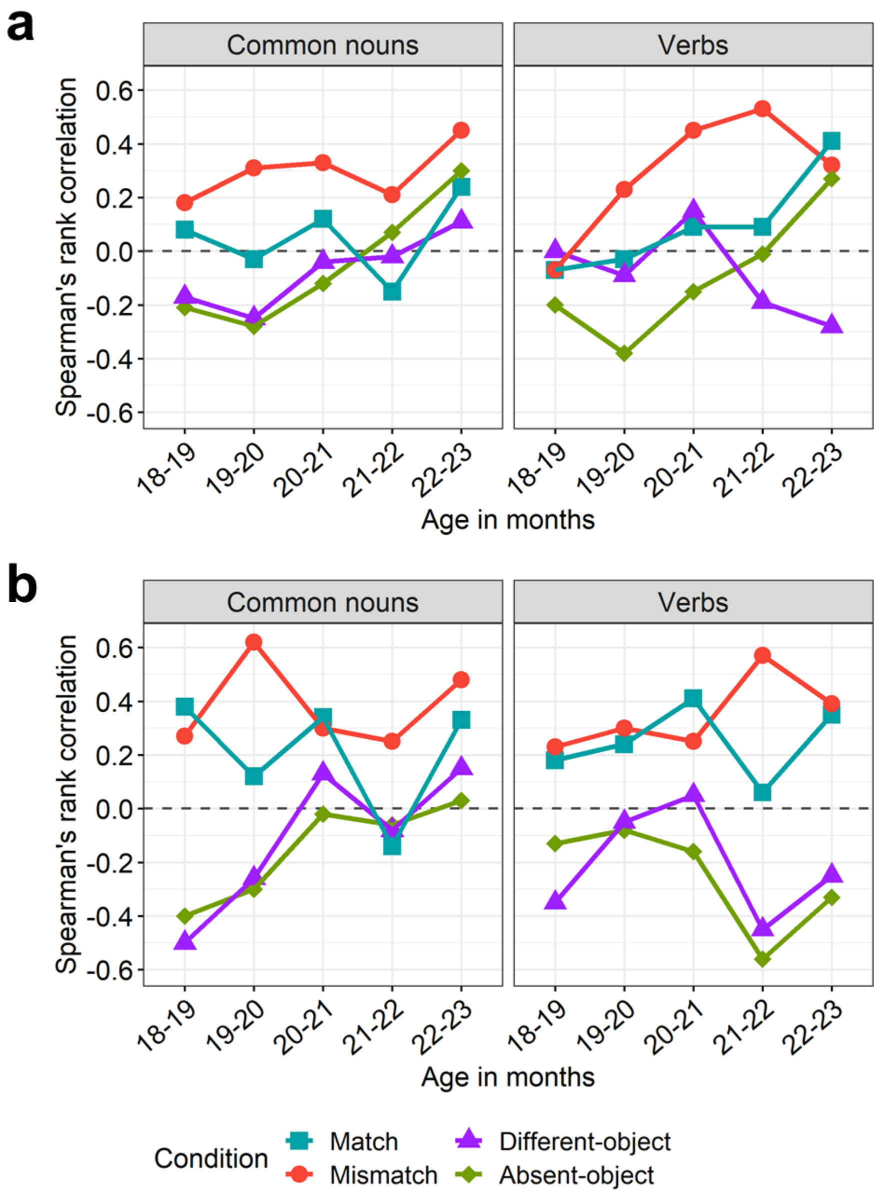

Figure 4.7 The correlation between task performance in each condition and the number of words in Study 1. The proportion of correct responses was calculated from (a) the majority-vote-based measurement using preferential looking and (b) the number of pointing responses. Correlation coefficients were calculated while delimiting the target age range using a two-month window.

completed the J-MCDI.

Stimuli and procedure

The stimulus materials, apparatus, and procedure were the same as in Study 1.

Data analyses

Indices and coding. The data were pre-processed, as in Study 1. We adopted the majority-vote-based method to calculate correct responses from preferential looking as the measure reflecting task performance to avoid increasing missing data (four among sixteen participants did not show clear pointing responses at all the first time). Treating this measure as the representative variable was considered reasonable since we confirmed in Study 1 that this measure could be interpreted as equivalent to pointing to the correct stimulus (see also Chapter 3).

To estimate vocabulary growth, the difference scores in the numbers of common nouns or verbs in the J-MCDI (the second time vocabulary minus the first time vocabulary) were calculated as in a previous similar study (Slone, Smith, & Yu, 2019).

Data rejection. No participants nor trials were excluded.

Bayesian statistical modeling. Three analyses were conducted using Bayesian hierarchical generalized linear models. The numbers of chains, iterations, and burn-in samples were identical to those in Study 1. Analyses 1 and 2 were the same as in Study 1, except for developmental indices and random intercepts. We used the time (first or second) as the developmental index. To account for within-participants effects for both the times and the conditions, we set random intercepts. The same weakly informative priors used in Study 1 were adopted for all random effects.

For Analysis 3, we focused on the effect of individual differences in toddlers’ understanding of object words on subsequent vocabulary growth. Thus, vocabulary growth for common nouns or verbs was used as the dependent variable, and task performance for each condition for the first time was used as the fixed effect (i.e., the predictor). We included participants’ age in months for the first time as control variables. We converted 18 and 19 months to 0 and 1, respectively, to match the scales with other fixed effects. The model was fitted using a Poisson distribution with overdispersion and a log link function. Weakly informative Student-*t* priors ($\upsilon = 4,\ \mu = 0,\ \sigma = 3.5$) were used for the fixed and random effects to stabilize the parameter estimates.

Predictions. We predicted that similar results to those of Study 1 would be obtained for Analyses 1 and 2 from the longitudinal data. For Analysis 3, we posited that the task performance in the mismatch condition, which reflected the degree of differentiation of object word meanings from global to specific object categories, would have a positive effect on verb growth, whereas task performance in the match condition would not.

4.3.2 Results

Analysis 1: Semantic differentiation from global event to specific object categories

The best model included the fixed effects of time and condition, but not their interaction, in parallel with Study 1 (see **Appendix E**, **Table E.5** for model comparison). The proportion of correct responses increased with age in months, as shown in **Figure 4.8-a**, and it was lower in the mismatch condition than in the match condition, although the credible interval of the parameter estimate straddled zero (**Table 4.4**). The proportion of correct responses was above chance level in the match condition for both the first time (MED = 0.700, CI [0.538, 0.833]) and the second time (MED = 0.835, CI [0.710, 0.922]). In contrast, this proportion in the mismatch condition exceeded chance level only for the second time (MED = 0.788, CI [0.647, 0.893]), but not for the first time (MED = 0.632, CI [0.463, 0.781]).

Analysis 2: The effect of object-specific actions on object word comprehension

As in Study 1, the null model was selected as the best model (see **Appendix E**, **Table E.5** for model comparison). The proportion of correct responses remained near chance level (MED = 0.476, CI [0.402, 0.549]), as shown in **Figure 4.8-b** (see also **Table 4.4**).

Analysis 3: Vocabulary growth prediction

Vocabulary growth of common nouns and verbs ranged from 0 to 59 (Median = 22.50, M = 28.44, SD = 22.89) and from 0 to 22 (Median = 1.50, M = 5.38, SD = 6.66), respectively. The best prediction models for both common noun and verb vocabulary growth included the fixed effects of the first-time task performance for all four conditions (see **Appendix E**, **Table E.6** for model comparison). For common nouns, the parameter estimates of all conditions straddled zero (**Figure 4.9** and **Table 4.5**). Relatively, there was a positive relationship between vocabulary growth and the proportion of correct responses in the mismatch (MED = 1.007, CI [-2.339, 4.430]), whereas a negative relationship was observed in the different-object condition (MED = -1.076, CI [-5.020, 2.833]). The effect size of the task performance in the match condition was the smallest (see the standardized parameter estimates in **Table 4.5**). For verbs, the proportion of correct responses in the mismatch condition was positively related to later vocabulary growth (MED = 4.695, CI [0.293, 10.126]; see **Figures 4.9** and **4.10** and **Table 4.5**). Other parameter estimates straddled zero, although there was a relatively negative relationship between vocabulary growth and task performance in the different-object (MED = -1.414, CI [-5.834, 2.997]) and absent-object conditions (MED = -1.577, CI [-5.495, 1.805]).

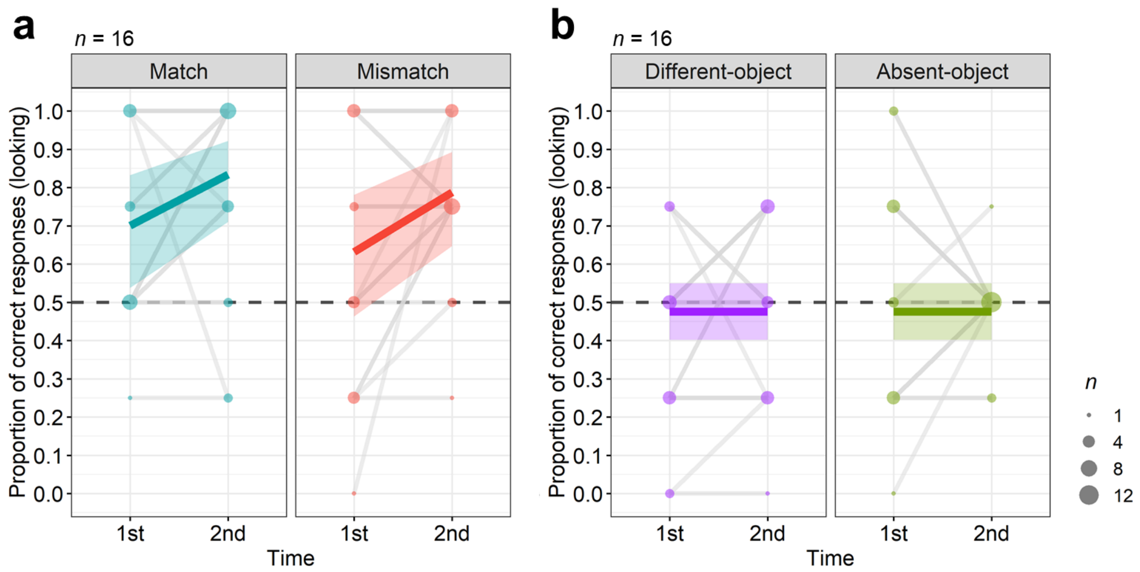

Figure 4.8 Developmental changes in the proportion of correct responses in Study 2. The dots connected by gray lines indicate the same individual at different times. Other legends are the same as in Figure 4.4-a. Comparison (a) between the match and the mismatch conditions and (b) between the different-object and the absent-object conditions.

Table 4.4 Posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates, and widely applicable information criterion (WAIC) for the best models in Analyses 1 and 2 in Study 2.

Analysis	Parameter	Posterior median [95% CI]	WAIC (SE)
1	***The match and the mismatch conditions***		
	time	0.775 [0.090, 1.508]	**172.43** (10.20)
	condition (mismatch)	-0.309 [-1.110, 0.474]	
	intercept	0.075 [-1.122, 1.276]	
	σ_1 (participant)	0.405 [0.028, 1.090]	
	σ_2 (participant and time)	0.396 [0.032, 1.023]	
	σ_3 (participant and condition)	0.629 [0.090, 1.280]	
2	***The different-object and the absent-object conditions***		
	intercept	-0.098 [-0.397, 0.198]	**173.39** (7.10)
	σ_1 (participant)	0.268 [0.027, 0.676]	

Note. Parameters σ_1, σ_2, and σ_3 represent the standard deviation of the random intercepts.

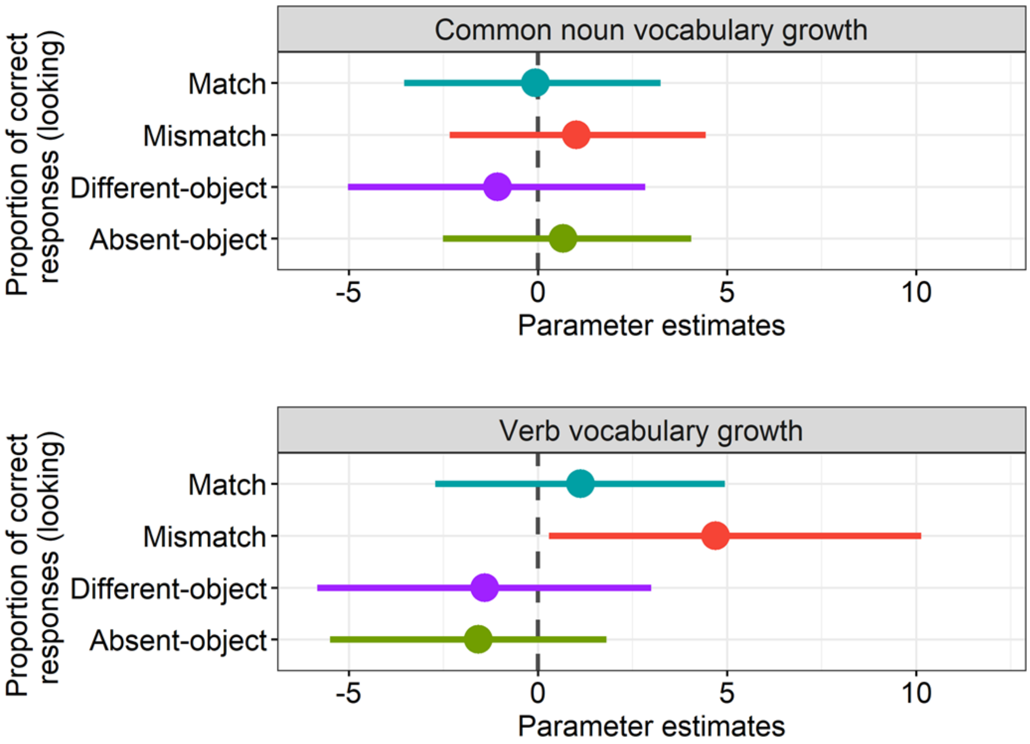

Figure 4.9 Posterior median and 95% Bayesian credible intervals of parameter estimates for the best models in Analysis 3 in Study 2. Dots and lines represent the posterior median and 95% Bayesian credible intervals of parameter estimates, respectively.

4.3.3 Discussion

The results of developmental change in the task performance obtained in Study 1 were roughly replicated in Study 2 using longitudinal data. The higher task performance in Study 2 in the first-time mismatch than in Study 1 and the credible interval for the parameter estimates between the match and the mismatch conditions straddling zero might be because the sample in Study 2 was a mix of 18- and 19-month-olds. This age period was just before and after the semantic differentiation of object words was actually observed in cross-sectional data. The results from Analysis 3 revealed that individual differences in semantic differentiation of object words predicted subsequent vocabulary growth of verbs in particular. This finding provided direct evidence supporting the hypothesis that the differentiation of object word meanings encouraged toddlers to develop new labels depicting specific actions.

Table 4.5 Posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates and widely applicable information criterion (WAIC) for the best models in Analysis 3 in Study 2.

Parameters	Posterior median [95% CI]		Standardized posterior median [95% CI]		WAIC (SE)
Common noun growth prediction					
match	-0.067	[-3.541, 3.235]	-0.076	[-1.469, 1.202]	**94.65**
mismatch	1.007	[-2.339, 4.430]	0.438	[-1.109, 2.054]	(8.51)
different-object	-1.076	[-5.020, 2.833]	-0.415	[-1.872, 0.970]	
absent-object	0.665	[-2.514, 4.048]	0.284	[-1.067, 1.874]	
age	1.901	[-0.504, 4.705]	0.984	[-0.620, 2.951]	
intercept	0.733	[-3.101, 4.376]	2.306	[0.772, 3.363]	
σ	1.829	[1.092, 3.297]	2.085	[1.183, 4.068]	
Verb growth prediction					
match	1.121	[-2.718, 4.939]	0.659	[-0.885, 2.660]	**55.78**
mismatch	4.695	[0.293, 10.126]	2.705	[0.513, 5.915]	(10.21)
different-object	-1.414	[-5.834, 2.997]	-0.558	[-2.356, 1.378]	
absent-object	-1.577	[-5.495, 1.805]	-0.478	[-2.280, 1.178]	
age	-0.298	[-3.686, 3.126]	-0.526	[-3.298, 2.280]	
intercept	-1.945	[-6.875, 2.120]	-0.597	[-3.161, 0.876]	
σ	1.898	[1.016, 3.932]	2.393	[1.222, 5.166]	

Note. Age in months for the first time was included in model candidates as the control variable. Parameter σ represents the standard deviation of the random intercepts.

4.4 General discussion

In addition to examining the robustness of the findings that early noun meanings developed from undifferentiated event categories to specific object categories (Chapter 2), we investigated whether toddlers could properly select object word referents solely by object-specific actions and how vocabulary grew along with the semantic change of object words in this Chapter.

Results from both cross-sectional and longitudinal data demonstrated that even younger toddlers were able to choose the correct stimuli in the match condition, and the accuracy became more precise with age. On the contrary, the task performance in the mismatch condition was near chance level for the younger period, and it exceeded chance level only later. These results show the robustness of the findings obtained in Chapter 2 and corroborate the semantic pluripotency hypothesis. The near-chance level of the task performance observed in the mismatch condition would be well interpreted by combining

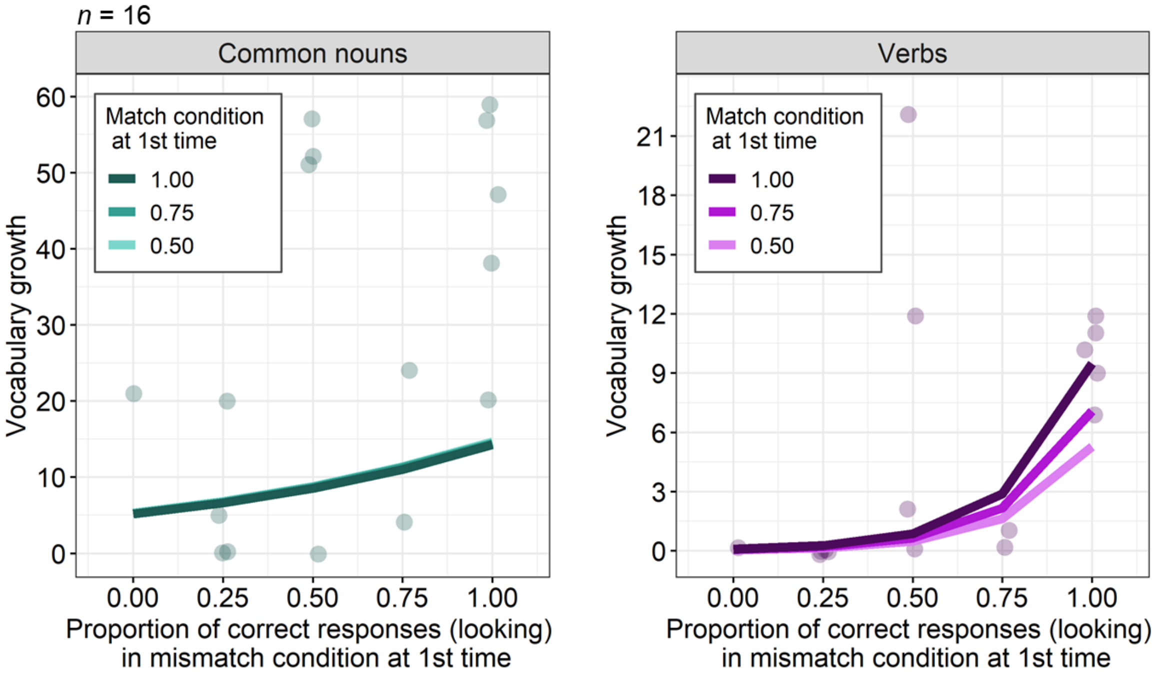

Figure 4.10 Individual differences in vocabulary growth of common nouns and verbs in Study 2. Dots represent observed data. Lines indicate the posterior median of expected values, color-coded by different task performance in the match condition for the first time. Task performance in the different-object and the absent-object conditions, which was also included in the best regression model, was fixed at 0.5. Age as the control variable was also fixed at the mean (= 18.6).

the results in the different-object and absent-object conditions. Considering the fact that toddlers were unable to detect object word referents based solely on object-specific actions even when two options did not compete with each other, it can be said that the failure of younger toddlers to select the correct stimulus in the mismatch condition occurred not because they understood object words with an emphasis on object-specific actions (according to the action-weight-based explanation), but because the initial meanings of object words were the undifferentiated fusion of both objects and actions (according to the holistic-fusion-based explanation). Thus, while it is true that object-specific actions play a crucial role even in early object category or object word learning (Booth, 2006; Booth, et al., 2010; Booth & Waxman, 2002; Hernik & Csibra, 2009; Horst et al., 2005; Kemler Nelson, 1995; Kemler Nelson, Frankenfield et al., 2000; Kemler Nelson, Russell et al., 2000; Kobayashi, 1992, 1997, 1998; Madole et al., 1993; McCune, 2008; Nelson, 1973a, 1974; Perone & Oakes, 2006; Ross et al., 1986; Träuble & Pauen, 2007, 2011; Ware & Booth, 2010; Werner and Kaplan, 1963), it does not mean that toddlers associate object-word-like labels relative to actions than objects, but rather associate such labels with holistic undifferentiated event categories in which the exact objects are embedded. Some might think that the task performance dropped in the mismatch condition because younger toddlers regarded filler objects (baskets and a toy shovel) as the target objects; however, this possibility would be ruled out because, if this had been the case, they would have been able to select the correct stimulus in the different-object condition.

Based on the semantic pluripotency hypothesis, we posited that the semantic differentiation of object words from global event categories to specific object categories would encourage children to form new label-meaning connections to be used for specific actions, thus resulting in the vocabulary growth of action words. In Study 1, we found that a steep increase in verbs occurred a few months after the task performance in the mismatch condition exceeded chance level, and this task performance was positively correlated with the number of verbs compared to the other conditions. In addition to this indirect evidence, Study 2 showed that the more object word meanings differentiated into specific object categories, the more the verb vocabulary size grew subsequently. Both semantic differentiation of object words and the invention of action words would allow children to combine object and action words to depict events more in detail. This study provided the first experimental evidence that developmental change in object word meanings predicted later vocabulary growth, which reinforced the previous observational findings that explored the cognitive development underlying the change in language use from one-word to two-word utterances. For instance, young children use gestures (Özçalışkan & Goldin-Meadow, 2005), intonation (Werner & Kaplan, 1963), or successive single-word utterances (Herr-Israel & McCune, 2011) to distinguish specific aspects of holistic events (objects or actions) before they can produce multi-word utterances.

Task performance in the mismatch condition was also positively related to the vocabulary size of common nouns in both Study 1 and Study 2, although careful interpretation is needed because of weak evidence, as the credible interval of parameter estimates straddled zero. Regarding this positive relationship as plausible would imply that semantic differentiation of object words might be related to the acquisition of metalinguistic knowledge that “objects have their own names” (i.e., “naming insight”; Dore, 1978; Kamhi, 1986; McShane, 1979; Nelson, 1983a) and the efficient use of a naming bias such as shape bias (Gershkoff-Stowe & Smith, 2004; Graham & Poulin-Dubois, 1999; Imai et al., 1994; Kucker et al., 2019; Landau et al., 1988, 1998; Perry & Samuelson, 2011; Samuelson, 2002; Samuelson & Smith, 1999; Smith et al., 2002; Yee et al., 2012).

In addition, task performance in the different-object and absent-object conditions had a negative relation to verb vocabulary size in particular. This might have occurred because individuals, who placed relatively great weight on actions in object word meanings, used object words to express actions, which resulted in the unnecessity of developing action words. In any case, further investigation is required to examine these possibilities.

The critical developmental period of semantic differentiation differed from the original preliminary study in Chapter 2. In this chapter, we detected that the period was around when toddlers reached 19 months or expressive vocabulary size of 40 or 50 words, although we initially reported that it occurred later, at around 21 months or 140 words in Chapter 2. This difference may be attributed to the small sample size and widely distributed age range in the previous study; however, it would still be true to say that cognitive abilities regarding word learning drastically develop during the latter half of the second year. Interestingly, the vocabulary size detected in this study was consistent with the rule of thumb of the critical period when vocabulary spurt occurred (Mervis & Bertrand, 1995; Nazzi & Bertoncini, 2003), suggesting that semantic differentiation of object words from global event categories to specific object categories might contribute, even partially, to the acceleration of word learning. Although some researchers assumed that there was no qualitative cognitive change before and after vocabulary spurt (e.g., McMurray, 2007), it is worth mentioning that the present study provided findings of qualitative change in early word meanings.

In this chapter, we utilized majority-vote-based measurements calculated from preferential looking based on the Simple Majority Vote model developed in Chapter 3 in order to avoid increasing missing data as well as to compare the results from looking and pointing responses directly. The results from Study 1 showed that the proportion of correct responses was slightly higher in the pointing model than in the preferential looking model both in the comparison between the match and mismatch conditions and between the different-object and the absent-object conditions. This tendency suggests that toddlers produced pointing responses only when they were confident or when object word meanings were comparatively robust (Hendrickson & Friend, 2013). Thus, it would be meaningful for more detailed interpretations and prevention of biased results if studies using pointing responses could also obtain concurrent gaze or head-turn responses. Particularly, we strongly recommend that studies focusing on the second half of the second year of life should collect and analyze both looking and pointing responses. In our data, the mean proportion of absent pointing responses was around 50% at about 18 months and it varied across individuals; however, this proportion steeply decreased during 18–20 months (see **Appendix G**). It would also be beneficial to pay attention to trials that lack pointing responses themselves (Hendrickson et al., 2015, 2017), or children's responses reflecting uncertainty, such as social reference to adults (Leckey et al., 2020).

Overall, the findings obtained in this chapter corroborate the semantic pluripotency hypothesis. Although we will discuss the limitation of this research and future directions comprehensively in Chapter 5, we briefly raise five points here. First, although we succeeded in identifying the interpretation of the results in the mismatch condition, an additional condition would be necessary to exhaustively examine the “core meanings” of object words. The absent-action condition, in which toddlers have to judge object word referents solely by objects, would be needed. Second, whether the semantic differentiation is a language-general phenomenon should be investigated. Common nouns account for the largest proportion of early vocabulary in Japanese (Ogura et al., 2016); however, Japanese is generally classified as a verb-friendly language because, for instance, it allows dropping the subject and the object of a sentence (Imai et al., 2005). A similar experiment to this study should be conducted for noun-friendly languages, such as English or French. Third, it is worthwhile to examine whether there is a semantically undifferentiated phase in verbs. Early words might initially encompass undifferentiated event categories, regardless of nouns or verbs. If so, the fact that in some languages, children learn verbs earlier than nouns (Frank et al., in press) could be explained as they do not learn verbs as words corresponding to specific actions but as words corresponding to global events. Fourth, what drives the semantic differentiation should be explored. Decontextualization (Lucariello, Kyratzis, & Engel, 1986; Sandhofer & Schonberg, 2020) would play a central role. Seeing shoes in various contexts (e.g., doorway or shop), or getting to know shoes used in unconventional ways (e.g., shoes as vases) might facilitate semantic differentiation. Fifth, it would be interesting to examine whether there is any adaptive significance of undifferentiated characteristics of the initial word meanings. Consideration of the semantic pluripotency hypothesis in the context of syntax development might be insightful and could contribute to answering it.

In conclusion, although early words often have been regarded as equivalent to those of adults (e.g., Huttenlocher & Smiley, 1987), the experimental findings obtained in this chapter shed light on the distinctive meanings of initial object words for younger children, which differed from those for older children and adults. Thus, we should be cautious about inferring that the meanings of early nouns are exactly the same as “ordinary object words” based merely on their forms.

4.5 Summary of this chapter

In many languages, nouns that semantically correspond to objects dominate children’s early vocabulary over verbs, which correspond to actions. However, some researchers argue that the meanings of early words cannot be easily classified into specific object or action categories. Based on the semantic pluripotency hypothesis developed to examine this possibility experimentally, we investigated whether noun meanings are inherently the exact object categories or are at first global event categories and subsequently differentiate into specific ones using a two-alternative forced-choice video task for 18–23-month-old toddlers (n = 69). Both the cross-sectional and longitudinal results suggested that unlike older toddlers, younger ones could not select the correct referents when objects and object-specific actions were presented separately (e.g., "doing a filler action with shoes" vs. "putting on filler objects"), despite success when objects and such actions were matched (e.g., "putting shoes on" vs. "doing a filler action with filler objects"). The results from additional conditions suggested that toddlers failed to judge object word referents solely by object-specific actions. These results indicated that the initial meanings of object words were the undifferentiated fusion of both objects and actions (e.g., "putting shoes on"), while they subsequently differentiated into specific object categories independent of actions. Furthermore, the degree of this semantic differentiation was related to the vocabulary growth of action words, suggesting that the differentiation of object word meanings encouraged toddlers to develop new labels that could depict specific actions apart from objects. These experimental findings shed light on the distinctiveness of initial object word meanings, which differ from adults' "ordinary object words."

Chapter 5

General Discussion

> There is probably no language that does not, in some way, categorize experience through formal means. In an Indo-European language, such as English, the general difference between a noun as a "thing word," a verb as an "action word," etc., is clearly expressed grammatically and syntactically. Long before the child masters the linguistic distinctions of such categories in his expressions of complex reference, he advances through a series of steps that only gradually culminate in the formation of nouns, verbs, etc. (Werner & Kaplan, 1963, p. 143).

5.1 Summary of key findings

The central purpose of this thesis was to experimentally investigate the semantic pluripotency hypothesis, which is a revision of the theoretical hypothesis proposed by Werner and Kaplan (1963), with a primary focus on object words. Throughout a series of chapters, we set out the findings we have obtained that support this hypothesis and provide directions for future research on the formation and differentiation of early word meanings.

The first and most important finding is that the initial meanings of object words are not already specific object categories from when they are first learned. Instead, they are global and holistic event categories in which both objects and actions are fused and subsequently differentiated into specific and discrete object categories independent of object-specific actions. In the two-alternative forced-choice task, even younger toddlers aged around 18 months were able to choose the appropriate referents of familiar object words, such as "shoe," when objects and object-specific actions were compatible on one side (e.g., "putting shoes on" vs. "rubbing baskets" in the match condition). On the contrary, unlike older toddlers aged around 23 months, the probability of choosing the appropriate referents remained at near chance level for younger toddlers when objects and object-specific actions were presented separately (e.g., "putting baskets on" vs. "rubbing shoes" in the mismatch condition). These results were robust throughout this thesis for the cross-sectional preliminary study (Chapter 2), and the cross-sectional and longitudinal confirmatory study (Chapter 4), regardless of developmental indices (age in months or total vocabulary size) and task performance indices (looking or pointing responses). Furthermore, given that toddlers failed to select appropriate object word referents solely by object-specific actions (i.e., the different-object and the absent-object conditions in Chapter 4), it would be indicated that the near-chance level of task performance in the mismatch condition for younger toddlers was not attributed to children's placement of relatively greater weight on object-specific actions than objects themselves in object word comprehension. Instead, it was attributed to children's characteristic understanding of object words that their initial meanings were the undifferentiated fusion entwining both objects and actions. Thus, we provide the first experimental evidence on the semantic pluripotency hypothesis which corroborates the theoretical assumptions that the initial meanings of children's words do not sufficiently differentiate into specific categories, such as objects (Nelson, 1983a, 1983b; Nelson & Lucariello, 1985; Werner & Kaplan, 1963).

Second, we found that the semantic differentiation of object words steeply occurred during the latter half of the second year of life, although the critical period of this occurrence was different between the preliminary (Chapter 2) and confirmatory (Chapter 4) experiments. A large body of research on children's early word learning has focused on this developmental period because it is a period when their cognitive and linguistic abilities develop notably in various respects, such as the vocabulary spurt (Kobayashi, Minami, & Sugiyama, 2012; Mervis & Bertrand, 1995), the use of naming for object categorization (Nazzi & Gopnik, 2001), and the use of known words to learn new words (Ferguson, Graf, & Waxman, 2014, 2018). Although it is still unknown how the semantic differentiation of object words is related to the improvement of such abilities, the fact that the transition of these factors is overlapping would open a new window for future research on investigating children's word learning mechanisms. Relatedly, the differentiation of word meanings may occur in conjunction with children's symbolic play because when engaging in pretend play, they have to independently operate object shape and its relevant action (Level 5, according to McCune's (2008) classification).

Based on the semantic pluripotency hypothesis, we further examined the relationship between the semantic differentiation of object words from global event categories to specific object categories and vocabulary growth. We posited that once object word meanings are sufficiently differentiated into specific object categories, toddlers should be encouraged (or even urged) to develop new label-meaning connections that could be used for specific actions because they could no longer use object words to designate actions. As expected, the results from Chapter 4 revealed that the degree of semantic differentiation measured by task performance in the mismatch condition was positively related to concurrent and later verb vocabulary sizes. Thus, we succeeded in providing the first experimental evidence to clarify how semantic differentiation affects vocabulary growth. The assumption that young children at first use early words for global events as a whole and subsequently differentiate word meanings into more specific categories, has been investigated mostly by careful observation of children's behavior and language use from one-word to multi-word utterance periods, such as gesture-speech combinations (Özçalışkan & Goldin-Meadow, 2005), different use of intonation (Werner & Kaplan, 1963), or successive single-word utterances (Herr-Israel & McCune, 2011). Our findings would not only experimentally support this view but also suggest a new method to examine the degree of semantic differentiation without demands for children's word production.

The methodological investigation of the equivalent interpretability of pointing and looking responses also played a significant role in a deep understanding of children's word meaning formation and their differentiation (Chapter 3). We constructed models that could predict pointing responses (left or right) from concurrent preferential looking, and the results on prediction accuracy revealed that children's looking behavior was closely related to their judgment of object word referents. Both the created models and these results make two major contributions to our subsequent research (Chapter 4). First, by leveraging the created model, we were able to apply the same two-alternative forced-choice task used in Chapter 2 to younger toddlers without increasing missing data. Second, the application of the developed model to discretize looking responses allowed us to conduct the same statistical analysis for both arm and gaze indices, which contributed to a direct comparison between two different measurements. In particular, similar results were obtained for both pointing and looking responses in the model selection approach, but task performance estimated by pointing responses tended to be slightly better than that estimated by looking responses. This suggests that although cognitive processes measured by these two indices are common and overlap to some extent, pointing and looking responses would reflect qualitatively different aspects of children's word comprehension. As discussed by Hendrickson et al. (Hendrickson & Friend, 2013; Hendrickson, Mitsven, Poulin-Dubois, Zesiger, & Friend, 2015; Hendrickson, Poulin-Dubois, Zesiger, & Friend, 2017), toddlers can produce both pointing and looking responses when word meanings are at least partially formed or robust, but they may likely show only looking responses when such meanings are still being formed or are rather fragile. If we had analyzed only pointing responses, we might have obtained biased results because trials in which toddlers did not produce clear pointing responses were excluded.

Overall, it appears convincing that the central purpose of this thesis, as the first step to verify the semantic pluripotency hypothesis, was fulfilled. However, as we will discuss later, there are some limitations.

5.2 The semantic pluripotency hypothesis: revisited

In Chapter 1, we proposed the semantic pluripotency hypothesis to address the semantic contents of children's early words and their development experimentally. This proposal consists of two sub-hypotheses. First, the initial meanings of words have pluripotent characteristics, as they correspond to the context-bounded fusion of various factors that emerge from children's experiences related to certain words. Second, the development of word meanings has plasticity, as word meanings dynamically differentiate into specific, discrete, and more decontextualized categories. To discuss future directions, we attempt to explain the semantic pluripotency hypothesis and our findings in the context of a dynamic systems approach (Thelen & Smith, 1994). We adopt this approach because it is a powerful tool to describe dynamic changes in stable patterns of observed phenomena with a focus on both children's internal (cognitive) and external (environmental) factors, and it would contribute to connecting the semantic pluripotency hypothesis to other related frameworks such as statistical word learning (Chen & Yu, 2017; Sandhofer & Schonberg, 2020; Scott & Fisher, 2012; Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007, 2011; Yurovsky, Yu, & Smith, 2013).

The dynamic systems approach originated in the fields of physics, chemistry, and mathematics that focused on nonlinear patterns observed in natural phenomena. In this approach, structures or patterns that seem ordered and complex emerge not from a kind of design intent or a blueprint but from interactions among individual components as a process of "self-organization." Thelen and Smith (1994) applied this view to child development in order to provide a unified account of ontogeny in various domains (e.g., action and cognition) and time scales. Although developmental research on dynamic systems primarily evolved in the field of perception and action, Thelen and Bates (2003) argued that it shared many assumptions with the connectionism approach widely used in research on linguistic processing (see also Smith & Samuelson, 2003). In fact, some researchers already utilized a dynamic systems approach to explain children's language development (McCune, 2008; McCune, Lennon, & Greenwood, 2020; McCune & Zlatev, 2015; Parladé & Iverson, 2011; Samuelson, Schutte, & Horst, 2009; Twomey, Ma, & Westermann, 2018; see also Spencer, Perone, & Buss, 2011; Spencer & Schöner, 2003).

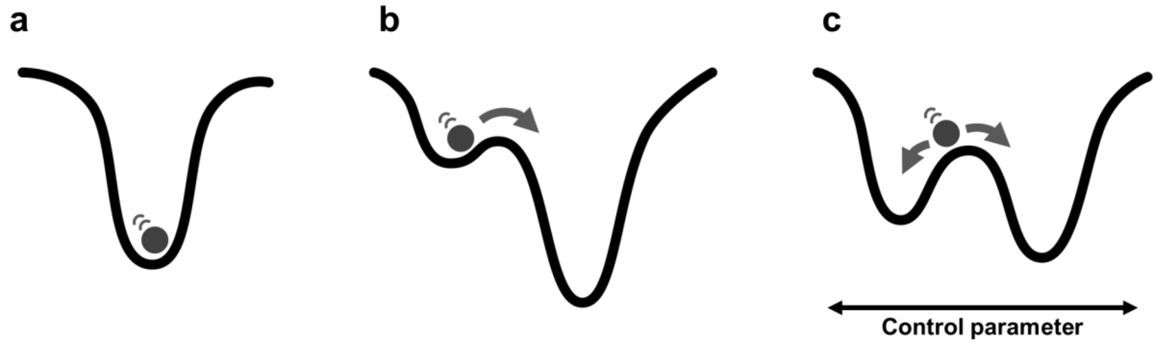

Figure 5.1 Schematic examples of attractor states. (a) The ball stays in a deep attractor, indicating the system demonstrates a stable pattern. (b) The ball at first stays in a shallow unstable attractor, which will result in the ball dropping into a deeper attractor. (c) As the value of the control parameter changes, the ball becomes less likely to stay in a certain attractor and it seeks more stable modes.

A key term of a dynamic systems approach is an "attractor," which refers to a (relatively) stable state in a complex system. The correspondence between the structure of attractors within a system and the phenomena observed at a given time point is often illustrated by the relationship between bowls and a rolling ball, and a factor that influences which attractor the ball is likely to stay in (i.e., phase transitions) is called a "control parameter" (Thelen & Smith, 1994). Although the location of the ball always fluctuates owing to noise derived from each component of the system, the ball stays around the bottom of the bowl if the attractor is sufficiently deep. In contrast, when the ball is temporally located around a shallower attractor and there is an adjacent deeper attractor, the ball will prefer to shift to the deeper well if enough time is given. A large change in the value of a control parameter also influences the preference of attractors[22] (**Figure 5.1**). For example,

[22] When moving the control parameter, there are critical points where the previous stable pattern cannot be maintained and transient behavior with no stable pattern becomes dominant. By further moving the parameter, the system evolves into a new or different stable attractor state and it becomes less fluctuated. This "critical fluctuation" is a potent tool to evaluate the stability of attractor states in a complex system and to detect a control parameter that has a large impact on the dynamism of the system. Also, a dynamic systems approach puts emphasis on increased fluctuations because it implies that a new stable pattern is emerging. For instance, Parladé and Iverson (2011) reported that children who showed a clear vocabulary spurt demonstrated a great instability in their communicative system during the period of the vocabulary spurt (e.g., words became less likely to be produced in coordination with other communicative behaviors such as gestures) compared to those who showed a less clear spurt, suggesting that a prominent transition in language development occurred before and after the spurt.

horses have three different but continuous gait patterns (i.e., attractors) of a walk, trot, and gallop, and when putting them on a treadmill, the treadmill's running speed is the control parameter that affects the phase transition (Hoyt & Taylor, 1981). As the speed of the treadmill increased, the gait pattern was more likely to shift from walking to trotting to galloping.

Let us explain the semantic pluripotency hypothesis using the terminology of a dynamic systems approach. When a word has an undifferentiated meaning, its semantic system is susceptible to various factors related to the word, and its meaning is less robust (i.e., a shallow attractor). Specifically, the meaning of the object word "shoe" would initially arise as a vague attractor for which both object and action features are the control parameters. Therefore, if both features match the word (e.g., "putting shoes on"), the ball can stay in the attractor (**Figure 5.2**), whereas it is difficult for it to stay in the attractor and it is easily bounced out of it if either feature is less shoe-like. This semantic system could result in the tendency of younger toddlers to comprehend the object word referents in the match condition but not in the mismatch condition (Chapters 2 and 4). With development, the attractor would become much deeper, indicating that the word meaning becomes more robust and stable so that it can be easily identified and distinguished from other word meanings. In response to this deepening, it would become more susceptible to object features such as shape, while being less susceptible to action features[23]. This means that object features are still control parameters, but action features lose the ability to affect object word comprehension when it comes to the judgment of referents of the word "shoe". This change in the structure of attractor states could be the process of cognitive development underlying the increase in task performance in the mismatch condition. The finding that task performance remained at chance level in the different-object and absent-object conditions

[23] This process may be similar to that of infants' specialization of phonetic perception into their native languages (Kuhl, 2011; Kuhl et al., 2006). Although both U.S. and Japanese infants aged 6–8 months could discriminate between English /r/ and /l/ sounds, the accuracy of this phonetic perception for infants aged 10–12 months increased only in U.S. children whereas it declined in those learning Japanese (the Japanese phonological system does not discriminate between /r/ and /l/ sounds). Interestingly, better discrimination performance in a phonetic contrast in the native language was related to a rapid increase in later language abilities, while that in the non-native language was related to slower growth in those abilities (Kuhl et al., 2008; Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). As we will see in the next section, Gogate and Maganti (2017) recently found a similar transition from the language-general to the language-specific tendency of vocabulary development.

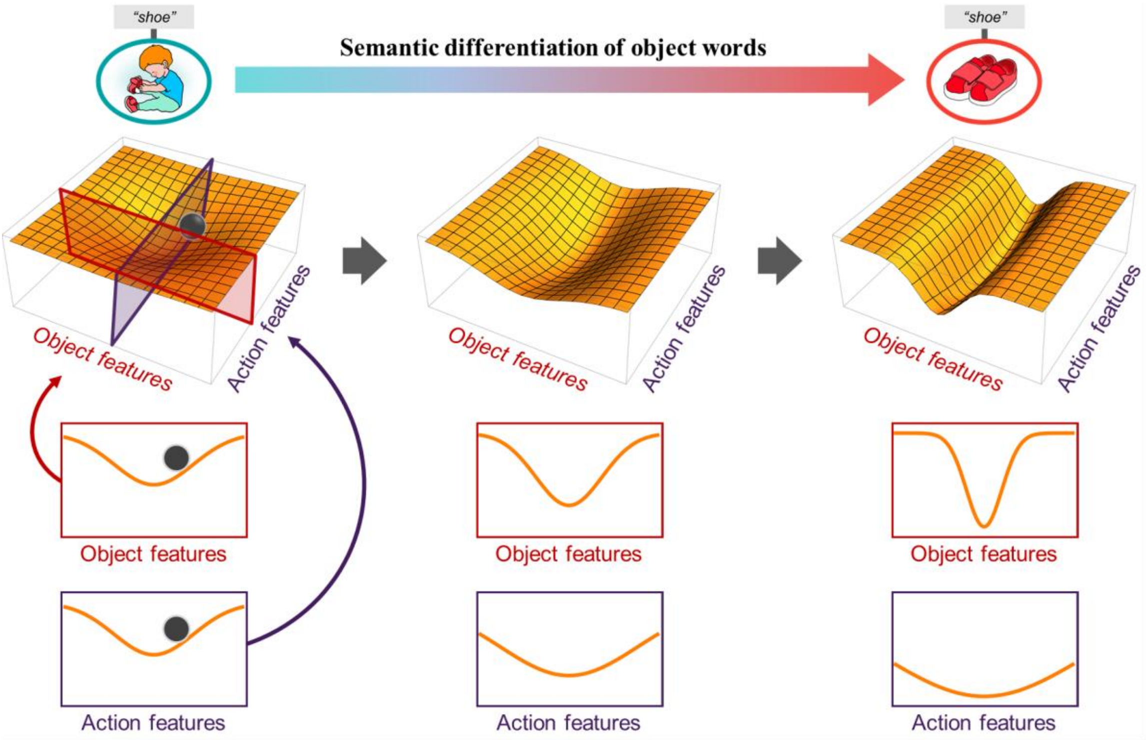

Figure 5.2 A schematic view of developmental change in attractor states of the semantic system. When a word has an undifferentiated global meaning, the attractor is still shallow, vague, and susceptible to both object and action features (left figures). The semantic differentiation would be explained as the structural change in attractor states so that the depth of attractor increases (i.e., the word meaning becomes robust) and the semantic system becomes more susceptible to object features (e.g., shape) while being less susceptible to action features (e.g., object-specific action; middle and right figures). The ball in the left figure indicates children's judgment of object word referents for a certain trial in the forced-choice task.

throughout the target age range could be explained by a similar process in which the ball did not stay around the attractor of object word meanings because object features as a control parameter were much less shoe-like (Chapter 4). In addition, the depth of attractors represents the robustness of word meanings, which can be estimated by whether toddlers show both looking and pointing responses (i.e., a deep attractor), or show only looking responses (i.e., a shallow attractor). Note that during this change in the semantic system, past history would not be completely lost. The structure of attractors "emerges from dynamic interactions of multiple, coupled components including the learner's body, learning history, and in-the-moment characteristics of the task" (Twomey et al., 2018, p. 415). This notion is compatible with Nelson's view that "earlier representational processes (e.g., event scripts) are not lost but are supplemented by more abstract levels of functioning"

(Nelson, 1983b, p. 147).

At this point, even if the explanation discussed above is plausible, it is unclear what occurs during semantic differentiation: whether the gradient of object features becomes steeper, that of action features becomes gentler, or both. It should also be determined how to quantify object and action features as continuous variables and how many dimensions are needed to quantify them (e.g., there are several factors that characterize object features, such as shape, size, or color). Despite such limitations, we believe that the refinement of the semantic pluripotency hypothesis using a dynamic systems approach is still beneficial for strengthening and directing future work, as well as providing a unified explanation for relevant prior studies. For example, Goldenberg and Sandhofer (2013) reported that two-year-old toddlers could learn novel object words more appropriately when learning in a situation of combining the same and varied background contexts compared to when learning under a situation of either the same or varied ones (see also Sandhofer & Schonberg, 2020). They interpreted from these results that the combined support for aggregation of instances and decontextualization of important features from other ones would be necessary for novice word learners to appropriately generalize novel labels. We expect that by leveraging the semantic pluripotency hypothesis, the effect of the same and varied contextual backgrounds could be more precisely separated and described. Learning under the same background might help in the rough estimation of the location of the attractor that corresponds to the word meaning and learning under the varied background might help to modulate the gradient of the attractor in values of various features related to the learning situation.

If we use a dynamic systems approach even more boldly, the unified developmental process underlying the differentiation of nouns and verbs is expected to be modeled. It could be discovered that there is a dimension-reduced single feature that characterizes both object and action features. This discovery would enable us to represent the process of semantic differentiation of early words into object and action words as a bifurcation model (**Figure 5.3**). In a bifurcation model, the deepening of one attractor is related to the formation of another attractor. Hence, the semantic differentiation of noun-like words from events to specific object categories directly predicts the emergence of other types of words that mean action categories.

These ideas are speculative at this point. However, such an explanation and prediction would contribute to the development of new research questions and to the construction of computational models of young children’s word learning.

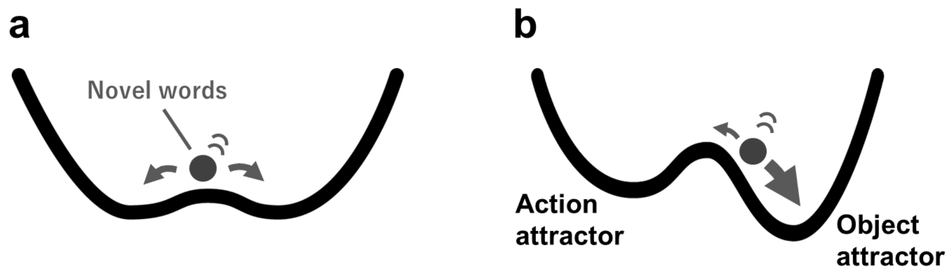

Figure 5.3 A schematic view of a bifurcation model representing the semantic differentiation of object and action words. (a) When words have undifferentiated meanings, the attractor looks vague and less branched. (b) Once the semantic differentiation of object words happens, a novel word becomes more likely to be understood as it means object category. This bifurcation predicts the formation of the other attractor that corresponds to action categories.

5.3 Limitations and future directions

To further verify the semantic pluripotency hypothesis, we raise several limitations of our experiments and provide our plan for future research in this section. We utilized only one experimental method in this thesis to examine the semantic pluripotency hypothesis, and we definitely have to adopt other methodologies such as eye-tracking, novel word learning, and natural environment studies. Specifically, since we have regarded it as evidence to support our hypothesis that the results of the probability in the mismatch condition remain near chance level, whereas those in the match condition are above chance, some may not be sufficiently convinced of the existence of the initial pluripotent state of words. In future work, we acknowledge that positive results must be shown that children whose words mean undifferentiated event category show higher performance compared to those whose words correspond to specific categories. In the following subsections, we discuss limitations with reference to such aspects as well.

5.3.1 What are the “core meanings” of object words?

First, in order to exhaustively examine the “core meanings” of object words, a new condition should be added in future experiments using the forced-choice task. Throughout this thesis, we revealed that object-word comprehension is influenced by object-specific actions for younger toddlers. Despite this fact, we also found that toddlers could not understand object word referents solely by object-specific actions (i.e., without the exact target objects). These results suggest that toddlers do not place relatively greater weight on object-specific actions than objects themselves in object word comprehension. Then, how much weight is placed on objects themselves? To address this question, the absent-action condition, in which toddlers have to judge object word referents based solely on objects, would be needed, as well as the absent-object condition. However, the preparation of this condition might be difficult because if it was made of static images of target and filler objects to exclude actions, task performance using static images and videos could not be easily compared due to the qualitative difference of stimuli. Instead, we are considering processing the video stimuli in the match condition so that only objects appear to be moving by making the girl transparent and invisible.

It seems plausible that the core meanings of object words are not object-specific actions but the exact objects; however, the extent to which object features are weighted compared to other factors and how such weightings change with development is still open to exploration. When using static images, even 6–9-month-old infants can direct their gaze to the referents that match familiar nouns such as food or body parts (Bergelson & Swingley, 2012; Tincoff & Jusczyk, 2012), suggesting that the presence or absence of exact objects is the most crucial control parameter of object word comprehension. However, this does not necessarily mean that infants of this age already have robust object-word connections. We assume that the attractor of such word meanings should be quite shallow at first, given that, as we disclosed in this thesis, at least actions affect object word comprehension for 18-month-old toddlers, as does who spoke the words (i.e., a mother or experimenter), for 9-month-old infants (Parise & Csibra, 2012). Although word learning is often regarded in an all-or-none fashion, it takes long and gradual processes for the initial meanings of object words to metamorphose into elaborate ordinary ones like those for adults (Werner & Kaplan, 1963). It would thus be interesting and essential to investigate how the attractor of object word meanings deepens with development and how the semantic system of such words becomes less susceptible to various factors other than object features, such as shape. For example, future studies should aim to clarify whether such changes are gradual and continuous or rather abrupt and discontinuous, as well as how much these developmental patterns vary among individuals.

5.3.2 Are the undifferentiated meanings of object words observed in noun-friendly languages as well?

Second, whether the semantic differentiation of object words from global to specific categories is a language-general phenomenon should be investigated. As in other noun-friendly languages such as English and French, common nouns account for the largest proportion of early vocabulary in Japanese (Ogura, Watamiki, & Inaba, 2016); however, Japanese is classified as a relatively verb-friendly language (Imai, Haryu, & Okada, 2005; Waxman et al., 2013) because, for example, verb meanings are narrower in Japanese than in English, as different verbs like "*kiru*," "*haku*," "*kaburu*," and "*kakeru*" are used to express wearing a shirt, pants, hat, and glasses, respectively. In addition, Japanese has other distinctive features that allow it to drop both the subject and object of a sentence (Imai et al., 2005; Imai et al., 2008), and its infant-directed speech contains richer onomatopoeic use than that of English, even when referring to objects (Fernald & Morikawa, 1993). Such onomatopoeic expressions embed ambiguous and intuitive information compared to ordinary words (Komatsu, 2012). Japanese would be one of the languages in which the semantic pluripotency hypothesis is highly applicable because of these language-specific characteristics. However, since Werner and Kaplan's (1963) hypothesis was developed mainly on the basis of observational studies of noun-friendly languages such as English, we conjecture that the semantic differentiation from global events to specific object categories would be a language-general phenomenon. Research using an experiment similar to ours for toddlers learning noun-friendly languages will reveal whether the semantic pluripotency hypothesis can be applied to other languages.

Although we assume that the existence of the initial pluripotent states of early words and the principles of semantic differentiation are cross-linguistically common, the timing of the differentiation of object word meanings may differ depending on language. English-speaking caregivers tend to utter object words more frequently and consistently than Japanese caregivers when interacting with their children (Fernald & Morikawa, 1993). Similar cross-cultural differences have been reported in other languages, such that English-speaking parents produced more nouns than verbs compared to parents speaking Korean (Choi, 2000) or Mandarin (Tardif, Shatz, & Naigles, 1997). This noun-biased input in English may enhance earlier differentiation of object words for children learning English compared to those learning verb-friendly languages[24]. In fact, English-learning infants start to understand familiar object words several months earlier than familiar relational words such as "eat" (Bergelson & Swingley, 2013). In addition, the timing when the semantic differentiation of object words occurs may also vary from word to word because the degree of dominance of object word input varies depending on the context, even within the same

[24] For the same chronological age, both comprehension and production vocabulary sizes of U.S. infants are larger than those of Japanese infants (Ogura, 2007; Tamis-LeMonda, Bornstein, Cyphers, Toda, & Ogino, 1992).

language. Cross-linguistically, caregivers likely produce more nouns than verbs in the book-reading context, whereas this balance is reversed in the toy-play context (Altınkamış, Kern, & Sofu, 2014; Ogura, 2007; Tardif, Gelman, & Xu, 1999; but see Choi, 2000). Furthermore, parental language use systematically differs depending on activities (e.g., feeding or grooming) in terms of, for example, the amount, diversity, and semantic content (Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019). Considering these findings, cross-linguistic and cross-context differences in the semantic differentiation of early words should be investigated.

5.3.3 Do other parts of speech also have pluripotent characteristics at first?

In addition to the cross-linguistic and cross-contextual examination of the semantic pluripotency hypothesis, it is worthwhile to investigate whether there is a semantically undifferentiated phase in other parts of speech (e.g., action words), as there is in object words. Surprisingly, Nomikou, Rohlfing, Cimiano, and Mandler (2019) reported that German 10-month-olds, unlike 9-month-olds, could recognize image stimuli corresponding to action words based solely on objects (e.g., an image of a banana for "eating"). Therefore, early words might initially encompass undifferentiated event categories, regardless of nouns or verbs. If so, the fact that in some languages, such as Mandarin, children learn verbs earlier than nouns (Brown, 1998; Choi & Gopnik, 1995; Tardif, 1996; Tardif et al., 1999, 2008; see also Frank, Braginsky, Yurovsky, & Marchman, in press) could be explained as they do not learn verbs as words corresponding to specific action categories but rather as words corresponding to global event categories. Cross-part-of-speech investigation on semantic differentiation may reveal whether noun dominance over verbs is observed because object categories are easier to form than action categories, or for other reasons: nouns in many languages have simpler and invariant phonetic features for easy pronunciation compared to verbs[25], nouns are more frequently produced by caregivers, and nouns are more likely to be produced in a particular context. Future studies that aim to distinguish between before and after the semantic differentiation of action words for Mandarin-speaking children would be helpful in seeking a reasonable explanation for the learning process of early words regardless of parts of speech and in suggesting a new solution to the long-standing

[25] Unlike English and Japanese, Mandarin has no verb conjugation.

controversy over the developmental order of nouns and verbs.

We assume that most words at first correspond to undifferentiated meanings in which various components are fused within certain contexts or situations regardless of nouns, verbs, adjectives, etc. Many observational case reports support this assumption, although very little experimental evidence has been provided. Church (1961) stated that children's initial adjective of "hot" may mean hot things such as soup and stoves rather than a single attribute. Kobayashi (1992) reported that a 19-month-old girl pointed toward the kitchen in response to her mother's question of "Which one is the spoon?" despite having a spoon and a dish in front of her at the time [26]. The Kyoto Scale of Psychological Development 2001, which is often used in Japan as a developmental assessment, includes the task of answering object names while looking at object illustrations, and an example of answering "father" to the image of clothing is introduced as an incorrect answer in the manual (Ikuzawa, Matsushita, & Nakase, 2002; Nakase & Nishio, 2001).

According to the case reports and the findings of this thesis, we expected that the semantic pluripotency hypothesis could be extended to prelinguistic phenomena, such as joint attention. Joint attention refers to a state where both members of a dyad focus on objects or events simultaneously, while being aware of the other members' shared attention; (Markus, Mundy, Morales, Delgado, & Yale, 2000; Morales et al., 2000; Mundy & Newell, 2007; Scaife & Bruner, 1975), the behaviors associated with joint attention (e.g., gaze following or pointing) play a role in later language development (Brooks & Meltzoff, 2008; Cameron-Faulkner et al., 2020; Colonnesi, Stams, Koster, & Noom, 2010; Tomasello & Farrar, 1986; Tomasello, Mannle, & Kruger, 1986). Although an attention overlap between a caregiver and a young child during joint attention is plausible, the focal point between them could differ. For instance, when a caregiver points to something and utters a word, an infant responding to it (i.e., engaging in joint attention) may capture a wider range of the

[26] During the experiments in Chapters 2 and 4, several participants showed similar responses, such that they pointed toward the shoe cupboard when asked "Which one is the shoes?" by the experimenter. Note that people do not usually wear shoes indoors in Japanese culture. Furthermore, when playing with toddlers in nursery schools in order to build a rapport with participants, the author observed several similar cases. For instance, during the "What is this?" game, several toddlers uttered "*Ame*" [rain] in response to the image of an umbrella. Even a 32-month-old girl answered "*Zubon*" [pants] and "*Origami*" to the images of a jacket and a pencil, respectively. Also, a 34-month-old girl said the attribute "*Midori*" [green] to the image of a pencil despite saying nouns like "*Megane*" [glasses] and "*Baiku*" [bike or motorcycle] in response to other images.

Figure 5.4 An example of joint attention and word learning that is based on the semantic pluripotency hypothesis. A caregiver and an infant are looking at trains passing a railroad crossing. When the caregiver points to the crossing and says "*Kankan*" (the onomatopoeic expression for a railroad crossing), the child might associate this word to the global event that includes vehicles, motions, sounds, etc., even though the caregiver intends only the crossing as a target object. Therefore, the dyad shares a scene, but their mental images could possibly differ from each other.

scene (objects, actions, etc.) than the caregiver's intended focal point (**Figure 5.4**). Currently this is merely an assumption; however, further investigations combining a novel word learning task and other experiments, similar to those utilized in this thesis, will disclose the characteristics of infants' instant formation of attractors of word meanings through the process of joint attention.

5.3.4 What drives the formation and differentiation of object word meanings?

We successfully clarified that object word meanings transformed from undifferentiated global event categories to differentiated specific object categories. Therefore, one of the most essential future steps would be to investigate what causes this semantic differentiation.

We hypothesized that decontextualization (Lucariello, Kyratzis, & Engel, 1986; Sandhofer & Schonberg, 2020; Twomey et al., 2018) would play a central role in this process. Sandhofer and Schonberg (2020) stated that variations in the learning context and the features within the target categories contribute to decontextualization or abstraction of word meanings. For example, the semantic differentiation of the object word "shoes" might be facilitated by hearing someone talking about shoes in their absence, seeing shoes in various environments (e.g., doorway or shop), or learning unconventional uses of shoes (e.g., shoes as vases). Furthermore, based on the semantic pluripotency hypothesis using a dynamic systems approach, variations in the contextual background may make the target word meaning attractor less vulnerable to irrelevant parameters (i.e., the slopes of the attractor along such parameters become shallower and flatter), while within-category variability may induce more sensitivity in the attractor toward important parameters (i.e., the attractor along critical parameters deepens).

Future research should explore the factors that foster the differentiation of initial word meanings alongside those that facilitate children to connect verbal sounds to undifferentiated holistic events. Contextual consistency and variability would contribute to the formation of undifferentiated words (aggregation) and the reorganization of differentiated words (decontextualization), respectively, and both play important roles in early word learning (Goldenberg & Johnson, 2015; Goldenberg & Sandhofer, 2013; Sandhofer & Schonberg, 2020). Caregivers may act as a scaffold to facilitate formation and differentiation of word meanings in children, while modulating the significance of those cues in accordance with the child's developmental age, regardless of whether they do it consciously or subconsciously. For example, initially, parents may encourage children to form connections between labels and undifferentiated events and gradually change their style to promote semantic differentiation. Similarly, caregivers reduce the complexity of their initial utterances when children are learning a word, and then gradually increase the level (Roy, 2009). In addition, mothers of infants (5–8 months) are more likely to utter words while moving the respective target objects in synchrony (i.e., abundant motionese) in a novel word learning task. However, mothers reduce this synchronized naming and motion with children's increasing age (Gogate, Bahrick, & Watson, 2000). Thus, adults can provide different types of cues for word learning (see Hollich et al., 2000 for classification), where some contribute to the formation of semantically undifferentiated words and others facilitate semantic differentiation. Future research on the identification and categorization of these cues is highly recommended to understand the ecological development of word learning, achieved through the joint cooperation of children and their caregivers, as well as to design effective and practical interventions for children with language development delays.

5.3.5 Do semantically undifferentiated words play a key role in word learning?

According to the pluripotency hypothesis, young children initially comprehend words as holistic, global, and undifferentiated concepts. Does this phenomenon occur merely because they cannot appropriately dissect word meanings due to immaturity, or because this process of early word learning results in some benefit to development of linguistic abilities later in life[27]? Several researchers hypothesized that children's limited capabilities may paradoxically be indicative of adaptive advantages for achieving more complex and elaborate abilities (Newport, 1988, 1990; Turkewitz & Kenny, 1982). For instance, recent computational research revealed that poor visual resolutions typically seen in infants are advantageous for better facial recognition during later stages of development (Vogelsang et al., 2018).

Additionally, we believe that undifferentiated event categories for initial word meanings might provide a basis for syntax development for children by facilitating the retention of implicit relationships between individual components embedded within a particular context; this retention may be sustained even after the word meanings are sufficiently differentiated. This idea is consistent with the theoretical assumptions that the meanings of early labels are “total situations in which agent, action, and object are intimately fused” (Werner & Kaplan, 1963, p. 116) and that early definition aspects of events are sustained even after their decontextualization (Nelson, 1983b). If children learn the meanings of each word as specific categories from the very beginning, this may burden children to learn how to connect individual categories to form sentences later in life. However, learning vague and undifferentiated early word meanings with interrelated characteristics might be efficient in learning how to combine words, due to the existing knowledge of semantic relationships between words retained after the semantic differentiation. This view could be indirectly and partially supported by the finding that the

[27] This insightful question was provided by Kawai (2019) where he intended to provide a suggestion bridging the semantic pluripotency hypothesis and computational modeling approach. Using a hidden Markov model, he constructed a computational model with his colleagues such that the disambiguation of novel word referents became more successful with the finer partitioning of categories (Kawai, Oshima, Sasamoto, Nagai, & Asada, 2015, 2020).

appropriate differentiation of word meaning affects children's initial syntactic comprehension. For instance, Werner and Kaplan (1963) cited the case reported by Guillaume (1927), in which a 2 year old was able to understand "*brosse papa*" [brush daddy] and "*brosse maman"* [brush mama], but not "*brosse le chapeau*" [brush the hat], probably because the word "*chapeau*" corresponded to the undifferentiated event of "putting on a hat" at that point for the child (Guillaume, 1927, p. 13). Furthermore, Uriu (1992) investigated children's responses to the instruction "*Ningyo o pōn-shite goran*" [Throw the doll][28], while presenting a pair of a doll as the target object and a ball as the perplexing object, simultaneously (see also Uriu, 1986). The resulting variety of actions implied the developmental changes in children's interpretation of such instructions: (1) they just threw the ball (up to 21 months), (2) they either throw the doll or throw the ball to the doll (up to 30 months), and (3) they appropriately throw the doll (after 4.5 years). These findings suggest it is much more difficult to independently interpret words from other contexts, initially, because of their undifferentiated meanings, which may seem disadvantageous for the pluripotent characteristics of early words. However, such context-bound meanings might be crucial for forming syntactic associations between words[29]. This hypothesis needs to be elaborated on and investigated further.

Additionally, semantic pluripotency may have another adaptive advantage in

[28] Unlike English, Japanese allows dropping the subject and object of a sentence, and its syntactic structure is understood primarily by case markers rather than the order of words. For example, in the sentence "*Kutsu o haku*" [putting shoes on], "*o*" is a postpositional particle indicating that the preceding word "*kutsu*" [shoes] is the object, and the overt expression of "*o*" represents the two other words (i.e., "*kutsu*" and "*haku*" in this case) in the transitive frame. Studies showed that 2 year old Japanese toddlers could utilize such case markings in learning whether novel words are transitive or intransitive verbs (Kobayashi & Suzuki, 2014a, 2014b; Matsuo, Kita, Shinya, Wood, & Naigles, 2012).

[29] Early syntax development and its relations to word meanings have been well investigated based on the "syntactic bootstrapping hypothesis," which focuses on young children's abilities to make use of argument structure while learning novel words (Fisher, Gertner, Scott, & Yuan, 2010; Fisher, Jin, & Scott, 2020; Gleitman, 1990; Lidz, Gleitman, & Gleitman, 2003; Yuan, Fisher, & Snedeker, 2012). For instance, the number of noun phrases in a sentence allows children to infer whether a novel verb in the sentence is transitive or intransitive (e.g., the sentence "Bill *larped* a ball" contains two noun phrases—"Bill" and "a ball"—implying that the word "*larp*" is a transitive verb). Although it may be difficult to integrate the syntactic bootstrapping hypothesis and the semantic pluripotency hypothesis because the former basically holds a nativist view and the latter holds an emergentist one. Future studies that intend to combine both these hypotheses will contribute to a deeper understanding of how semantics and syntax are related during infancy and toddlerhood.

language development; unlike discrete and specific word meanings, vague and undifferentiated meanings may facilitate flexibility in learning, regardless of semantic categories, which enables children to learn any language. Gogate and Maganti (2017) demonstrated that infants (8–9 months), learning English or Spanish (noun-friendly languages), were able to learn novel label-action connections, whereas older infants (12–14 months) did not. Additionally, they found marginal evidence that older infants with a poor noun vocabulary could relatively learn such action words, but those with a larger vocabulary range could not. These findings suggest children temporarily lose their flexibility and general abilities to learn various types of words, as they specialize in native noun-friendly languages (see also Gogatea & Hollich, 2016). We assume that a similar developmental pattern can be observed in future experiments based on the semantic pluripotency hypothesis. For instance, based on our findings that semantic differentiation of object words occurs in the latter half of the second year of life, after learning the connections between novel labels and objects, 15-month-olds may be able to relearn the same words as labels corresponding to actions because their semantic attractors are still shallow. However, 24-month-olds may be unable to relearn because deeper and steeper attractors are formed during their initial learning. This "meaning-switch" task (**Figure 5.5**) would be potent to examine whether the initial word meanings can differentiate into various specific categories[30]. If so, it can be said that early words have not only undifferentiated but also literally pluripotent characteristics. Thus, the successful discovery of advantageous pluripotent features in early words will provide crucial support to the semantic pluripotency hypothesis.

5.4 Implications for caregivers, educators, and clinicians

We did not conduct any pedagogical or clinical investigations in this thesis. However, we believe that the semantic pluripotency hypothesis is potentially helpful for all practitioners in deeply understanding the language development during infancy and early childhood and planning subsequent strategies for effective treatment. In this section, we attempt to

[30] For this investigation, novel labels should be morpho-syntactically neutral since morpho-syntactic aspects of words would also provide cues for learning. When developing the experimental design and materials, referring to Tomasello and Akhtar (1995) would be helpful. They reported that when using a syntactically ambiguous novel label, 2-year-olds could learn object and action words equally well depending on whether the object or action was emphasized under instructions.

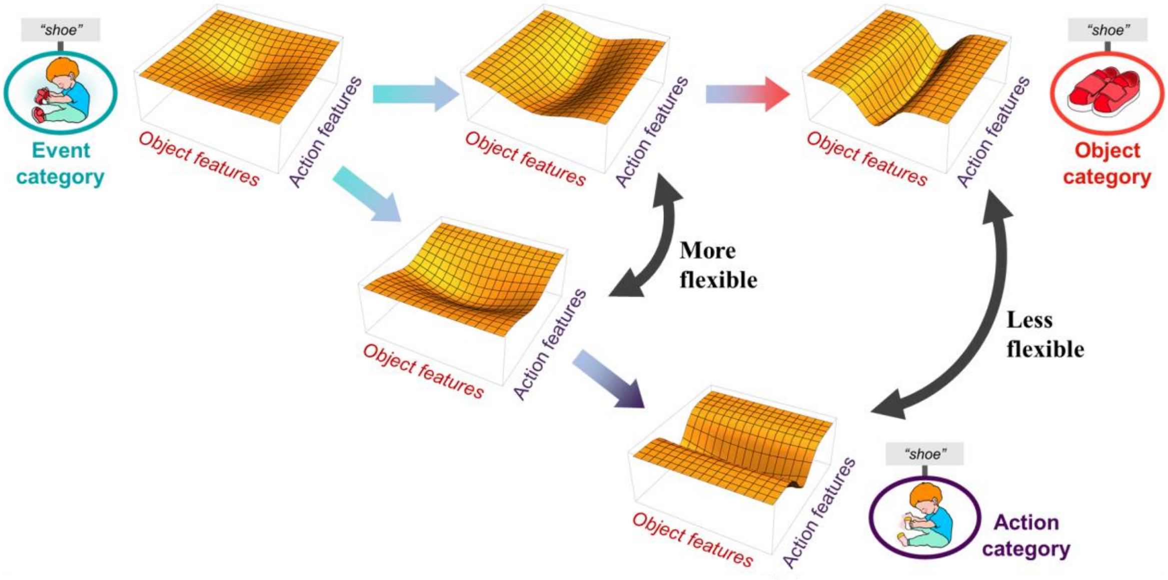

Figure 5.5 A schematic view of the hypothesis to be examined using a "meaning-switch" task. For younger toddlers, the structure of the attractor is still shallow although slightly differentiated, which allows the word meanings to be flexibly interchangeable between object and action categories. For older toddlers, the structure of the attractor is sufficiently deep and rigid, resulting in inflexible word meanings.

summarize the findings of studies potentially relevant to the semantic pluripotency hypothesis that could benefit caregivers, educators, and clinicians.

For readers' comprehension, we introduce the iceberg analogy for children's language (**Figure 5.6**), proposed by a speech therapist, Nakagawa (1986, 2009), which facilitates a clinical and practical perspective. She developed it as a simple explanation for worried caregivers whose children did not start speaking or those who wanted to know of preferred ways to foster their child's language development. This perspective is highly compatible with the semantic pluripotency hypothesis, particularly due to its emphasis on the importance of increasing the number of comprehensible holistic events for children, as a foundation for early language development, rather than teaching individual label-meaning connections or forcing them to utter words. The iceberg analogy by Nakagawa (1986, 2009) explained that children's productive words are only the tip of the iceberg, and there are many more words and events that children can understand, which lie beneath the surface of the water. As an example, she suggests that when a child drinks a glass of juice in front of them, rather than throwing the glass or splashing the juice on their head, it indicates that the child understands the "drinking-juice" event. Knowledge about such events would be termed as a "script" by Nelson (1983a, 1983b). During this event, caregivers would have

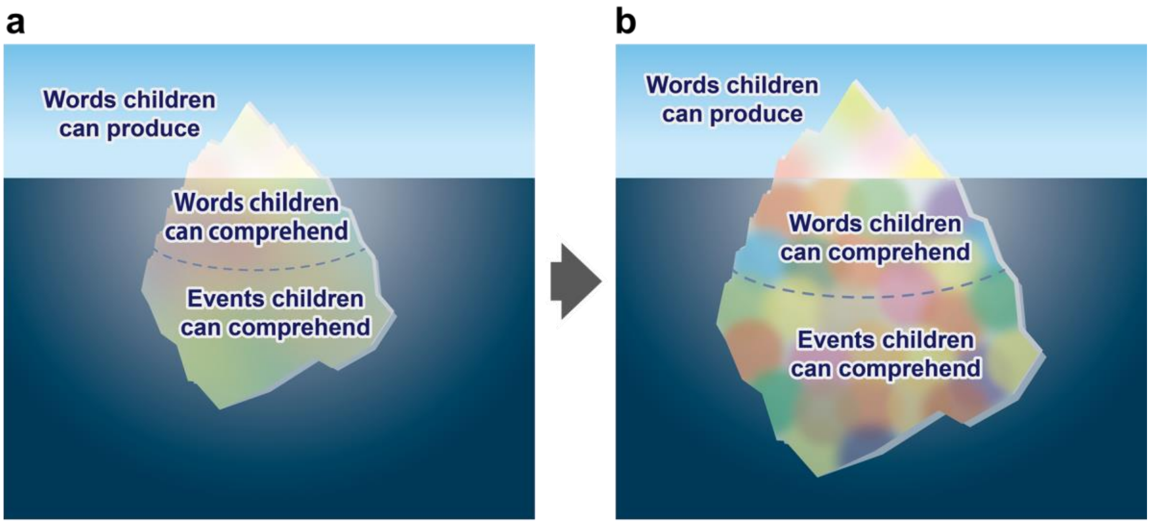

Figure 5.6 A modified perspective on young children's language as an iceberg. Adults can only observe children's productive words (the tip of the iceberg). However, there are many more comprehensible words than productive ones, and they are supported by comprehensible events that are embedded in particular contexts or situations (underneath the surface of the water). The original analogy was proposed by Nakagawa (1986, p. 109). We integrated the semantic pluripotency hypothesis into this perspective: (a) children's comprehensible words and events are initially blurry and undifferentiated, which would result in ambiguous productive words (of course there are other factors, such as the development of vocal motor control). (b) With development, the quantity as well as the quality of children's language change. In particular, comprehensible words and events are partitioned and become independent of other words or events, which would result in the improvement of word manipulability (represented by the color figure). This would help children combine words and produce multi-word utterances.

often said "Let's have some juice" or "Be careful not to spill your glass," which builds the foundation for increasing children's comprehensible words. Thus, this iceberg analogy implies that for developing the upper tip, it is essential to effectively foster its lower part. In **Figure 5.6**, we have integrated the semantic pluripotency hypothesis with this iceberg perspective. Therefore, it suggests that children's comprehensible words and events develop and expand quantitatively and qualitatively from holistic undifferentiated categories to specific independent categories. Thus, the attempts to nurture the lower part of the iceberg could be executed through two different strategies: (1) fostering the holistic word meaning formation or (2) promoting the semantic differentiation of words.

We assume that enhancing the saliency[31] of contexts as a whole or contextual coherency may be more effective to provide a scaffolding for the semantic formation of the initial words, rather than emphasizing only a particular aspect in context. Although it is possible that naming or labeling itself plays a privileged role in children's categorization (Althaus & Plunkett, 2016; Fairchild, Mathis & Papafragou, 2018; Ferry, Hespos, & Waxman, 2010; Fulkerson & Waxman, 2007; Havy & Waxman, 2016; Markman & Hutchinson, 1984; Pickron, Iyer, Fava, & Scott, 2018; Twomey & Westermann, 2018; Xu, 2002; but see Deng & Sloutsky, 2015), combining labeling with multiple cues may be more helpful in forming global event categories as word meanings[32] (or aggregation). Moreover, increasing the number of correlated contextual cues may increase the likelihood of learning (Sandhofer & Schonberg, 2020). In addition, multiple correlated cues would facilitate parallel learning of different categories in young children (Sloutsky & Robinson, 2013), and redundant use of socio-pragmatic cues (e.g., pointing and gazing) combined with labeling would positively influence toddlers' novel word learning (Booth, McGregor, & Rohlfing, 2008; see also Gogate, Maganti, & Bahrick, 2015 for the abundant use of the multimodal motherese).

According to the semantic pluripotency hypothesis and its relevant theories (Nelson, 1983a, 1983b, 1985), children's event category formation is primarily important for later language development. Hence, the articulation of event boundaries might encourage the organization of children's ongoing events. For instance, for 20-month-olds, task performance in delayed memory retrieval of cartoons dropped when the perception of event boundaries within the cartoons was disturbed (however, the same was not true for 16-month-olds) (Sonne, Kingo, & Krøjgaard, 2016). Nakagawa (2009) also noted that it would be preferable to use verbal cues indicating event boundaries from a clinical perspective (e.g.,

---

[31] Despite not being motivated by practical purposes, Wildt, Rohlfing, & Scharlau's (2019) theoretical classification of saliency might help in considering the integration of redundant cues: perceptual saliency attributed to objects (e.g., color, size), perceptual saliency generated in perceivers (e.g., novelty), social saliency given to receivers (e.g., gaze, gesture), and social saliency emerging in active interaction (e.g., joint action).

[32] Even in learning situations where verbal labels are not necessarily required, redundant use of verbal and nonverbal cues may be useful; for example, 12-month-old infants showed high memory retrieval in a delayed imitation task when using verbal input (Taylor, Liu, & Herbert, 2016). In parenting support practices, Nakagawa (2009) argued that it is important to utter contextually relevant words slowly, repeatedly, and in as concise sentences as possible to facilitate children's comprehension of words and events.

utterances “Put on your shoes” and “You did it!” before and after a child put her shoes on, respectively).

Another possible strategy to promote young children’s initial formation of words might be to use onomatopoeic or gestural expressions rather than conventional words. For instance, Imai and Kita (2014) claimed that sound symbolic associations, such as relations between onomatopoeias and referents, foster multisensory integration and semantic processing. Kita (1997) also hypothesized that mimetic words, iconic gestures, and expressive prosodies are highly relevant to mental information of subjective experiences, accompanied by emotional imagery; this informational unit is referred to as the “proto-eventuality.” Furthermore, mimetic words bridge the proto-eventuality dimension with the decontextualized or analytic language dimension. Such primitive expressions often lead to linguistic ambiguity because they seem to lack morpho-syntactic markers compared to conventional words. However, such expressions could provide a powerful scaffolding for detecting rough locations of the attractor corresponding to their meanings[33] because onomatopoeic or gestural words are more easily related to bodily perceptions than ordinary linguistic ones[34].

Next, we briefly introduce cues that would potentially promote the semantic differentiation of words. As discussed above, contextual and within-category variabilities would play an important role in decontextualizing word meanings (Sandhofer & Schonberg, 2020). Additionally, utterances that combine known words with novel words might be beneficial in helping children understand the meanings of the known word more precisely

[33] Furthermore, it might also be useful to adopt sound patterns, indicative of a particular event, that a child can express as a label. According to the case report by Hagihara (2018), a pediatric occupational therapist failed to share conventional short phrases such as “*Mouikkai*” [Again] with a child with autism spectrum disorder during the play session, but the child could gradually understand and make use of the bricolage word “*Kika*” as a label that indicated “Do it again!” once the therapist started using the child’s spontaneous vocalization as a communicative sign (i.e., “*Kika*”). Such bricolage words seem inconvenient due to the lack of generalization, but they could contribute to organizing the initial word meanings for children.

[34] During the experiments in this thesis, we often observed that a participant responded to the question of “Which ones are shoes?” by touching or looking at his or her bare feet, implying that initial word comprehension is related to embodiment. Note that people usually do not wear shoes indoors in the Japanese culture. Similarly, we observed in our recent preliminary study using an eye-tracker that a toddler looked at a filler object and a girl’s mouth in a video stimulus alternately when asked to choose the stimulus that corresponded to the word “cup.”

and infer the meaning of the unknown word. For example, unlike 15-month-olds, 19-month-old toddlers could identify the referents of artificial novel nouns (e.g., vep) by making use of known verbs (e.g., crying) in an intermodal preferential looking task (Ferguson et al., 2014). Specifically, after hearing sentences such as "The vep is crying," toddlers looked at animate images for longer than at inanimate ones. It was also revealed that children's ability to leverage semantic knowledge was more efficient at 24 months (Ferguson et al., 2018). Therefore, once children can understand a few word meanings to a certain extent, it would be beneficial to repeat children's utterances with a little expansion or structural change (for example, when a child says, "This is a cup," a caregiver would say "This is a *red* cup"). Such expansions of adults' utterances would promote the realization in children that word meanings can be divided further into partitioned and discrete pieces, and that co-occurring words depict different aspects of the event. This strategy, called the "recast," might also be effective for early syntax development because children would be able to compare their own utterances with those of adults and learn how to combine words (Nelson, 1977).

Children face several difficulties on their journey of language development, such as referential indeterminacy (i.e., "Gavagai problem;" Quine, 1960) and a chaotic perception of environments initially (i.e., "one great blooming, buzzing confusion;" James, 1890, p. 488). "Fortunately, infants are not alone in this endeavor" (Mason, Goldstein, & Schwade, 2019, p. 60); adults are good companions to confront these difficulties with children. Becoming a careful observer like those who conducted classic diary-based studies (e.g., Guillaume, 1927; Kamhi, 1986; Lewis, 1936; Stern & Stern, 1907) may be difficult; however, we speculate that some readers have subconsciously been providing adaptive support for children's language development, as discussed above. Thus, it may not be necessary to do something special to nurture children's language in most cases. Instead, it would be much more important to discover the subconscious strategies during everyday involvement with children, which may empower you as well.

Needless to say, the extent of the effect from each assistance on the initial formation and differentiation of word meanings is still open to verification. In addition, the process of semantic differentiation among children with various developmental patterns (e.g., autism spectrum disorder) should be further explored and investigated[35] as in other related topics such as shape bias (Abdelaziz, Kover, Wagner, & Naigles, 2018; Field, Allen, & Lewis, 2016; Tek & Naigles, 2017; Tovar, Rodríguez-Granados, & Arias-Trejo, 2020), to suggest

[35] The experimental tasks utilized in this thesis would be useful to evaluate the extent to which object word meanings differentiate into specific categories.

ways of effectively supporting early word learning. Nevertheless, we believe that the semantic pluripotency hypothesis has significant applications in several fields as it provides insight into the internal mechanisms of children's language that cannot be captured solely by utterances. This thesis disclosed that the word meanings for young children may not necessarily be identical to those for adults and that semantic contents of early words dynamically change with development, which is of scientific and practical significance.

5.5 Concluding remarks

In this thesis, we proposed the semantic pluripotency hypothesis as a revision of Werner and Kaplan's (1963) theoretical hypothesis on children's early words. Our experimentally verifiable hypothesis opens a new window to explore the developmental process of word learning. This hypothesis casts doubt on the seemingly plausible but misleading assumption that children understand labels as words meaning specific categories similar to those for adults. This thesis elucidates the uniqueness of children's initial word meanings and their differentiation process. It implies that noun-like words do not mean specific object categories initially, instead they correspond to an undifferentiated fusion of both objects and actions; this finding provides the first experimental support for Werner and Kaplan's (1963) theoretical hypothesis. In addition, this thesis theoretically contributes to early language development by providing empirical evidence of the developmental connection between object word and action word learning such that the semantic differentiation of object words predicts the vocabulary growth of action words. Furthermore, the semantic pluripotency hypothesis provides a new possible explanation for how young children solve the "Quine's problem" (**Figure 5.7**). Traditionally, it has been believed that children face difficulty determining the correct combination of novel labels and their referents. However, the semantic pluripotency hypothesis assumes that children may solve this problem and achieve successful word learning by temporarily and ambiguously associating novel labels with the holistic situations in which they experience, and gradually differentiate word meanings; this process is similar to the development of facial recognition (Vogelsang et al., 2018). Addressing this assumption would make significant contributions to the relevant research fields.

Quine's discussion regarding the Gavagai problem was originally a thought experiment examining how a field linguist would translate an unknown local language (Quine, 1960). He explained that there were always other potential frameworks for understanding an unknown language (i.e., manual of translation) than the one used by the

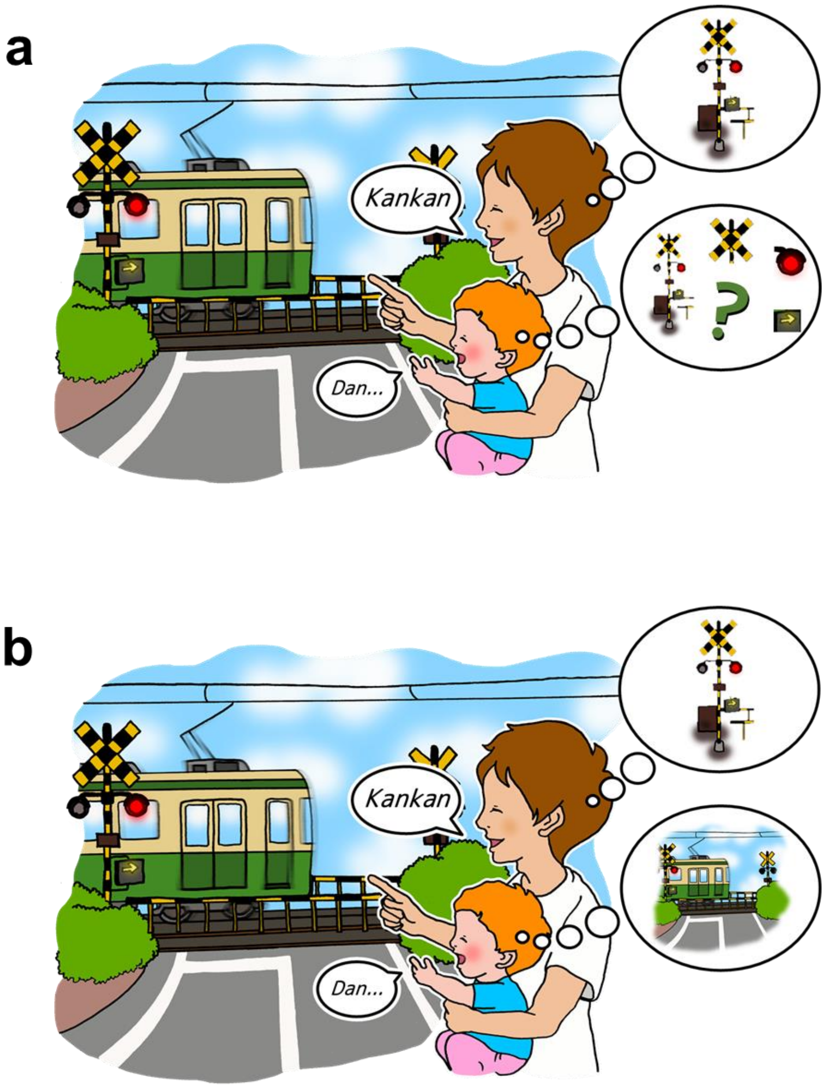

Figure 5.7 Different solutions for "Quine's problem." (a) The traditional view of the problem which young children face. Children have to determine which part of the scene is the appropriate referent upon hearing a novel word. (b) The semantic pluripotency hypothesis perspective. When hearing a novel word, children associate novel labels temporarily and ambiguously with the holistic scene. Gradually, they make differentiated word meanings.

linguist while translating (Furuta, 1998). In the field of language development, the children have been regarded as field linguists, as termed by Quine; however, it could be said that us researchers (or adults) are ourselves the field linguists who translate children's language as unknown beings. Therefore, while investigating and discussing children's language, we must consider that frameworks other than the adults' completed ones (e.g., specific semantic categories or parts of speech) may exist. We believe that the semantic pluripotency hypothesis could be the new manual of translation for redrawing the distinctive characteristics of early language development from children’s points of view despite being in its foundational stage.

References

Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three-year-old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences? A permutation analysis. *PloS One*, *12*(10), e0186129. https://doi.org/10.1371/journal.pone.0186129

Abdelaziz, A., Kover, S. T., Wagner, M., & Naigles, L. R. (2018). The shape bias in children with autism spectrum disorder: Potential sources of individual differences. *Journal of Speech, Language, and Hearing Research*, *61*(11), 2685–2702. https://doi.org/10.1044/2018_JSLHR-L-RSAUT-18-0027

Acredolo, L. P., & Goodwyn, S. W. (1985). Symbolic gesturing in language development: A case study. *Human Development*, *28*(1), 40–49. https://doi.org/10.1159/000272934

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, *59*(2), 450–466. https://doi.org/10.2307/1130324

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Akobeng, A. K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatrica*, *96*(5), 644–647. https://doi.org/10.1111/j.1651-2227.2006.00178.x

Aktan-Erciyes, A., & Göksun, T. (2019). Early event understanding predicts later verb comprehension and motion event lexicalization. *Developmental Psychology*, *55*(11), 2249–2262. https://doi.org/10.1037/dev0000804

Althaus, N., & Plunkett, K. (2016). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*, *19*(5), 770–780. https://doi.org/10.1111/desc.12358

Altınkamış, N. F., Kern, S., & Sofu, H. (2014). When context matters more than language: Verb or noun in French and Turkish caregiver speech. *First Language*, *34*(6), 537–550. https://doi.org/10.1177/0142723714560179

Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *WIREs Cognitive Science Interdisciplinary Reviews*, *4*(2), 149–168. https://doi.org/10.1002/wcs.1215

Ahmed, A., & Ruffman, T. (1998). Why do infants make A not B errors in a search task, yet show memory for the location of hidden objects in a nonsearch task?. *Developmental Psychology*, *34*(3), 441–453. https://doi.org/10.1037/0012-1649.34.3.441

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, *17*(2), 1265–1282. https://doi.org/10.1016/S0885-2014(02)00116-8

Baillargeon, R., & Wang, S. H. (2002). Event categorization in infancy. *Trends in Cognitive Sciences*, *6*(2), 85–93. https://doi.org/10.1016/S1364-6613(00)01836-2

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development*, *72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Baltrušaitis, T., Zadeh, A., Lim, Y. C., & Morency, L.-P. (2018). Openface 2.0: Facial behavior analysis toolkit. *Proceedings of 13th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2018)* (pp. 59–66). Xi'an, China: IEEE. https://doi.org/10.1109/FG.2018.00019

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6*(2), 183–200. https://doi.org/10.1017/S0305000900002245

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, *127*(3), 391–397. https://doi.org/10.1016/j.cognition.2013.02.011

Black, A., & Bergmann, C. (2017). Quantifying infants' statistical word segmentation: A meta-analysis. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 124–129). Austin, TX: Cognitive Science Society. Retrieved from https://cogsci.mindmodeling.org/2017/papers/0035/

Blinkoff, E., & Hirsh-Pasek, K. (2019). Supporting language in the home. *International Journal of Birth and Parent Education*, 6(4), 13–15. Retrieved from https://www.early-education.org.uk/sites/default/files/Joint%20EE%20IJBPE%20Summer%202019.pdf#page=15

Bloom, P. (2004). Myths of word learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 205–224). Cambridge, MA: MIT Press.

Bloom, L., Tinker, E., & Margulis, C. (1993). The words children learn: Evidence against a noun bias in early vocabularies. *Cognitive Development*, *8*(4), 431–450. https://doi.org/10.1016/S0885-2014(05)80003-6

Booth, A. E. (2006). Object function and categorization in infancy: Two mechanisms of facilitation. *Infancy*, *10*(2), 145–169. https://doi.org/10.1207/s15327078in1002_3

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, *4*(3), 179–202. https://doi.org/10.1080/15475440802143091

Booth, A. E., Schuler, K., & Zajicek, R. (2010). Specifying the role of function in infant categorization. *Infant Behavior and Development*, *33*(4), 672–684. https://doi.org/10.1016/j.infbeh.2010.09.003

Booth, A. E., & Waxman, S. (2002). Object names and object functions serve as cues to categories for infants. *Developmental Psychology*, *38*(6), 948–957. https://doi.org/10.1037/0012-1649.38.6.948

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., ... & Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115–1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Bornstein, M. H., & Hendricks, C. (2012). Basic language comprehension and production in > 100,000 young children from sixteen developing nations. *Journal of Child Language*, *39*(4), 899–918. https://doi.org/10.1017/S0305000911000407

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for ‘motionese’: Modifications in mothers’ infant-directed action. *Developmental Science*, *5*(1), 72–83. https://doi.org/10.1111/1467-7687.00211

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(1), 207–220. https://doi.org/10.1017/S030500090700829X

Brown, P. (1998). Children’s first verbs in Tzeltal: Evidence for an early verb category. *Linguistics*, *36*(4), 713–753. https://doi.org/10.1515/ling.1998.36.4.713

Bruner, J. S. (1983). *Child’s talk: Learning to use language*. New York, NY: Norton.

Butler, J., & Rovee-Collier, C. (1989). Contextual gating of memory retrieval. *Developmental Psychobiology*, *22*(6), 533–552. https://doi.org/10.1002/dev.420220602

Cameron-Faulkner, T., Malik, N., Steele, C., Coretta, S., Serratrice, L., & Lieven, E. (2020). A cross-cultural analysis of early prelinguistic gesture development and its relationship to language development. *Child Development*, *92*(1), 273–290. https://doi.org/10.1111/cdev.13406

Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers’ word retrieval. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1468–1480. https://doi.org/10.1044/1092-4388(2005/102)

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carey, S. & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, *15*, 17–29.

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26*(1), 69–111. https://doi.org/10.1017/S0305000998003687

Chan, A., Meints, K., Lieven, E., & Tomasello, M. (2010). Young children's comprehension of English SVO word order revisited: Testing the same children in act-out and intermodal preferential looking tasks. *Cognitive Development*, *25*(1), 30–45. https://doi.org/10.1016/j.cogdev.2009.10.002

Charles, E. P., & Rivera, S. M. (2009). Object permanence and method of disappearance: Looking measures further contradict reaching measures. *Developmental Science*, *12*(6), 991–1006. https://doi.org/10.1111/j.1467-7687.2009.00844.x

Chen, C. H., & Yu, C. (2017). Grounding statistical learning in context: The effects of learning and retrieval contexts on cross-situational word learning. *Psychonomic Bulletin & Review*, *24*, 920–926. https://doi.org/10.3758/s13423-016-1163-x

Chen, J., Tardif, T., Pulverman, R., Casasola, M., Zhu, L., Zheng, X., & Meng, X. (2015). English- and Mandarin-learning infants' discrimination of actions and objects in dynamic events. *Developmental Psychology*, *51*(10), 1501–1515. https://doi.org/10.1037/a0039474

Childers, J. B., Porter, B., Dolan, M., Whitehead, C. B., & McIntyre, K. P. (2020). Does children's visual attention to specific objects affect their verb learning?. *First Language*, *40*(1), 21–40. https://doi.org/10.1177/0142723719875575

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, 27(1), 69–96. https://doi.org/10.1017/S0305000999004018

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, *22*, 497–529. https://doi.org/10.1017/S0305000900009934

Chouinard, B., Scott, K., & Cusack, R. (2019). Using automatic face analysis to score infant behavior from video collected online. *Infant Behavior and Development*, *54*, 1–12. https://doi.org/10.1016/j.infbeh.2018.11.004

Christiansen, M. H., & Monaghan, P. (2006). Discovering verbs through multiple-cueintegration. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 88–107). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.001.0001

Church, J. (1961). *Language and the discovery of reality: A developmental psychology of cognition*. New York, NY: Random House. https://doi.org/10.1037/14589-000

Church, J. (1966). *Three babies*. New York, NY: Random House.

Clark, E. V. (1973) What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York, NY: Academic Press. https://doi.org/10.1016/B978-0-12-505850-6.50009-8

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review*, *30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1149–1157. https://doi.org/10.1098/rstb.2010.0319

Delle Luche, C., Durrant, S., Poltrock, S., & Floccia, C. (2015). A methodological investigation of the Intermodal Preferential Looking paradigm: Methods of analyses, picture selection and data rejection criteria. *Infant Behavior and Development*, *40*, 151–172. https://doi.org/10.1016/j.infbeh.2015.05.005

Deng, W. S., & Sloutsky, V. M. (2015). Linguistic labels, dynamic visual features, and attention in infant category learning. *Journal of Experimental Child Psychology*, *134*, 62–77. https://doi.org/10.1016/j.jecp.2015.01.012

DePaolis, R. A., Keren-Portnoy, T., & Vihman, M. (2016). Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in Psychology*, *7*, 715. https://doi.org/10.3389/fpsyg.2016.00715

Dewey, J. (1894). The psychology of infant language. *Psychological Review*, *1*(1), 63–66. https://doi.org/10.1037/h0065788

Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: An alternative account of Samuelson and Smith. *Developmental Science*, *7*(1), 33–41. https://doi.org/10.1111/j.1467-7687.2004.00320.x

Dore, J. (1975). Holophrases, speech acts and language universals. *Journal of Child Language*, *2*(1), 21–40. https://doi.org/10.1017/S0305000900000878

Dore, J. (1978). Conditions for the acquisition of speech acts. In I. Markova (Ed.), *The social context of language* (pp. 87–111). Chichester, UK: John Wiley.

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, *42*(2), 447–465. https://doi.org/10.1017/S0305000914000063

Eiteljoerge, S. F., Adam, M., Elsner, B., & Mani, N. (2019a). Consistency of co-occurring actions influences young children's word learning. *Royal Society Open Science*, *6*(8), 190097. https://doi.org/10.1098/rsos.190097

Eiteljoerge, S. F., Adam, M., Elsner, B., & Mani, N. (2019b). Word-object and action-object association learning across early development. *PLoS One*, *14*(8), e0220317. https://doi.org/10.1371/journal.pone.0220317

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye-tracking study. Are 18-month-old late talkers really different from their typical peers?. *Journal of Communication Disorders*, *58*, 143–157. https://doi.org/10.1016/j.jcomdis.2015.06.011

Estes, K. G., & Lew-Williams, C. (2015). Listening through voices: Infant statistical word segmentation across multiple speakers. *Developmental Psychology*, *51*(11), 1517–1528. https://doi.org/10.1037/a0039725

Fairchild, S., Mathis, A., & Papafragou, A. (2018). Linguistic cues are privileged over non-linguistic cues in young children's categorization. *Cognitive Development*, *48*, 167–175. https://doi.org/10.1016/j.cogdev.2018.08.007

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*(3644), 668–670. https://doi.org/10.1126/science.146.3644.668

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5, Serial No. 242). https://doi.org/10.2307/1166093

Fenson, L., Dale, P. S., Reznick, J. S., Thai, D., Bates, E., Hartung, J. P., Pethick, S., Relly, J. S. (1993). *MacArthur communicative development inventories: User's guide and technical manual*. San Diego, CA: Singular Press.

Fenson, L., Marchman, V. A., Thai, D. J., Dale, P. S., Reznick, J. S., Bates, E. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual* (2nd ed). Baltimore, MD: Brookes.

Ferguson, B., Graf, E., & Waxman, S. R. (2014). Infants use known verbs to learn novel nouns: Evidence from 15- and 19-month-olds. *Cognition*, *131*(1), 139–146. https://doi.org/10.1016/j.cognition.2013.12.014

Ferguson, B., Graf, E., & Waxman, S. R. (2018). When veps cry: Two-year-olds efficiently learn novel words from linguistic contexts alone. *Language Learning and Development*, *14*(1), 1–12. https://doi.org/10.1080/15475441.2017.1311260

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*(3), 637–656. https://doi.org/10.2307/1131208

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231. https://doi.org/10.1111/1467-9280.00044

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics: On-line methods in children's language processing* (pp. 97–135). Philadelphia, PA: John Benjamins.

Fernandes, K. J., Marcus, G. F., Di Nubila, J. A., & Vouloumanos, A. (2006). From semantics to syntax and back again: Argument structure in the third year of life. *Cognition*, *100*(2), B10–B20. https://doi.org/10.1016/j.cognition.2005.08.003

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Field, C., Allen, M. L., & Lewis, C. (2016). Are children with autism spectrum disorder initially attuned to object function rather than shape for word learning?. *Journal of Autism and Developmental Disorders*, *46*(4), 1210–1219. https://doi.org/10.1007/s10803-015-2657-5

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*(1), 41–81. https://doi.org/10.1006/cogp.1996.0012

Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *WIREs Cognitive Science Interdisciplinary Reviews*, *1*(2), 143–149. https://doi.org/10.1002/wcs.17

Fisher, C., Jin, K. S., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77. https://doi.org/10.1111/tops.12447

Frank, M. C. (2019). Towards a more robust and replicable science of infant development. *Infant Behavior and Development*, *57*, 101349. https://doi.org/10.1016/j.infbeh.2019.101349

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (in press). *Variability and consistency in early language learning: The Wordbank project*. Cambridge, MA: MIT Press. Retrieved from https://langcog.github.io/wordbank-book/index.html

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24. https://doi.org/10.1080/15475441.2012.707101

Friend, M., & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 302–309. https://doi.org/10.3758/BF03202556

Friend, M., & Keplinger, M. (2008). Reliability and validity of the Computerized Comprehension Task (CCT): data from American English and Mexican Spanish infants. *Journal of Child Language*, *35*(1), 77–98. https://doi.org/10.1017/S0305000907008264

Friend, M., & Pace, A. (2011). Beyond event segmentation: Spatial- and social-cognitive processes in verb-to-action mapping. *Developmental Psychology*, *47*(3), 867–876. https://doi.org/10.1037/a0021107

Friend, M., & Pace, A. E. (2016). Action interrupted: Movement and breakpoints in the processing of motion violations in toddlers and adults. *Journal of Cognition and Development*, *17*(1), 105–121. https://doi.org/10.1080/15248372.2015.1016611

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218–228. https://doi.org/10.1016/j.cognition.2006.09.005

Furuta, T. (1998). What is the indeterminacy of translation?. *Kagaku Kisoron Kenkyu*, *25*(2), 83–88 (in Japanese). https://doi.org/10.4288/kisoron1954.25.83

Ganger, J., & Brent, M. R. (2004). Reexamining the vocabulary spurt. *Developmental Psychology*, *40*(4), 621–632. https://doi.org/10.1037/0012-1649.40.4.621

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development, Vol. 2: Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Lawrence Erlbaum.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, UK: Cambridge University Press.

Gershkoff-Stowe, L., & Hahn, E. R. (2013). Word comprehension and production asymmetries in children and adults. *Journal of Experimental Child Psychology*, *114*(4), 489–509. https://doi.org/10.1016/j.jecp.2012.11.005

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, *75*(4), 1098–1114. https://doi.org/10.1111/j.1467-8624.2004.00728.x

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science*, *1*(1), 31–35. https://doi.org/10.1111/1467-8721.ep10767853

Glick, J. A. (1992). Werner's relevance for contemporary developmental psychology. *Developmental Psychology*, *28*(4), 558–565. https://doi.org/10.1037/0012-1649.28.4.558

Gogate, L., & Hollich, G. (2016). Early verb-action and noun-object mapping across sensory modalities: A neuro-developmental view. *Developmental Neuropsychology*, *41*(5–8), 293–307. https://doi.org/10.1080/87565641.2016.1243112

Gogate, L., & Maganti, M. (2017). The origins of verb learning: Preverbal and postverbal infants' learning of word-action relations. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3538–3550. https://doi.org/10.1044/2017_JSLHR-L-17-0085

Gogate, L., Maganti, M., & Bahrick, L. E. (2015). Cross-cultural evidence for multimodal motherese: Asian Indian mothers' adaptive use of synchronous words and gestures. *Journal of Experimental Child Psychology*, *129*, 110–126. https://doi.org/10.1016/j.jecp.2014.09.002

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Goldenberg, E. R., & Johnson, S. P. (2015). Category generalization in a new context: The role of visual attention. *Infant Behavior and Development*, *38*, 49–56. https://doi.org/10.1016/j.infbeh.2014.12.001

Goldenberg, E. R., & Sandhofer, C. M. (2013). Same, varied, or both? Contextual support aids young children in generalizing category labels. *Journal of Experimental Child Psychology*, *115*(1), 150–162. https://doi.org/10.1016/j.jecp.2012.11.011

Goldfield, B. A., & Reznick, J. S. (1996). Measuring the vocabulary spurt: a reply to Mervis & Bertrand. *Journal of Child Language*, *23*(1), 241–246. https://doi.org/10.1017/S0305000900010175

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition*, *4*(2), 189–202. https://doi.org/10.1016/0010-0277(76)90004-4

Goldstein, M., Elmlinger, S., & Schwade, J. (2020, July). *Relative contributions of infant-directed speech and motion when learning new words*. Paper presented at the Virtual International Congress of Infant Studies (vICIS) 2020. Abstract retrieved from https://infantstudies.org/wp-content/uploads/2020/07/vICIS-Symposia-Abstract-BookletJuly1FINAL.pdf

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, *15*(1), 30–33. https://doi.org/10.1111/j.0963-7214.2006.00401.x

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, *28*(1), 99–108. https://doi.org/10.1037/0012-1649.28.1.99

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, *14*(1), 23–45. https://doi.org/10.1017/s030500090001271x

Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science*, *8*(3), 316–339. https://doi.org/10.1177/1745691613484936

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*(1), 125–155. https://doi.org/10.1017/S0305000900008692

Goodwyn, S., & Acredolo, L. (1993). Symbolic gesture versus word: Is there a modality advantage for onset of symbol use? *Child Development*, *64*(3), 688–701. https://doi.org/10.2307/1131211

Göksun, T., Hirsh-Pasek, K., & Golinkoff, R. M. (2010). Trading spaces: Carving up events for learning language. *Perspectives on Psychological Science*, *5*(1), 33–42. https://doi.org/10.1177/1745691609356783

Göksun, T., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Konishi, H., & Okada, H. (2011). Who is crossing where? Infants' discrimination of figures and grounds in events. *Cognition*, *121*(2), 176–195. https://doi.org/10.1016/j.cognition.2011.07.002

Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, *26*(2), 295–320. https://doi.org/10.1017/S0305000999003815

Guillaume, P. (1927). Les débuts de la phrase dans le langage de l'enfant [The beginnings of the sentence in the child's language]. *Journal de Psychologie*, *24*, 1–25.

Gurteen, P. M., Horne, P. J., & Erjavec, M. (2011). Rapid word learning in 13-and 17-month-olds in a naturalistic two-word procedure: Looking versus reaching measures. *Journal of Experimental Child Psychology*, *109*(2), 201–217. https://doi.org/10.1016/j.jecp.2010.12.001

Hagihara, H. (2018). Bodily interactions with environment: A foundation of language acquisition. *Papers from the 18th National Conference of the Japanese Cognitive Linguistics Association*, *18*, 561–566 (in Japanese).

Hahn, E. R., & Cantrell, L. (2012). The shape-bias in Spanish-speaking children and its relationship to vocabulary. *Journal of Child Language*, *39*(2), 443–455. https://doi.org/10.1017/S030500091100016X

Hahn, E. R., & Gershkoff-Stowe, L. (2010). Children and adults learn actions for objects more readily than labels. *Language Learning and Development*, *6*(4), 283–308. https://doi.org/10.1080/15475441003635315

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development*, *21*(2), 167–179. https://doi.org/10.1016/S0163-6383(98)90001-7

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*(7169), 557–559. https://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*(6), 923–929. https://doi.org/10.1111/j.1467-7687.2010.00951.x

Hammer, R., & Diesendruck, G. (2005). The role of dimensional distinctiveness in children's and adults' artifact categorization. *Psychological Science*, *16*(2), 137–144. https://doi.org/10.1111/j.0956-7976.2005.00794.x

Haryu, E. (2006). Meta-knowledge about word meanings that enables young children to fast-map a various kinds of words to their meanings. *Japanese Psychological Review*, *49*(1), 78–90 (in Japanese). https://doi.org/10.24602/sjpr.49.1_78

Hauschild, K. M., Pomales-Ramos, A., & Strauss, M. S. (2020). The visual array task: A novel gaze-based measure of object label and category knowledge. *Developmental Science*, e13015. https://doi.org/10.1111/desc.13015

Havy, M., & Waxman, S. R. (2016). Naming influences 9-month-olds' identification of discrete categories along a perceptual continuum. *Cognition*, *156*, 41–51. https://doi.org/10.1016/j.cognition.2016.07.011

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, *58*(4), 1021–1034. https://doi.org/10.2307/1130543

Hendrickson, K., & Friend, M. (2013). Quantifying the relationship between infants' haptic and visual response to word-object pairings. In S. Biaz, N. Goldman, & R. Hawkes (Eds.), *Online Proceedings Supplement of BUCLD 37: The 37th Annual Boston University Conference on Language Development*. Boston, MA: Boston University Conference on Language Development. Retrieved from: http://www.bu.edu/bucld/supplementvol37

Hendrickson, K., Mitsven, S., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2015). Looking and touching: What extant approaches reveal about the structure of early word knowledge. *Developmental Science*, *18*(5), 723–735. https://doi.org/10.1111/desc.12250

Hendrickson, K., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2017). Assessing a continuum of lexical-semantic knowledge in the second year of life: A multimodal approach. *Journal of Experimental Child Psychology*, *158*, 95–111. https://doi.org/10.1016/j.jecp.2017.01.003

Hernik, M., & Csibra, G. (2009). Functional understanding facilitates learning about tools in human children. *Current Opinion in Neurobiology*, *19*(1), 34–38. https://doi.org/10.1016/j.conb.2009.05.003

Herr-Israel, E., & McCune, L. (2011). Successive single-word utterances and use of conversational input: a pre-syntactic route to multiword utterances. *Journal of Child Language*, *38*(1), 166–180. https://doi.org/10.1017/S0305000909990237

Hespos, S. J., Grossman, S. R., & Saylor, M. M. (2010). Infants' ability to parse continuous actions: Further evidence. *Neural Networks*, *23*(8–9), 1026–1032. https://doi.org/10.1016/j.neunet.2010.07.010

Hespos, S. J., Saylor, M. M., & Grossman, S. R. (2009). Infants' ability to parse continuous actions. *Developmental Psychology*, *45*(2), 575–585. https://doi.org/10.1037/a0014145

Higuchi, K., Matsuda, S., Kamikubo, R., Enomoto, T., Sugano, Y., Yamamoto, J., & Sato, Y. (2018). Visualizing gaze direction to support video coding of social attention for children with autism spectrum disorder. *Proceedings of 23rd International Conference on Intelligent User Interfaces* (pp. 571–582). Japan, Tokyo: Association for Computing Machinery. https://doi.org/10.1145/3172944.3172960

Hollich, G., Hirsh-Pasek, K., Golinkoff, R., Brand, R., Brown, E., Chung, H. L., Hennon, E., & Rocroi, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, *65*(3, Serial No. 262). https://doi.org/10.1111/1540-5834.00090

Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infants categorize. *Child Development*, *76*(3), 614–631. https://doi.org/10.1111/j.1467-8624.2005.00867.x

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research*, *24*(5), 503–509. https://doi.org/10.1111/jsr.12306

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. *Nature*, *292*(5820), 239–240. https://doi.org/10.1038/292239a0

Huttenlocher, J., & Smiley, P. (1987). Early word meanings: The case of object names. *Cognitive Psychology*, *19*(1), 63–89. https://doi.org/10.1016/0010-0285(87)90004-1

Ikuzawa, M., Matsushita, Y., & Nakase, A. (2002). *Shimpan K-shiki hattatsu kensa 2001 jisshi tebikisho* [*Kyoto scale of psychological development 2001: Administration manual*]. Kyoto, Japan: Kyoto International Social Welfare Exchange Centre.

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: The role of shape similarity in early acquisition. *Cognitive Development*, *9*(1), 45–75. https://doi.org/10.1016/0885-2014(94)90019-1

Imai, M., & Haryu, E. (2014). *Kotoba o oboeru shikumi: Bogo kara gaikokugo made* [*Mechanisms for learning words: From native language to foreign language*]. Tokyo, Japan: Chikumashobo.

Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for Japanese children?. *Child Development*, *76*(2), 340–355. https://doi.org/10.1111/j.1467-8624.2005.00849_a.x

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, *79*(4), 979–1000. https://doi.org/10.1111/j.1467-8624.2008.01171.x

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month-olds. *PLoS One*, *10*(2), e0116494. https://doi.org/10.1371/journal.pone.0116494

Iverson, J. M., Capirci, O., & Caselli, M. C. (1994). From communication to language in two modalities. *Cognitive Development*, *9*(1), 23–43. https://doi.org/10.1016/0885-2014(94)90018-3

James, W. (1890). *The principles of psychology*. New York, NY: Henry Holt.

Ji, L. J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*(1), 1–23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207. https://doi.org/10.1006/cogp.1999.0716

Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *Journal of Cognitive Neuroscience*, *14*(5), 795–805. https://doi.org/10.1162/08989290260138681

Kamhi, A. G. (1986). The elusive first word: The importance of the naming insight for the development of referential speech. *Journal of Child Language*, *13*(1), 155–161. https://doi.org/10.1017/S0305000900000362

Kauschke, C., & Hofmeister, C. (2002). Early lexical development in German: A study on vocabulary growth and vocabulary composition during the second and third year of life. *Journal of Child Language*, *29*(4), 735–757. https://doi.org/10.1017/S0305000902005330

Kawai, Y. (2019). Naze haisei nanoka?: Keisanron no shiten kara [Why pluripotent?: From the perspectives of the computational approach]. *Baby Science*, 18, 29–30.

Kawai, Y., Oshima, Y., Sasamoto, Y., Nagai, Y., & Asada, M. (2015). A model for syntactic development of children: Acquisition processes of syntactic categories reflecting structures of Japanese, English, and Chinese languages. *Cognitive Studies*, *22*(3), 475–479 (in Japanese). https://doi.org/10.11225/jcss.22.475

Kawai, Y., Oshima, Y., Sasamoto, Y., Nagai, Y., & Asada, M. (2020). A computational model for child inferences of word meanings via syntactic categories for different ages and languages. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(3), 401–416. https://doi.org/10.1109/TCDS.2018.2883048

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Chen, W., Ma, W., Ye, Q., & Liu, T-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 3146–3154). Long Beach, CA: Curran Associates. Retrieved from https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree

Kemler Nelson, D. G. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. *Cognitive Development*, *10*(3), 347–380. https://doi.org/10.1016/0885-2014(95)90002-0

Kemler Nelson, D. G., Frankenfield, A., Morris, C., & Blair, E. (2000). Young children's use of functional information to categorize artifacts: Three factors that matter. *Cognition*, *77*(2), 133–168. https://doi.org/10.1016/S0010-0277(00)00097-4

Kemler Nelson, D. G., Russell, R., Duke, N., & Jones, K. (2000). Two-year-olds will name artifacts by their functions. *Child Development*, *71*(5), 1271–1288. https://doi.org/10.1111/1467-8624.00228

Kingo, O. S., & Krøjgaard, P. (2012). Object function facilitates infants' object individuation in a manual search task. *Journal of Cognition and Development*, *13*(2), 152–173. https://doi.org/10.1080/15248372.2011.575424

Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *Linguistics*, *35*(2), 379–415. https://doi.org/10.1515/ling.1997.35.2.379

Kobayashi, H. (1992). Affordance ga sasaeru goikakutoku [Language acquisition supported by affordances]. *Gengo*, *21*(4), 37–45.

Kobayashi, H. (1997). The role of actions in making inferences about the shape and material of solid objects among Japanese 2 year-old children. *Cognition*, *63*(3), 251–269. https://doi.org/10.1016/S0010-0277(97)00007-3

Kobayashi, H. (1998). How 2-year-old children learn novel part names of unfamiliar objects. *Cognition*, *68*(2), B41-B51. https://doi.org/10.1016/S0010-0277(98)00044-4

Kobayashi, T., Minami, Y., & Sugiyama, H. (2012). Goibakuhatsu no atarashii shiten: Nihongo gakushu-ji no shokigoihattatsu ni kansuru juudan-deta kaiseki [A new perspective on the vocabulary spurt: Longitudinal data analysis on early vocabulary development in children learning Japanese]. *Baby Science*, 12, 40–64. Retrieved from https://www.crn.or.jp/LABO/BABY/LEARNED/12/KOBAYASHI_GAKKAISHI.pdf

Kobayashi, T., & Suzuki, T. (2014a). Can Japanese-speaking 2-year-olds use the accusative case marker to infer the meaning on novel verbs?. *IEICE Technical Report*, *113*(426), 67–72 (in Japanese). Retrieved from https://www.ieice.org/ken/paper/201402017BjK/

Kobayashi, T., & Suzuki, T. (2014b). Inferring verb meanings from syntactic frames by Japanese two-year-old children: An experimental approach using an IPL paradigm with a dialogue phase. *Japanese/Korean Linguistics*, *22*, 71–82. Retrieved from https://web.stanford.edu/group/cslipublications/cslipublications/ja-ko-contents/toc-jako22.shtml

Komatsu, T. (2012). Quantifying Japanese onomatopoeias: Toward augmenting creative activities with onomatopoeias. In J.-M. Seigneur (Ed.), *Proceedings of the 3rd Augmented Human International Conference* (Article No. 15). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2160125.2160140

Kominsky, J. F. (2019). PyHab: Open-source real time infant gaze coding and stimulus presentation software. *Infant Behavior and Development*, *54*, 114–119. https://doi.org/10.1016/j.infbeh.2018.11.006

Konishi, H., Stahl, A. E., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Individual differences in nonlinguistic event categorization predict later motion verb comprehension. *Journal of Experimental Child Psychology*, *151*, 18–32. https://doi.org/10.1016/j.jecp.2016.03.012

Kucker, S. C., Samuelson, L. K., Perry, L. K., Yoshida, H., Colunga, E., Lorenz, M. G., & Smith, L. B. (2019). Reproducibility and a unifying explanation: Lessons from the shape bias. *Infant Behavior and Development*, *54*, 156–165. https://doi.org/10.1016/j.infbeh.2018.09.011

Kuhl, P. K. (2011). Early language learning and literacy: Neuroscience implications for education. *Mind, Brain, and Education*, *5*(3), 128–142. https://doi.org/10.1111/j.1751-228X.2011.01121.x

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the “critical period”. *Language Learning and Development*, *1*(3–4), 237–264. https://doi.org/10.1080/15475441.2005.9671948

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Landau, B., Smith, L., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, *38*(1), 1–27. https://doi.org/10.1006/jmla.1997.2533

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Leckey, S., Selmeczy, D., Kazemi, A., Johnson, E. G., Hembacher, E., & Ghetti, S. (2020). Response latencies and eye gaze provide insight on how toddlers gather evidence under uncertainty. *Nature Human Behaviour*, *4*(9), 928–936. https://doi.org/10.1038/s41562-020-0913-y

Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2

Lewis, M., M. (1936). *Infant speech; A study of the beginnings of language*. London, UK: Routledge & Kegan Paul.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, *87*(3), 151–178. https://doi.org/10.1016/S0010-0277(02)00230-5

Lucariello, J., Kyratzis, A., & Engel, S. (1986). Event representations, context, and language. In K. Nelson (Ed.), *Event knowledge: Structure and function in development* (pp. 137–160). Hillsdale, NJ: Lawrence Erlbaum.

Maguire, M. J., & Dove, G. O. (2008). Speaking of events: Event word learning and event representation. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 193–218). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195188370.003.0009

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364–391). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195170009.001.0001

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–634. https://doi.org/10.1111/j.1467-7687.2008.00707.x

Mandler, J. M., Fivush, R., & Reznick, J. S. (1987). The development of contextual categories. *Cognitive Development*, *2*(4), 339–354. https://doi.org/10.1016/S0885-2014(87)80012-6

Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*(5), 445–470. https://doi.org/10.1111/j.1532-7078.2009.00027.x

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language*, *38*(3), 606–627. https://doi.org/10.1017/S0305000910000243

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *The Minnesota symposia on child psychology, Vol. 25. Modularity and constraints in language and cognition* (pp. 59–101). Hillsdale, NJ: Lawrence Erlbaum.

Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, *92*, 199–227. https://doi.org/10.1016/0024-3841(94)90342-5

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27. https://doi.org/10.1016/0010-0285(84)90002-1

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, *9*(3), 302–315. https://doi.org/10.1111/1467-9507.00127

Mason, G. M., Goldstein, M. H., & Schwade, J. A. (2019). The role of multisensory development in early language learning. *Journal of Experimental Child Psychology*, *183*, 48–64. https://doi.org/10.1016/j.jecp.2018.12.011

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*(5), 922–934. https://doi.org/10.1037/0022-3514.81.5.922

Matatyaho, D. J., & Gogate, L. J. (2008). Type of maternal object motion during synchronous naming predicts preverbal infants' learning of word-object relations. *Infancy*, *13*(2), 172–184. https://doi.org/10.1080/15250000701795655

Matatyaho-Bullaro, D. J., Gogate, L., Mason, Z., Cadavid, S., & Abdel-Mottaleb, M. (2014). Type of object motion facilitates word mapping by preverbal infants. *Journal of Experimental Child Psychology*, *118*, 27–40. https://doi.org/10.1016/j.jecp.2013.09.010

Matsuo, A., Kita, S., Shinya, Y., Wood, G. C., & Naigles, L. (2012). Japanese two-year-olds use morphosyntax to learn novel verb meanings. *Journal of Child Language*, *39*(3), 637–663. https://doi.org/10.1017/S0305000911000213

McCune, L. (2008). *How children learn to learn language*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195177879.001.0001

McCune, L., Lennon, E. M., & Greenwood, A. (2020). Gestures, grunts, and words: Development in a dynamic system. *First Language*. https://doi.org/10.1177/0142723720966820

McCune, L., & Zlatev, J. (2015). Dynamic systems in semiotic development: The transition to reference. *Cognitive Development*, *36*, 161–170. https://doi.org/10.1016/j.cogdev.2015.09.010

McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, *14*(2), 181–189. https://doi.org/10.1111/j.1467-7687.2010.00968.x

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631. https://doi.org/10.1126/science.1144073

McShane, J. (1979). The development of naming. *Linguistics*, *17*(9–10), 879–905. https://doi.org/10.1515/ling.1979.17.9-10.879

Mervis, C. B., & Bertrand, J. (1995). Early lexical acquisition and the vocabulary spurt: A response to Goldfield & Reznick. *Journal of Child Language*, *22*(2), 461–468. https://doi.org/10.1017/S0305000900009880

Madole, K. L., Oakes, L. M., & Cohen, L. B. (1993). Developmental changes in infants' attention to function and form-function correlations. *Cognitive Development*, *8*(2), 189–209. https://doi.org/10.1016/0885-2014(93)90014-V

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 639–648. https://doi.org/10.1098/rstb.2006.2000

Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*, *55*(8), 1579–1593. https://doi.org/10.1037/dev0000738

Morales, M., Mundy, P., Delgado, C. E., Yale, M., Messinger, D., Neal, R., & Schwartz, H. K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology*, *21*(3), 283–298. https://doi.org/10.1016/S0193-3973(99)00040-4

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7), 309–315. https://doi.org/10.1016/S1364-6613(00)01682-X

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*(4), 413–429. https://doi.org/10.1111/1467-7687.00296

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274. https://doi.org/10.1111/j.1467-8721.2007.00518.x

Nakagawa, N. (1986). *Kotoba o hagukumu: Hattatsu ni okure no aru kodomotachi no tameni* [*Nurturing language: For children with developmental delay*]. Tokyo, Japan: Budousha.

Nakagawa, N. (2009). *Hattatsu shougai to kotoba no soudan: Kodomo no sodachi o sasaeru gengochoukakushi no apurochi* [*Developmental disorders and consultation on language: Speech therapists' approach to supporting children's development*]. Tokyo, Japan: Shogakukan.

Nakase, A., & Nishio, H. (2001). *Shimpan K-shiki hattatsu kensa hannou jitsurei shu* [*Kyoto scale of psychological development 2001: Responses of examples*]. Kyoto, Japan: Nakanishiya Shuppan.

Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy*, *2*(1), 73–86. https://doi.org/10.1207/S15327078IN0201_5

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, *69*(2), 295–308. https://doi.org/10.1111/j.1467-8624.1998.tb06189.x

Namy, L. L., & Waxman, S. R. (2002). Patterns of spontaneous production of novel words and gestures within an experimental setting in children ages 1;6 and 2;2. *Journal of Child Language*, *29*(4), 911–921. https://doi.org/10.1017/S0305000902005305

Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: Two modes of word acquisition?. *Developmental Science*, *6*(2), 136–142. https://doi.org/10.1111/1467-7687.00263

Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorization?. *Cognition*, *80*(3), B11–B20. https://doi.org/10.1016/S0010-0277(01)00112-3

Nelson, K. (1973a). Some evidence for the cognitive primacy of categorization and its functional basis. *Merrill-Palmer Quarterly of Behavior and Development*, *19*(1), 21–39. Retrieved from https://www.jstor.org/stable/23083791

Nelson, K. (1973b). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*(1/2, Serial No. 149). https://doi.org/10.2307/1165788

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, *81*(4), 267–285. https://doi.org/10.1037/h0036592

Nelson, K. (1983a). The conceptual basis for language. In Th. B. Seiler & W. Wannenmacher (Eds.), *Concept development and the development of word meaning* (pp. 173–188). Berlin Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-642-69000-6_11

Nelson, K. (1983b). The derivation of concepts and categories from event representations. In E. K. Scholnick (Ed.), *New trends in conceptual representation: Challenges to Piaget's theory?* (pp. 129–149). Hillsdale, NJ: Lawrence Erlbaum.

Nelson, K. (1986). Event knowledge and cognitive development. In K. Nelson (Ed.), *Event knowledge: Structure and function in development* (pp. 1–19). Hillsdale, NJ: Lawrence Erlbaum.

Nelson, K. (1995). The dual category problem in the acquisition of action words. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 223–249). Mahwah, NJ: Lawrence Erlbaum.

Nelson, K., Hampson, J., & Shaw, L. K. (1993). Nouns in early lexicons: Evidence, explanations and implications. *Journal of Child Language*, *20*(1), 61–84. https://doi.org/10.1017/S0305000900009120

Nelson, K., & Lucariello, J. (1985). The development of meaning in first words. In M. D. Barrett (Ed.), *Children's single-word speech* (pp. 59–86). New York, NY: John Wiley.

Nelson, K. E. (1977). Facilitating children's syntax acquisition. *Developmental Psychology*, *13*(2), 101–107. https://doi.org/10.1037/0012-1649.13.2.101

Newport, E. L. (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American sign language. *Language Sciences*, *10*(1), 147–172. https://doi.org/10.1016/0388-0001(88)90010-1

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1016/0364-0213(90)90024-Q

Noble, C. H., Rowland, C. F., & Pine, J. M. (2011). Comprehension of argument structure and semantic roles: Evidence from English-learning children and the forced-choice pointing paradigm. *Cognitive Science*, *35*(5), 963–982. https://doi.org/10.1111/j.1551-6709.2011.01175.x

Nomikou, I., Rohlfing, K. J., Cimiano, P., & Mandler, J. M. (2019). Evidence for early comprehension of action verbs. *Language Learning and Development*, *15*(1), 64–74. https://doi.org/10.1080/15475441.2018.1520639

Ogura, T. (2007) Early lexical development in Japanese children. *Gengo Kenkyu*, *132*, 29–53 (in Japanese). Retrieved from http://www.ls-japan.org/modules/documents/LSJpapers/journals/132_ogura.pdf

Ogura, T., & Watamaki, T. (2004). *Nihongo MacArthur nyuyoji gengo hattatsu shitumonshi* [*Japanese MacArthur communicative development inventories*]. Kyoto, Japan: Kyoto International Social Welfare Exchange Centre.

Ogura, T., Watamaki, T., & Inaba, T. (2016). *Nihongo MacArthur nyuyoji gengo hattatsu shitumonshi no kaihatsu to kenkyu* [*The development and research of Japanese MacArthur communicative development inventories*]. Kyoto, Japan: Nakanishiya Shuppan.

Okamoto, N. (1982). *Kodomo to kotoba* [*Children and language*]. Tokyo, Japan: Iwanami Shoten.

Okumura, Y., Kanakogi, Y., Kobayashi, T., & Itakura, S. (2020). Ostension affects infant learning more than attention. *Cognition*, *195*, 104082. https://doi.org/10.1016/j.cognition.2019.104082

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?. *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Oviatt, S. L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, *51*(1), 97–106. https://doi.org/10.2307/1129595

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition*, *96*(3), B101–B113. https://doi.org/10.1016/j.cognition.2005.01.001

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. In S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 3839–3845). New York, NY: AAAI Press / International Joint Conferences on Artificial Intelligence. Retrieved from https://www.ijcai.org/Proceedings/2016

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

Parladé, M. V., & Iverson, J. M. (2011). The interplay between language, gesture, and affect during communicative transition: A dynamic systems approach. *Developmental Psychology*, *47*(3), 820–833. https://doi.org/10.1037/a0021811

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Perone, S., & Oakes, L. M. (2006). It clicks when it is rolled and it squeaks when it is squeezed: What 10-month-old infants learn about object function. *Child Development*, *77*(6), 1608–1622. https://doi.org/10.1111/j.1467-8624.2006.00962.x

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, *2*, 345. https://doi.org/10.3389/fpsyg.2011.00345

Perry, L. K., Samuelson, L. K., & Burdinie, J. B. (2014). Highchair philosophers: The impact of seating context-dependent exploration on children's naming biases. *Developmental Science*, *17*(5), 757–765. https://doi.org/10.1111/desc.12147

Pickron, C. B., Iyer, A., Fava, E., & Scott, L. S. (2018). Learning to individuate: The specificity of labels differentially impacts infant visual attention. *Child Development*, *89*(3), 698–710. https://doi.org/10.1111/cdev.13004

Pruden, S. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2008). Current events: How infants parse the world and events for language. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 160–192). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195188370.003.0008

Puccini, D., & Liszkowski, U. (2012). 15-month-old infants fast map words but not representational gestures of multimodal labels. *Frontiers in Psychology*, *3*, 101. https://doi.org/10.3389/fpsyg.2012.00101

Quine, W. V. O. (1960). *Word & object*. Cambridge, MA: MIT Press.

Ratcliff, R., Love, J., Thompson, C. A., & Opfer, J. E. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, *83*(1), 367–381. https://doi.org/10.1111/j.1467-8624.2011.01683.x

Ratner, N., & Bruner, J. (1978). Games, social exchange and the acquisition of language. *Journal of Child Language*, *5*(3), 391–401. https://doi.org/10.1017/S0305000900002063

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy*, *1*(4), 491–507. https://doi.org/10.1207/S15327078IN0104_9

Ross, G., Nelson, K., Wetstone, H., & Tanouye, E. (1986). Acquisition and generalization of novel object concepts by young language learners. *Journal of Child Language*, *13*(1), 67–83. https://doi.org/10.1017/S0305000900000301

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Roy, D. (2009). New horizons in the study of child language acquisition. In M. Uther, R. Moore, & S. Cox (Eds.), *Proceedings of the 10th Annual Conference of the International Speech Communication Association* (pp. 13–20). Brighton, UK: International Speech Communication Association. Retrieved from: https://www.isca-speech.org/archive/interspeech_2009/i09_0013.html

Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology*, *80*(3), 201–224. https://doi.org/10.1006/jecp.2001.2633

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, *38*(6), 1016–1037. https://doi.org/10.1037/0012-1649.38.6.1016

Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition*, *110*(3), 322–345. https://doi.org/10.1016/j.cognition.2008.10.017

Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter, and Tomasello. *Child Development*, *69*(1), 94–104. https://doi.org/10.1111/j.1467-8624.1998.tb06136.x

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond?. *Cognition*, *73*(1), 1–33. https://doi.org/10.1016/S0010-0277(99)00034-7

Sandhofer, C., & Schonberg, C. (2020). Multiple examples support children's word learning: The roles of aggregation, decontextualization, and memory dynamics. In J. B. Childers (Ed.), *Language and concept acquisition from infancy through childhood: Learning from multiple examples* (pp. 159–178). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-35594-4_8

Santolin, C., Garcia, G.G., Zettersten, M., Sebastian-Galles, N., & Saffran, J. R. (2021), Experience with research paradigms relates to infants' direction of preference. *Infancy*, *26*(1), 39–46. https://doi.org/10.1111/infa.12372

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition and Development*, *8*(1), 113–128. https://doi.org/10.1080/15248370709336996

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. https://doi.org/10.1038/253265a0

Schafer, G., & Plunkett, K. (1998). Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development*, *69*(2), 309–320. https://doi.org/10.1111/j.1467-8624.1998.tb06190.x

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. *Developmental Science*, *12*(4), 583–601. https://doi.org/10.1111/j.1467-7687.2009.00809.x

Scott, K., Chu, J., & Schulz, L. (2017). Lookit (Part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind: Discoveries in Cognitive Science*, *1*(1), 15–29. https://doi.org/10.1162/OPMI_a_00001

Scott, K., & Schulz, L. (2017). Lookit (Part 1): A new online platform for developmental research. *Open Mind: Discoveries in Cognitive Science*, *1*(1), 4–14. https://doi.org/10.1162/OPMI_a_00002

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. https://doi.org/10.1016/j.cognition.2011.10.010

Semmelmann, K., Hönekopp, A., & Weigelt, S. (2017). Looking tasks online: utilizing webcams to collect video data from home. *Frontiers in Psychology*, *8*, 1582. https://doi.org/10.3389/fpsyg.2017.01582

Sheehan, E. A., Namy, L. L., & Mills, D. L. (2007). Developmental changes in neural activity to familiar words and gestures. *Brain and Language*, *101*(3), 246–259. https://doi.org/10.1016/j.bandl.2006.11.008

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173–189. https://doi.org/10.1016/j.jml.2004.04.004

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), e12816. https://doi.org/10.1111/desc.12816

Sloutsky, V. M., & Robinson, C. W. (2013). Redundancy matters: Flexible learning of multiple contingencies in infants. *Cognition*, *126*(2), 156–164. https://doi.org/10.1016/j.cognition.2012.09.016

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism?. *Cognition*, *60*(2), 143–171. https://doi.org/10.1016/0010-0277(96)00709-3

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*(1), 13–19. https://doi.org/10.1111/1467-9280.00403

Smith, L. B., & Samuelson, L. K. (2003). Different is good: Connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science*, *6*(4), 434–439. https://doi.org/10.1111/1467-7687.00298

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Song, L., Pruden, S. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Prelinguistic foundations of verb learning: Infants discriminate and categorize dynamic human actions. *Journal of Experimental Child Psychology*, *151*, 77–95. https://doi.org/10.1016/j.jecp.2016.01.004

Sonne, T., Kingo, O. S., & Krøjgaard, P. (2016). Occlusions at event boundaries during encoding have a negative effect on infant memory. *Consciousness and Cognition*, *41*, 72–82. https://doi.org/10.1016/j.concog.2016.02.006

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Spencer, J. P., Perone, S., & Buss, A. T. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child Development Perspectives*, *5*(4), 260–266. https://doi.org/10.1111/j.1750-8606.2011.00194.x

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, *6*(4), 392–412. https://doi.org/10.1111/1467-7687.00295

Stan Development Team (2019). RStan: The R interface to Stan. R package version 2.19.3. Retrieved from http://mc-stan.org

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*(4620), 179–181. https://doi.org/10.1126/science.6623069

Stern, C., & Stern, W. (1907). *Die Kindersprache: Eine psychologische und sprachtheoretische Untersuchung* [*Children's speech: A psychological and language-theoretical investigation*]. Leipzig, Germany: Barth.

Stevenson, A. (1893). The speech of children. *Science*, *21*(526), 118–120. https://doi.org/10.1126/science.ns-21.526.118-a

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The signal in the noise: The visual ecology of parents' object naming. *Infancy*, *24*(3), 455–476. https://doi.org/10.1111/infa.12278

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Suanda, S. H., Walton, K. M., Broesch, T., Kolkin, L., & Namy, L. L. (2013). Why two-year-olds fail to learn gestures as object labels: Evidence from looking time and forced-choice measures. *Language Learning and Development*, *9*(1), 50–65. https://doi.org/10.1080/15475441.2012.723189

Swallow, K. M., & Wang, Q. (2020). Culture influences how people divide continuous sensory experience into events. *Cognition*, *205*, 104450. https://doi.org/10.1016/j.cognition.2020.104450

Swets J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285–1293. https://doi.org/10.1126/science.3287615

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0

Tafreshi, D., Thompson, J. J., & Racine, T. P. (2014). An analysis of the conceptual foundations of the infant preferential looking paradigm. *Human Development*, *57*(4), 222–240. https://doi.org/10.1159/000363487

Tamis-LeMonda, C. S., Bornstein, M. H., Cyphers, L., Toda, S., & Ogino, M. (1992). Language and play at one year: A comparison of toddlers and mothers in the United States and Japan. *International Journal of Behavioral Development*, *15*(1), 19–42. https://doi.org/10.1177/016502549201500102

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152. https://doi.org/10.1111/cdev.13089

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology*, *32*(3), 492–504. https://doi.org/10.1037/0012-1649.32.3.492

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929–938. https://doi.org/10.1037/0012-1649.44.4.929

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of English and Mandarin. *Child Development*, *70*(3), 620–635. https://doi.org/10.1111/1467-8624.00045

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*(3), 535–565. https://doi.org/10.1017/S030500099700319X

Taylor, G., Liu, H., & Herbert, J. S. (2016). The role of verbal labels on flexible memory retrieval at 12-months of age. *Infant Behavior and Development*, *45*(Part A), 11–17. https://doi.org/10.1016/j.infbeh.2016.08.002

Tek, S., & Naigles, L. (2017). The shape bias as a word-learning principle: Lessons from and for autism spectrum disorder. *Translational Issues in Psychological Science*, *3*(1), 94–103. https://doi.org/10.1037/tps0000104

Thelen, E., & Bates, E. (2003). Connectionism and dynamic systems: Are they really different?. *Developmental Science*, *6*(4), 378–391. https://doi.org/10.1111/1467-7687.00294

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*(4), 432–444. https://doi.org/10.1111/j.1532-7078.2011.00084.x

Tippenhauer, N., & Saylor, M. M. (2019). Effects of context variability on 2-year-olds' fact and word learning. *Journal of Experimental Child Psychology*, *179*, 248–259. https://doi.org/10.1016/j.jecp.2018.11.011

Tomasello, M. (2002). Things are what they do: Katherine Nelson's functional approach to language and cognition. *Journal of Cognition and Development*, *3*(1), 5–19. https://doi.org/10.1207/S15327647JCD0301_2

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2014). *A natural history of human thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*(2), 201–224. https://doi.org/10.1016/0885-2014(95)90009-8

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Tomasello, M., Mannle, S., & Kruger, A. C. (1986). Linguistic environment of 1- to 2-year-old twins. *Developmental Psychology*, *22*(2), 169–176. https://doi.org/10.1037/0012-1649.22.2.169

Torkildsen, J. V. K., Svangstu, J. M., Hansen, H. F., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience*, *20*(7), 1266–1282. https://doi.org/10.1162/jocn.2008.20087

Tovar, Á. E., Rodríguez-Granados, A., & Arias-Trejo, N. (2020). Atypical shape bias and categorization in autism: Evidence from children and computational simulations. *Developmental Science*, *23*(2), e12885. https://doi.org/10.1111/desc.12885

Tran, M., Cabral, L., Patel, R., & Cusack, R. (2017). Online recruitment and testing of infants with Mechanical Turk. *Journal of Experimental Child Psychology*, *156*, 168–178. https://doi.org/10.1016/j.jecp.2016.12.003

Träuble, B., & Pauen, S. (2007). The role of functional information for infant categorization. *Cognition*, *105*(2), 362–379. https://doi.org/10.1016/j.cognition.2006.10.003

Träuble, B., & Pauen, S. (2011). Cause or effect: What matters? How 12-month-old infants learn to categorize artifacts. *British Journal of Developmental Psychology*, *29*(3), 357–374. https://doi.org/10.1348/026151009X479547

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, *15*(4), 357–368. https://doi.org/10.1002/dev.420150408

Twomey, K. E., & Hilton, M. (2020). Word learning. In S. Hupp & J. D. Jewell (Eds.), *The encyclopedia of child and adolescent development, Vol. 3: Cognition in childhood*. Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9781119171492.wecad126

Twomey, K. E., Ma, L., & Westermann, G. (2018). All the right noises: Background variability helps early word learning. *Cognitive Science*, *42*(52), 413–438. https://doi.org/10.1111/cogs.12539

Twomey, K. E., & Westermann, G. (2018). Learned labels shape pre-speech Infants' object representations. *Infancy*, *23*(1), 61–73. https://doi.org/10.1111/infa.12201

Tomasello, M., & Rakoczy, H. (2003). What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, *18*(2), 121–147. https://doi.org/10.1111/1468-0017.00217

Ueda, Y., Chen, L., Kopecky, J., Cramer, E. S., Rensink, R. A., Meyer, D. E., ... & Saiki, J. (2018). Cultural differences in visual search for geometric figures. *Cognitive Science*, *42*(1), 286–310. https://doi.org/10.1111/cogs.12490

Uriu, Y. (1986). The very young children's comprehension of the unusual agent-action instructions. *Japanese Journal of Educational Psychology*, *34*(4), 306–314 (in Japanese). https://doi.org/10.5926/jjep1953.34.4_306

Uriu, Y. (1992). Young children's comprehension of unusual "object-action" sentences. *Japanese Journal of Educational Psychology*, *40*(4), 411–419 (in Japanese). https://doi.org/10.5926/jjep1953.40.4_411

Valleau, M. J., Konishi, H., Golinkoff, R. M., Hirsh-Pasek, K., & Arunachalam, S. (2018). An eye-tracking study of receptive verb knowledge in toddlers. *Journal of Speech, Language, and Hearing Research*, *61*(12), 2917–2933. https://doi.org/10.1044/2018_JSLHR-L-17-0363

van Schaik, J. E., Meyer, M., van Ham, C. R., & Hunnius, S. (2020). Motion tracking of parents' infant- versus adult-directed actions reveals general and action-specific modulations. *Developmental Science*, *23*(1), e12869. https://doi.org/10.1111/desc.12869

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, *108*(2), 394–401. https://doi.org/10.1016/j.jecp.2010.09.011

Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences*, *115*(44), 11333–11338. https://doi.org/10.1073/pnas.1800901115

Volterra, V., Bates, E., Benigni, L., Bretherton, I., & Camaioni, L. (1979). First words in language and action: A qualitative look. In E. Bates, L. Benigni, I. Bretherton, L. Camaioni & V. Volterra (Eds.), *The emergence of symbols: cognition and communication in infancy* (pp. 141–222). New York, NY: Academic Press.

Vygotsky, L. S. (1986). *Thought and Language* (A. Kozulin, Ed., Trans.). Cambridge, MA: MIT Press. (Original work published 1934)

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. https://doi.org/10.1016/j.cognition.2010.11.014

Ware, E. A., & Booth, A. E. (2010). Form follows function: Learning about function helps children learn about shape. *Cognitive Development*, *25*(2), 124–137. https://doi.org/10.1016/j.cogdev.2009.10.003

Watanabe, S. (2010a). Equations of states in singular statistical estimation. *Neural Networks*, *23*(1), 20–34. https://doi.org/10.1016/j.neunet.2009.08.002

Watanabe, S. (2010b). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*(116), 3571–3594. Retrieved from http://jmlr.org/papers/v11/watanabe10a.html

Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H. J. (2013). Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives*, *7*(3), 155–159. https://doi.org/10.1111/cdep.12032

Waxman, S. R., Fu, X., Ferguson, B., Geraghty, K., Leddon, E., Liang, J., & Zhao, M. F. (2016). How early is infants' attention to objects and actions shaped by culture? New evidence from 24-month-olds raised in the US and China. *Frontiers in Psychology*, *7*, 97. https://doi.org/10.3389/fpsyg.2016.00097

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child Development*, *85*(2), 429–436. https://doi.org/10.1111/cdev.12149

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology*, *34*(6), 1289–1309. https://doi.org/10.1037/0012-1649.34.6.1289

Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought*. New York, NY: John Wiley.

West, K. L., & Iverson, J. M. (2017). Language learning is hands-on: Exploring links between infants' object manipulation and verbal input. *Cognitive Development*, *43*, 190–200. https://doi.org/10.1016/j.cogdev.2017.05.004

Whorf, B. L. (1956). Science and linguistics. In J. B. Carroll (Ed.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (pp. 207–219). Cambridge, MA: MIT Press. (Reprinted from *MIT Technology Review*, *42*, 229–231, 247–248, 1940).

Wildt, E., Rohlfing, K. J., & Scharlau, I. (2019). The role of saliency in learning first words. *Frontiers in Psychology*, *10*, 1150. https://doi.org/10.3389/fpsyg.2019.01150

Winters, S., Dubuc, C., & Higham, J. P. (2015). Perspectives: the looking time experimental paradigm in studies of animal visual perception and cognition. *Ethology*, *121*(7), 625–640. https://doi.org/10.1111/eth.12378

Wojcik, E. H. (2017). 2.5-year-olds' retention and generalization of novel words across short and long delays. *Language Learning and Development*, *13*(3), 300–316. https://doi.org/10.1080/15475441.2016.1275644

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*(4), 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250. https://doi.org/10.1016/S0010-0277(02)00109-9

Yamamoto, H. (2020). An ecological approach to infants' visual experiences. *Japanese Psychological Review*, *63*(1), 102–117 (in Japanese).

Yee, M. N., Jones, S. S., & Smith, L. B. (2012). Changes in visual object recognition precede the shape bias in early noun learning. *Frontiers in Psychology*, *3*, 533. https://doi.org/10.3389/fpsyg.2012.00533

Yoshida, H. (2006). Goigakushu no mekanizumu: Gakushu shisutemu, seiyaku, soukankankei-gakushu no kasetsu: Haryu rombun eno comento [Mechanisms of word learning: Learning systems, constraints, and correlational account: Commentary to Haryu's article]. *Japanese Psychological Review*, *49*(1), 91–95. https://doi.org/10.24602/sjpr.49.1_91

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13–15), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yuan, S., & Fisher, C. (2009). "Really? She blicked the baby?" Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399. https://doi.org/10.1111/j.1467-8624.2012.01783.x

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891–921. https://doi.org/10.1111/cogs.12035

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349. https://doi.org/10.1111/desc.12349

Zmyj, N., Daum, M. M., Prinz, W., & Aschersleben, G. (2012). Infants' attentional preference for object-related actions of older children compared to object-related actions of peers and adults. *Infant Behavior and Development*, *35*(3), 533–542. https://doi.org/10.1016/j.infbeh.2012.03.003

Appendices

Chapter 2

Appendix A   Two-way ANOVA for different age and vocabulary size groups

Appendix B   Reaction time analysis

Chapter 3

Appendix C   Examples of face and gaze angle trajectories

Chapter 4

Appendix D   Equivalence confirmation of looking and pointing measures for Study 1

Appendix E   Model comparison details

Appendix F   Correlation analysis details

Appendix G   Developmental changes in the proportion of absent pointing responses

Appendix A Two-way ANOVA for different age and vocabulary size groups

A two-way analysis of variance (ANOVA) for different age groups, although conducted as a preliminary analysis, revealed a significant main effect of both age ($F(1,34) = 19.21,\ p < 0.001,\ \eta_G^2 = 0.28$) and condition ($F(1,34) = 11.46,\ p = 0.018,\ \eta_G^2 = 0.09$), as well as a significant interaction ($F(1,34) = 5.37,\ p = 0.27,\ \eta_G^2 = 0.04$; **Figure A.1**). Post-hoc analysis using a modified sequentially rejective Bonferroni procedure showed a significant simple effect of age group in both the match ($F(1,34) = 9.88,\ p = 0.0035,\ \eta_G^2 = 0.23$) and mismatch ($F(1,34) = 17.33,\ p < 0.001,\ \eta_G^2 = 0.34$) conditions. However, no significant simple effects of either condition for those aged less than 21 months ($F(1,3) = 6.00,\ p = 0.092,\ \eta_G^2 = 0.20$) or 21-months and older ($F(1,31) = 2.71,\ p = 0.11,\ \eta_G^2 = 0.03$) was found. A two-way ANOVA for vocabulary-size groups revealed a significant main effect of both vocabulary size ($F(1,34) = 17.75,\ p < 0.01,\ \eta_G^2 = 0.28$) and condition ($F(1,34) = 20.48,\ p < 0.001,\ \eta_G^2 = 0.13$), as well as a significant interaction ($F(1,34) = 15.38,\ p < 0.001,\ \eta_G^2 = 0.10$). Post-hoc analysis revealed a significant simple effect of vocabulary size group for both the match ($F(1,34) = 4.15,\ p = 0.049,\ \eta_G^2 = 0.11$) and mismatch ($F(1,34) = 24.34,\ p < 0.001,\ \eta_G^2 = 0.42$) conditions. A significant simple effect of conditions for the group with a vocabulary size of less than 140 words ($F(1,7) = 28.00,\ p = 0.0011,\ \eta_G^2 = 0.25$) was also found; however, no significant effect for the group with a vocabulary of at least 140 words ($F(1,27) = 0.39,\ p = 0.54,\ \eta_G^2 = 0.01$) was found.

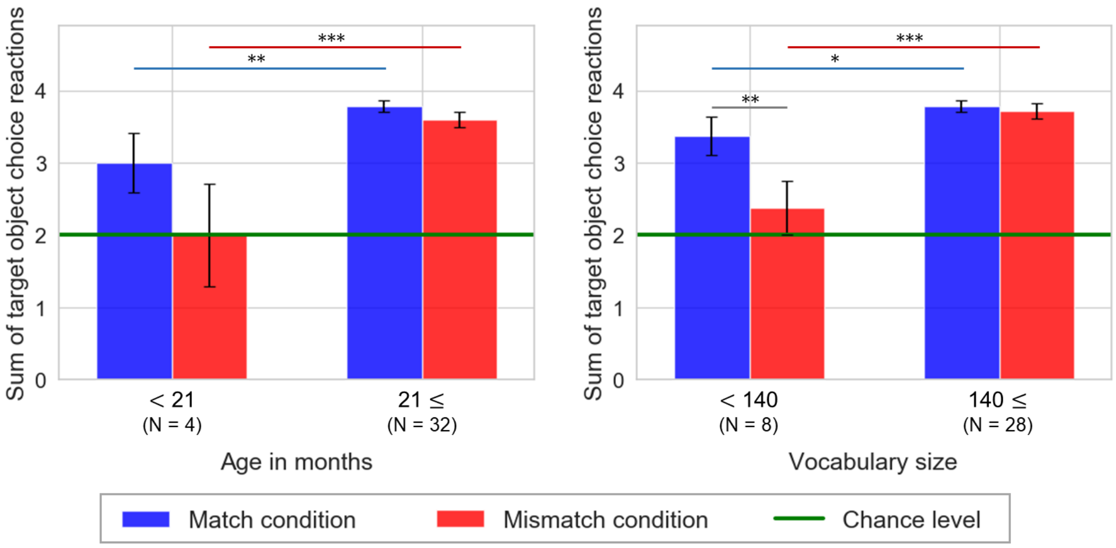

Figure A.1 Mean proportion of target-object choice reactions for each group before and after the developmental boundary. Note that three out of four children in the younger age group were included in the lower vocabulary size group. Error bars indicate one standard error. * p < 0.05, ** p < 0.01, *** p < 0.001.

Appendix B Reaction time analysis

A supplemental analysis was conducted of the time elapsed between when each experimental question was completed and when the participant made a pointing reaction during each trial. Although this study did not adopt a methodology that could strictly deal with reaction times, reaction times may reflect the degree of the cognitive load or answer for confidence during the task, even if this index was subject to noise from various factors, ranging from the perception of the stimuli to the execution of pointing movements. In fact, the reaction times in one case were occasionally delayed, as the participant appeared to ponder over the choice. Two raters independently measured reaction times in all trials; substantial interrater reliability was confirmed (ICC(2,1) = 0.72). The average reaction time between the raters was used for analysis.

The correlation analysis for each condition using Pearson's correlation coefficients showed no significant correlation between reaction times and age in months ($r = -0.13, p = 0.12$ for the match condition; $r = -0.06, p = 0.50$ for the mismatch condition) or vocabulary size ($r = -0.14, p = 0.10$ for the match condition; $r = -0.13, p = 0.13$, for the mismatch condition; **Figure B.1**). These results indicate that for both conditions, the reaction time in this task was less susceptible to overall developmental changes that showed dramatic improvement during toddler age (e.g., motor control).

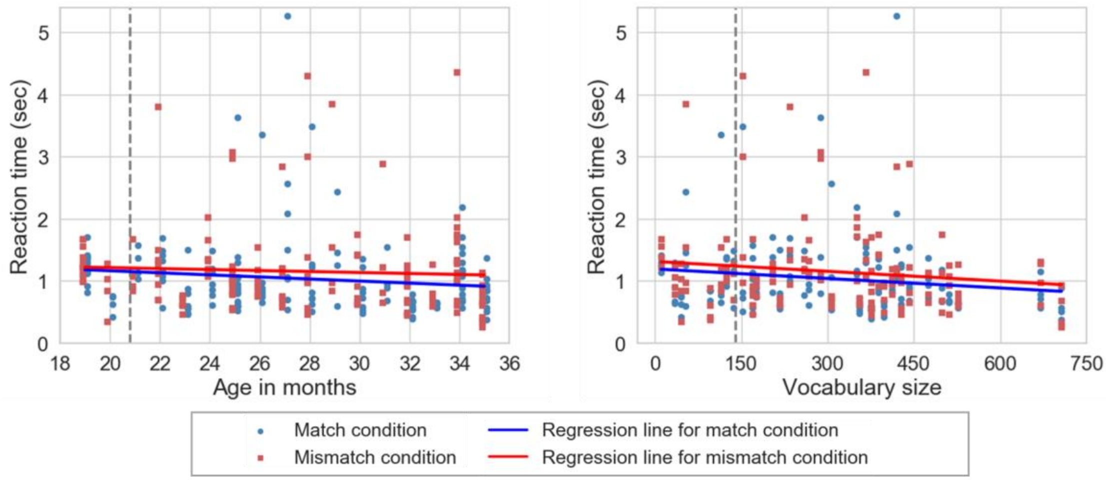

Figure B.1 The relation between reaction time and developmental indices (age in months or vocabulary size) for each condition. The vertical dashed line represents the developmental boundary, predicted as the semantic differentiation of noun-like words.

Assuming that each pointing reaction was independent during each trial, the mean reaction times were calculated for each condition and developmental group (**Figure B.2**). A two-way ANOVA of reaction times revealed no significant main effect of age group ($F(1, 284) = 0.13,\ p = 0.71,\ \eta_G^2 \approx 0.00$) or condition ($F(1, 284) = 0.67,\ p = 0.41,\ \eta_G^2 \approx 0.00$), and no significant interaction ($F(1, 284) = 0.011,\ p = 0.92,\ \eta_G^2 \approx 0.00$). A similar two-way ANOVA revealed no significant main effect of vocabulary size ($F(1,284) = 0.049,\ p = 0.83,\ \eta_G^2 \approx 0.00$) or condition ($F(1,284) = 0.79$, $p = 0.37,\ \eta_G^2 \approx 0.00$), and no significant interaction ($F(1,284) = 0.29,\ p = 0.59,\ \eta_G^2 \approx 0.00$). Thus far, these results indicate that cognitive processing load or answer confidence during tasks barely differed between conditions among all groups in terms of reaction time, although this evidence was indirect and weak due to the coverage of noise from various other factors.

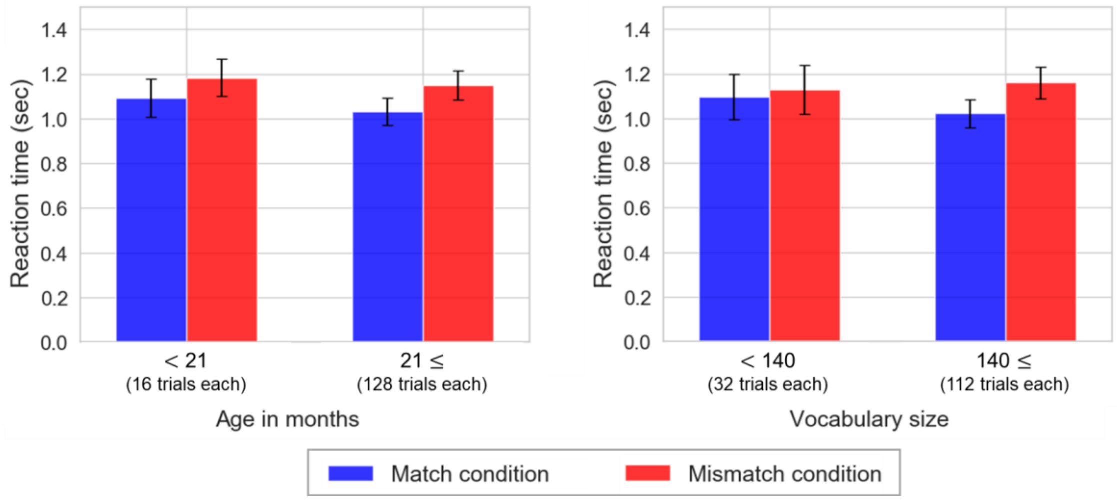

Figure B.2 Mean reaction time for each condition before and after the developmental boundary. Error bars indicate one standard error.

Appendix C Examples of face and gaze angle trajectories

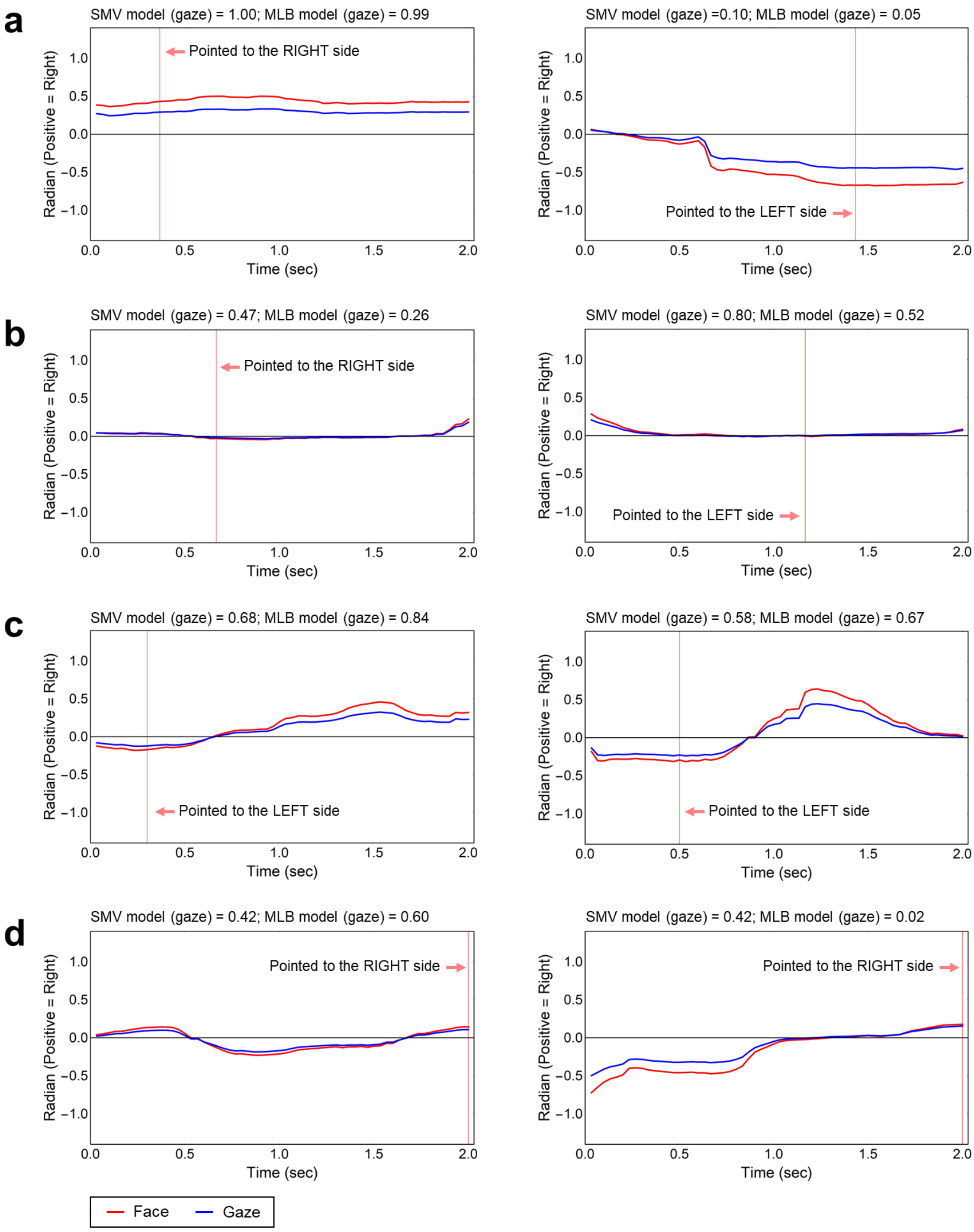

Figure C.1 Examples of face and gaze angle trajectories. (a) Examples of trials in which the predictions were correct; (b–d) Examples of trials in which the predictions were incorrect because: (b) face and gaze direction mostly remained around the center of the monitor; (c) pointing responses occurred within the first second and then a participant looked at the other stimulus for a while; and (d) looking was at the opposite side of a stimulus to the one chosen until just before pointing. Note that the vertical pink line indicates the timing of when participants produced pointing responses (but if the timing was over 2 seconds, the line remained at the location of 2 seconds). Prediction scores for the SMV and MLB models using gaze direction trajectories in each trial are also shown. Scores ranged from 0 to 1, where 0 indicates the left and 1 the right pointing prediction.

Appendix D Equivalence confirmation of looking and pointing measures for Study 1

Table D.1 Correlations of the proportion of correct responses calculated from looking and pointing measures in Study 1.

Condition	*n*	*ρ*	*p*
For participants who showed clear pointing responses in at least one trial for each condition			
match	50	0.67	< 0.0001
mismatch	51	0.72	< 0.0001
different-object	51	0.49	0.0003
absent-object	48	0.58	< 0.0001
For participants who showed clear pointing responses in all four trials for each condition			
match	40	0.69	< 0.0001
mismatch	36	0.83	< 0.0001
different-object	36	0.78	< 0.0001
absent-object	31	0.75	< 0.0001

Note. Spearman's rank correlation coefficient was used.

Appendix E Model comparison details

Table E.1 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates, using age in months as a developmental index for Analysis 1 in Study 1.

Table E.2 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using total vocabulary size as a developmental index for Analysis 1 in Study 1.

Table E.3 Widely applicable information criterion (WAIC), posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using age in months as a developmental index for Analysis 2 in Study 1.

Table E.4 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using total vocabulary size as a developmental index for Analysis 2 in Study 1.

Table E.5 Widely applicable information criterion (WAIC) posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates for Analyses 1 and 2 in Study 2.

Table E.6 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for the best five models for Analysis 3 in Study 2.

Table E.1 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates, using age in months as a developmental index for Analysis 1 in Study 1.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			age	condition	age × condition	intercept	σ
The preferential looking model with the match and the mismatch conditions (n = 69)							
1	age, condition	**369.94** (13.03)	0.181 [0.059, 0.308]	-0.634 [-1.016, -0.258]	–	-2.472 [-5.049, 0.022]	0.402 [0.055, 0.753]
2	age, condition, age × condition	372.23 (13.10)	0.178 [0.004, 0.358]	-0.790 [-5.279, 3.703]	0.008 [-0.215, 0.228]	-2.403 [-6.043, 1.147]	0.411 [0.069, 0.761]
3	condition	375.13 (12.45)	–	-0.634 [-1.014, -0.256]	–	1.224 [0.922, 1.556]	0.497 [0.109, 0.833]
4	age	379.81 (13.40)	0.177 [0.057, 0.302]	–	–	-2.732 [-5.254, -0.264]	0.371 [0.048, 0.718]
5	(intercept only)	385.35 (13.05)	–	–	–	0.884 [0.670, 1.121]	0.457 [0.096, 0.798]
The pointing model with the match and the mismatch conditions (n = 51)							
1	age, condition	**229.50** (12.76)	0.152 [-0.030, 0.343]	-1.014 [-1.564, -0.490]	–	-1.316 [-5.245, 2.522]	0.568 [0.087, 1.041]
2	condition	229.95 (12.73)	–	-1.001 [-1.541, -0.484]	–	1.836 [1.399, 2.337]	0.566 [0.136, 1.040]
3	age, condition, age × condition	231.97 (12.93)	0.186 [-0.080, 0.462]	0.092 [-6.510, 6.613]	-0.054 [-0.369, 0.263]	-2.021 [-7.675, 3.547]	0.562 [0.092, 1.056]
4	age	244.40 (12.15)	0.144 [-0.027, 0.326]	–	–	-1.733 [-5.481, 1.873]	0.499 [0.036, 0.985]
5	(intercept only)	244.52 (12.02)	–	–	–	1.270 [0.979, 1.609]	0.515 [0.109, 0.978]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of "condition" indicates the effect of the mismatch condition compared to the match condition. Parameter σ represents the standard deviation of the random intercept.

Table E.2 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using total vocabulary size as a developmental index for Analysis 1 in Study 1.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			vocabulary size	condition	vocabulary size × condition	intercept	σ
The preferential looking model with the match and the mismatch conditions (n = 69)							
1	vocabulary size, condition	**369.32** (13.41)	0.004 [0.002, 0.007]	-0.632 [-1.018, -0.244]	–	0.797 [0.410, 1.211]	0.384 [0.063, 0.734]
2	vocabulary size, condition, vocabulary size × condition	371.72 (13.65)	0.005 [0.001, 0.010]	-0.550 [-1.150, 0.047]	-0.001 [-0.006, 0.004]	0.745 [0.277, 1.232]	0.373 [0.056, 0.736]
3	vocabulary size	378.94 (13.93)	0.004 [0.002, 0.007]	–	–	0.468 [0.145, 0.799]	0.347 [0.062, 0.701]
The pointing model with the match and the mismatch conditions (n = 51)							
1	vocabulary size, condition	**224.11** (13.28)	0.006 [0.002, 0.010]	-1.009 [-1.559, -0.471]	–	1.188 [0.606, 1.795]	0.411 [0.056, 0.898]
2	vocabulary size, condition, vocabulary size × condition	225.37 (13.51)	0.009 [0.002, 0.017]	-0.656 [-1.551, 0.226]	-0.004 [-0.013, 0.004]	0.942 [0.204, 1.699]	0.417 [0.055, 0.901]
3	vocabulary size	237.61 (13.09)	0.006 [0.002, 0.010]	–	–	0.638 [0.172, 1.124]	0.368 [0.059, 0.841]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of “condition” indicates the effect of the mismatch condition compared to the match condition. Model candidates that did not include vocabulary size in the fixed effects are the same as in Table E1. Parameter σ represents the standard deviation of the random intercept.

Table E.3 Widely applicable information criterion (WAIC), posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using age in months as a developmental index for Analysis 2 in Study 1.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			age	condition	age × condition	intercept	σ
The preferential looking model with the match and the mismatch conditions (n = 69)							
1	(intercept only)	**376.18** (11.75)	–	–	–	-0.022 [-0.195, 0.152]	0.141 [0.019, 0.412]
2	condition	376.48 (11.43)	–	0.222 [-0.112, 0.568]	–	-0.133 [-0.376, 0.108]	0.149 [0.023, 0.416]
3	age	376.61 (11.90)	0.060 [-0.039, 0.160]	–	–	-1.254 [-3.304, 0.789]	0.141 [0.013, 0.402]
4	age, condition	376.89 (11.56)	0.060 [-0.040, 0.159]	0.223 [-0.116, 0.558]	–	-1.361 [-3.406, 0.694]	0.140 [0.012, 0.407]
5	age, condition, age × condition	378.58 (11.58)	0.087 [-0.051, 0.226]	1.356 [-2.604, 5.309]	-0.055 [-0.248, 0.138]	-1.916 [-4.779, 0.933]	0.134 [0.011, 0.416]
The pointing model with the match and the mismatch conditions (n = 51)							
1	(intercept only)	**258.12** (9.30)	–	–	–	0.252 [0.016, 0.495]	0.281 [0.044, 0.663]
2	age	259.46 (9.22)	0.051 [-0.090, 0.188]	–	–	-0.815 [-3.682, 2.142]	0.284 [0.021, 0.681]
3	condition	260.16 (9.33)	–	0.079 [-0.358, 0.521]	–	0.216 [-0.098, 0.536]	0.291 [0.051, 0.675]
4	age, condition	261.53 (9.31)	0.051 [-0.090, 0.190]	0.076 [-0.360, 0.511]	–	-0.847 [-3.754, 2.102]	0.280 [0.027, 0.682]
5	age, condition, age × condition	263.55 (9.60)	0.025 [-0.162, 0.211]	-1.037 [-6.440, 4.282]	0.053 [-0.200, 0.311]	-0.304 [-4.191, 3.604]	0.289 [0.025, 0.690]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of "condition" indicates the effect of the absent-object condition compared to the different-object condition. Parameter σ represents the standard deviation of the random intercept.

Table E.4 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using total vocabulary size as a developmental index for Analysis 2 in Study 1.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			vocabulary size	condition	vocabulary size × condition	intercept	σ
The preferential looking model with the match and the mismatch conditions (n = 69)							
1	vocabulary size	378.19 (11.77)	0.000 [-0.002, 0.002]	–	–	-0.035 [-0.305, 0.239]	0.156 [0.023, 0.439]
2	vocabulary size, condition	378.42 (11.45)	0.000 [-0.002, 0.002]	0.220 [-0.112, 0.560]	–	-0.146 [-0.470, 0.176]	0.162 [0.025, 0.431]
3	vocabulary size, condition, vocabulary size × condition	380.49 (11.55)	0.000 [-0.003, 0.003]	0.255 [-0.274, 0.796]	0.000 [-0.005, 0.004]	-0.165 [-0.554, 0.220]	0.160 [0.018, 0.431]
The pointing model with the match and the mismatch conditions (n = 51)							
1	vocabulary size	260.01 (9.48)	0.000 [-0.003, 0.002]	–	–	0.304 [-0.086, 0.701]	0.310 [0.045, 0.705]
2	vocabulary size, condition	262.14 (9.54)	0.000 [-0.003, 0.002]	0.080 [-0.357, 0.521]	–	0.266 [-0.175, 0.716]	0.316 [0.046, 0.711]
3	vocabulary size, condition, vocabulary size × condition	263.16 (9.26)	0.001 [-0.003, 0.004]	0.352 [-0.351, 1.068]	-0.002 [-0.007, 0.002]	0.135 [-0.379, 0.658]	0.317 [0.050, 0.706]

Note. The fixed effect of "condition" indicates the effect of the absent-object condition compared to the different-object condition. Model candidates that did not include vocabulary size in the fixed effects are the same as in Table E.3. There were no best models among model candidates that included vocabulary size. Parameter σ represents the standard deviation of the random intercept.

Table E.5 Widely applicable information criterion (WAIC) posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates for Analyses 1 and 2 in Study 2.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]						
			time	condition	time × condition	intercept	σ_1	σ_2	σ_3
Analysis 1: The preferential looking model with the match and the mismatch conditions (n = 16)									
1	time, condition	**172.43** (10.20)	0.775 [0.090, 1.508]	-0.309 [-1.110, 0.474]	–	0.075 [-1.122, 1.276]	0.405 [0.028, 1.090]	0.396 [0.032, 1.023]	0.629 [0.090, 1.280]
2	time, condition, time × condition	174.38 (9.91)	0.574 [-0.341, 0.534]	-0.863 [-2.783, 1.017]	0.378 [-0.813, 1.597]	0.366 [-1.112, 0.883]	0.433 [0.036, 1.113]	0.377 [0.036, 1.018]	0.634 [0.074, 1.277]
3	time	177.57 (12.07)	0.709 [0.137, 1.292]	–	–	-0.068 [-1.005, 0.880]	0.652 [0.211, 1.204]	–	–
4	(intercept only)	181.83 (12.26)	–	–	–	0.958 [0.541, 1.444]	0.630 [0.203, 1.167]	–	–
5	condition	183.22 (12.10)	–	-0.287 [-0.847, 0.271]	–	1.110 [0.605, 1.686]	0.635 [0.196, 1.174]	–	–
Analysis 2: The preferential looking model with the different-object and the absent-object conditions (n = 16)									
1	(intercept only)	**173.39** (7.10)	–	–	–	-0.098 [-0.397, 0.198]	0.268 [0.027, 0.676]	–	–
2	condition	174.29 (6.85)	–	0.261 [-0.242, 0.759]	–	-0.228 [-0.626, 0.160]	0.273 [0.024, 0.687]	–	–
3	time	175.30 (7.31)	0.064 [-0.435, 0.559]	–	–	-0.192 [-0.992, 0.615]	0.269 [0.026, 0.682]	–	–
4	time, condition	176.74 (6.28)	0.067 [-0.461, 0.598]	0.265 [-0.281, 0.818]	–	-0.334 [-1.244, 0.564]	0.231 [0.015, 0.666]	0.159 [0.009, 0.517]	0.224 [0.015, 0.657]
5	time, condition, time × condition	177.58 (6.33)	0.338 [-0.414, 1.084]	1.069 [-0.544, 2.716]	-0.534 [-1.556, 0.482]	-0.739 [-1.952, 0.458]	0.229 [0.012, 0.668]	0.163 [0.012, 0.529]	0.240 [0.017, 0.670]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of “condition” indicates the effect of the mismatch condition compared to the match condition in Analysis 1 and the absent-object condition compared to the different-object condition in Analysis 2. Parameters σ_1, σ_2, and σ_3 represent the standard deviation of the random intercepts that correspond to participants, the combination of participants and times, and the combination of participants and conditions, respectively.

Table E.6 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for the best five models for Analysis 3 in Study 2.

Model rank	The number of conditions included	WAIC (SE)	Parameter estimates: posterior median [95% CI]						
			match	mismatch	different-object	absent-object	age	intercept	σ
Common noun growth prediction									
1	4	**94.65** (8.51)	-0.067 [-3.541, 3.235]	1.007 [-2.339, 4.430]	-1.076 [-5.020, 2.833]	0.665 [-2.514, 4.048]	1.901 [-0.504, 4.705]	0.733 [-3.101, 4.376]	1.829 [1.092, 3.297]
2	3	95.03 (8.41)	-0.062 [-3.534, 3.199]	0.738 [-2.355, 3.916]	–	0.470 [-2.562, 3.627]	2.081 [-0.176, 4.699]	0.470 [-3.158, 3.939]	1.779 [1.084, 3.164]
3	3	95.09 (8.38)	0.027 [-3.394, 3.271]	–	-0.695 [-4.367, 2.937]	0.635 [-2.460, 3.967]	2.329 [0.406, 4.618]	0.888 [-2.888, 4.471]	1.777 [1.083, 3.190]
4	3	95.26 (8.33)	–	1.016 [-2.185, 4.238]	-1.065 [-4.881, 2.656]	0.648 [-2.385, 3.906]	1.859 [-0.454, 4.571]	0.759 [-2.321, 3.371]	1.744 [1.052, 3.121]
5	2	95.26 (8.22)	0.035 [-3.313, 3.164]	–	–	0.491 [-2.437, 3.590]	2.366 [0.551, 4.548]	0.685 [-2.825, 3.955]	1.706 [1.050, 3.034]
Verb growth prediction									
1	4	**55.78** (10.21)	1.121 [-2.718, 4.939]	4.695 [0.293, 10.126]	-1.414 [-5.834, 2.997]	-1.577 [-5.495, 1.805]	-0.298 [-3.686, 3.126]	-1.945 [-6.875, 2.120]	1.898 [1.016, 3.932]
2	3	55.84 (10.17)	1.084 [-2.728, 4.943]	4.097 [0.090, 9.085]	–	-1.701 [-5.605, 1.705]	0.143 [-2.956, 3.292]	-2.344 [-7.041, 1.566]	1.905 [1.047, 3.902]
3	2	56.18 (9.99)	1.337 [-2.441, 5.091]	4.021 [0.048, 8.903]	–	–	0.171 [-2.869, 3.155]	-3.258 [-7.549, 0.174]	1.828 [1.017, 3.702]
4	2	56.20 (10.15)	–	4.217 [0.211, 9.155]	–	-1.821 [-5.599, 1.474]	-0.068 [-3.044, 3.013]	-1.444 [-4.958, 1.362]	1.878 [1.038, 3.747]
5	3	56.20 (10.10)	–	4.824 [0.416, 10.113]	-1.407 [-5.734, 2.999]	-1.684 [-5.547, 1.617]	-0.505 [-3.796, 2.898]	-1.034 [-4.954, 2.024]	1.862 [0.999, 3.831]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of each condition indicates the effect of the proportion of correct responses for the first time. Age in months for the first time was included in model candidates as the control variable. We converted 18 and 19 months to 0 and 1, respectively, to match the scales with other fixed effects. Parameter σ represents the standard deviation of the random intercept.

Appendix F Correlation analysis details

Table F.1 Correlation between task performance in each condition and number of words for Analysis 3 in Study 1.

Age range	Condition											
	Match			Mismatch			Different-Object			Absent-Object		
	n	*ρ*	*p*	*n*	*ρ*	*p*	*n*	*ρ*	*p*	*n*	*ρ*	*p*
The preferential-looking-based task performance and common nouns												
18–19-months	25	0.08	0.69	25	0.18	0.40	25	-0.17	0.42	25	-0.21	0.32
19–20-months	24	-0.03	0.88	24	0.31	0.14	24	-0.25	0.25	24	-0.28	0.19
20–21-months	21	0.12	0.61	21	0.33	0.14	21	-0.04	0.87	21	-0.12	0.61
21–22-months	20	-0.15	0.54	20	0.21	0.36	20	-0.02	0.94	20	0.07	0.78
22–23-months	23	0.24	0.27	23	0.45	0.030	23	0.11	0.63	23	0.30	0.16
The preferential-looking-based task performance and verbs												
18–19-months	25	-0.07	0.75	25	-0.07	0.75	25	-0.00	0.99	25	-0.20	0.33
19–20-months	24	-0.03	0.88	24	0.23	0.28	24	-0.09	0.68	24	-0.38	0.069
20–21-months	21	0.09	0.70	21	0.45	0.043	21	0.15	0.53	21	-0.15	0.52
21–22-months	20	0.09	0.71	20	0.53	0.017	20	-0.19	0.43	20	-0.01	0.97
22–23-months	23	0.41	0.052	23	0.32	0.13	23	-0.28	0.19	23	0.27	0.22
The pointing-based task performance and common nouns												
18–19-months	14	0.38	0.18	15	0.27	0.33	15	-0.50	0.056	14	-0.40	0.15
19–20-months	16	0.12	0.66	16	0.62	0.010	16	-0.26	0.33	13	-0.30	0.32
20–21-months	17	0.34	0.18	17	0.30	0.24	17	0.13	0.61	15	-0.02	0.95
21–22-months	17	-0.14	0.59	17	0.25	0.33	17	-0.08	0.76	17	-0.06	0.81
22–23-months	19	0.33	0.17	19	0.48	0.039	19	0.15	0.54	19	0.03	0.90
The pointing-based task performance and verbs												
18–19-months	14	0.18	0.53	15	0.23	0.42	15	-0.35	0.20	14	-0.13	0.67
19–20-months	16	0.24	0.38	16	0.30	0.25	16	-0.05	0.86	13	-0.08	0.79
20–21-months	17	0.41	0.10	17	0.25	0.33	17	0.05	0.83	15	-0.16	0.58
21–22-months	17	0.06	0.83	17	0.57	0.018	17	-0.45	0.068	17	-0.56	0.019
22–23-months	19	0.35	0.14	19	0.39	0.10	19	-0.25	0.17	19	-0.33	0.31

Note. Spearman's rank correlation coefficient was used.

Appendix G Developmental changes in the proportion of absent pointing responses

A cross-sectional study

Using Bayesian hierarchical generalized linear models, we estimated the effects of developmental indices (age in months and total vocabulary size) and conditions on the proportion of absent pointing responses. The analysis procedure used was the same as in Analysis 1 and 2 in Study 1 (see Chapter 4), except for the dependent variable. The number of absent pointing responses for all trials within the same participants and conditions was used as the dependent variable. Other settings, such as the number of chains, iterations, and burn-in samples, were identical to those used in Study 1.

The results of the model comparison are shown in **Tables G.1** and **G.2**. Consistently, developmental indices of both age in months and vocabulary size were negatively related to the proportion of absent pointing responses for both the match and mismatch conditions, whereas only vocabulary size had a negative relationship with the absence of pointing responses for the different-object and the absent-object conditions (**Figure G.1**). Although some of the best models included the effects of the conditions, their credible intervals straddled zero. The interaction between vocabulary size and conditions was included in the best model with the match and mismatch conditions.

A longitudinal study

We also conducted a similar analysis for the longitudinal data. The settings were identical to those of a cross-sectional study, except for developmental indices and random intercepts. We used the time (first or second) as the developmental index and set random intercepts to account for within-participants effects for both time and condition.

As shown in **Figure G.2** and **Table G.3**, the proportion of absent pointing responses was higher for the first time than for the second time for the best model, with both the match and mismatch conditions. In other parameter estimates, the credible intervals straddled zero, although they were included in the best models.

Table G.1 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using age in months as a developmental index in a cross-sectional study.

Model rank	and fixed effects	WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			age	condition	age × condition	intercept	σ
The match and the mismatch conditions							
1	age	**166.78** (18.28)	-1.446 [-2.464, -0.713]	–	–	26.638 [11.903, 46.720]	4.278 [3.030, 6.365]
2	(intercept only)	167.61 (18.08)	–	–	–	-2.774 [-4.577, -1.442]	4.805 [3.407, 7.058]
3	age, condition	169.54 (18.63)	-1.458 [-2.456, -0.724]	-0.004 [-0.647, 0.646]	–	26.904 [12.155, 46.555]	4.297 [3.021, 6.344]
4	condition	169.72 (18.40)	–	0.002 [-0.644, 0.647]	–	-2.797 [-4.695, -1.392]	4.833 [3.439, 7.132]
5	age, condition, age × condition	171.64 (18.98)	-1.519 [-2.577, -0.739]	-1.755 [-11.184, 7.516]	0.089 [-0.381, 0.566]	28.067 [12.430, 48.853]	4.328 [3.060, 6.416]
The different-object and the absent-object conditions							
1	condition	**222.22** (17.49)	–	0.487 [-0.069, 1.060]	–	-1.893 [-3.048, -0.907]	3.505 [2.596, 4.857]
2	age, condition	223.46 (17.78)	-0.727 [-1.311, -0.215]	0.490 [-0.070, 1.067]	–	12.974 [2.486, 24.810]	3.279 [2.407, 4.607]
3	(intercept only)	225.21 (17.69)	–	–	–	-1.625 [-2.720, -0.713]	3.453 [2.556, 4.811]
4	age, condition, age × condition	226.00 (18.11)	-0.736 [-1.361, -0.188]	0.343 [-6.672, 7.338]	0.007 [-0.335, 0.351]	13.167 [1.952, 25.804]	3.296 [2.431, 4.612]
5	age	226.35 (17.91)	-0.720 [-1.300, -0.215]	–	–	13.129 [2.778, 24.783]	3.237 [2.376, 4.541]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of "condition" indicates the effect of the mismatch condition compared to the match condition for the upper half of the table and the absent-object condition compared to the different-object condition for the lower half of the table. Parameter σ represents the standard deviation of the random intercept.

Table G.2 Widely applicable information criterion (WAIC), posterior median, and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates using total vocabulary size as a developmental index in a cross-sectional study.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]				
			vocabulary size	condition	vocabulary size × condition	intercept	σ
The match and the mismatch conditions							
1	vocabulary size, condition, vocabulary size × condition	**165.69** (18.06)	-0.025 [-0.049, -0.007]	-0.776 [-1.836, 0.252]	0.010 [0.000, 0.022]	-0.591 [-3.001, 1.633]	4.898 [3.445, 7.350]
2	vocabulary size	167.29 (18.20)	-0.019 [-0.040, -0.002]	–	–	-0.998 [-3.338, 1.086]	4.731 [3.340, 7.071]
3	vocabulary size, condition	169.86 (18.56)	-0.019 [-0.040, -0.002]	-0.002 [-0.646, 0.647]	–	-1.028 [-3.325, 1.060]	4.748 [3.356, 7.041]
The different-object and the absent-object conditions							
1	vocabulary size, condition	**221.81** (17.60)	-0.013 [-0.028, -0.001]	0.491 [-0.069, 1.061]	–	-0.602 [-2.184, 0.962]	3.442 [2.542, 4.804]
2	vocabulary size, condition, vocabulary size × condition	223.03 (17.76)	-0.016 [-0.032, -0.002]	0.233 [-0.732, 1.184]	0.003 [-0.006, 0.013]	-0.430 [-2.083, 1.208]	3.468 [2.561, 4.859]
3	vocabulary size	224.75 (17.78)	-0.013 [-0.028, -0.001]	–	–	-0.346 [-1.861, 1.179]	3.384 [2.501, 4.716]

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of "condition" indicates the effect of the mismatch condition compared to the match condition for the upper half of the table and the absent-object condition compared to the different-object condition for the lower half of the table. Model candidates that did not include vocabulary size in the fixed effects are the same as in Table G.1. Parameter σ represents the standard deviation of the random intercept.

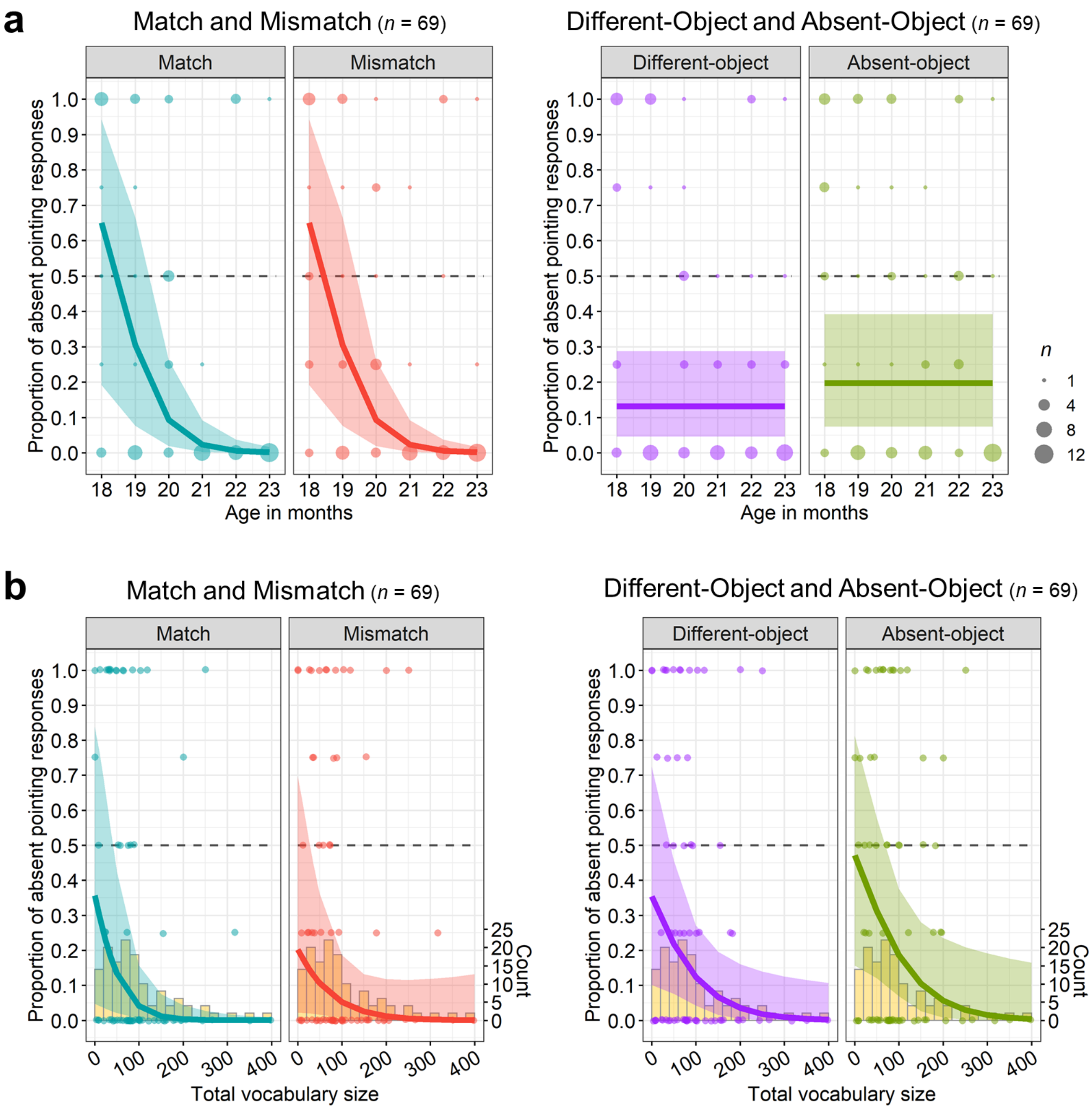

Figure G.1 Developmental changes in the proportion of absent pointing responses in a cross-sectional study. Thick lines and shaded areas represent the posterior median and 95% Bayesian credible intervals of expected values estimated from the selected best models, respectively. The dashed horizontal line indicates the chance level. (a) Age in months is used as a developmental index. The size of data points indicates the number of participants located at the same coordinates. (b) Total vocabulary size is used as a developmental index. Dots represent observed data. The histogram shows the distribution of the number of participants regarding total vocabulary size.

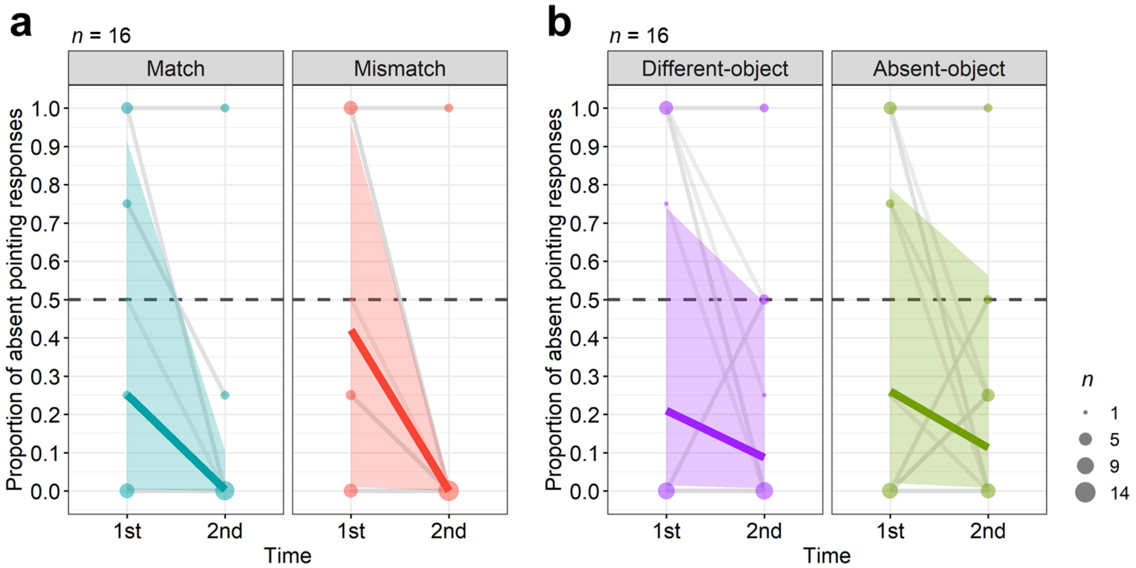

Figure G.2 Developmental changes in the proportion of absent pointing responses in a longitudinal study. The dots connected by gray lines indicate the same individual at different times. Other legends are the same as in Figure G.1-a. Comparisons are (a) between the match and the mismatch conditions and (b) between the different-object and the absent-object conditions.

Table G.3 Widely applicable information criterion (WAIC) posterior median and 95% Bayesian credible intervals (CIs) of parameter estimates for model candidates in a longitudinal study.

Model rank and fixed effects		WAIC (SE)	Parameter estimates: posterior median [95% CI]						
			time	condition	time × condition	intercept	σ_1	σ_2	σ_3
The match and the mismatch conditions									
1	time, condition, time × condition	**54.22** (9.75)	-4.641 [-7.440, -2.467]	3.586 [-0.461, 9.591]	-2.767 [-8.204, 0.363]	3.519 [-1.281, 8.393]	5.767 [3.102, 11.310]	0.410 [0.026, 2.207]	0.266 [0.012, 0.981]
2	time	55.32 (9.69)	-5.313 [-7.708, -3.598]	–	–	4.597 [0.780, 8.706]	5.158 [3.119, 9.459]	–	–
3	time, condition	57.67 (10.21)	-5.431 [-8.039, -3.387]	0.183 [-1.032, 1.419]	–	4.599 [0.498, 9.126]	5.189 [2.709, 9.758]	0.429 [0.038, 2.373]	0.238 [0.015, 0.889]
4	(intercept only)	143.82 (21.11)	–	–	–	-1.881 [-4.062, -0.264]	3.008 [1.819, 5.528]	–	–
5	condition	148.55 (21.99)	–	0.082 [-0.701, 0.870]	–	-1.948 [-4.181, -0.233]	3.030 [1.845, 5.546]	–	–
The different-object and the absent-object conditions									
1	time, condition	**89.59** (11.37)	-1.015 [-4.306, 2.388]	0.275 [-0.770, 1.398]	–	-0.316 [-6.062, 4.825]	0.551 [0.033, 3.211]	3.802 [2.016, 6.683]	0.373 [0.023, 1.360]
2	time, condition, time × condition	91.21 (11.72)	-1.249 [-4.792, 2.414]	-0.414 [-4.126, 3.217]	0.429 [-1.658, 2.598]	0.050 [-6.136, 5.576]	0.568 [0.039, 3.196]	3.871 [2.056, 6.747]	0.384 [0.023, 1.395]
3	time	158.39 (17.70)	-1.495 [-2.298, -0.739]	–	–	0.991 [-0.814, 2.832]	2.555 [1.626, 4.300]	–	–
4	(intercept only)	174.86 (20.48)	–	–	–	-1.130 [-2.539, 0.149]	2.331 [1.475, 3.935]	–	–
5	condition	179.21 (21.32)	–	0.124 [-0.565, 0.827]	–	-1.208 [-2.671, 0.121]	2.331 [1.481, 3.935]	–	–

Note. WAIC values for the best models among model candidates are shown in bold. The fixed effect of “condition” indicates the effect of the mismatch condition compared to the match condition for the upper half of the table and the absent-object condition compared to the different-object condition for the lower half of the table. Parameters σ_1, σ_2, and σ_3 represent the standard deviation of the random intercepts that correspond to participants, the combination of participants and times, and the combination of participants and conditions, respectively.

Summary (in Japanese)

論文題目

The Differentiation of Early Word Meanings from Global to Specific Categories: Towards a Verification of the “Semantic Pluripotency Hypothesis”

（言語発達初期における語の意味の未分化性と可塑的変化 :「胚性詞」仮説の検証に向けて）

論文要約

日本語を含めた多くの言語圏において，初期の語の多くはいわゆる〈モノの名前〉に対応する名詞が多くを占めている。このため，従来の発達科学者たちは，〈モノの名前〉の学習がなぜ早期に起こるのか，また，〈行為の名前〉に対応する動詞の学習はなぜ名詞の学習よりも難しいのか，といった観点から研究を進めてきた。一方，乳幼児期の語は，成人の語と比較して未分化な意味をもっており，見かけ上名詞だからといって安易に〈モノ〉を指示しているとは限らないと考える理論家もいたが，十分に実証されることなく今日に至っている。言語発達の過程およびメカニズムを真に理解するためには，語の意味の分節性を自明視せず，子どもの独特な語音－意味の構造を実証的に捉える必要がある。「語の意味は未分化な全体から特定の明確なカテゴリーへと変化する」と主張したウェルナーとカプラン（1963）の理論的仮説をもとに，筆者らは新たに「胚性詞」仮説を提案し，この仮説の検証を通じて，語の意味の発達を実証的に調べることを目指した。本論文では，特に初期に習得される「名詞的な語」に着目して，この胚性詞仮説を検証する。具体的には，①名詞の意味は初期には《靴を履く》のように少なくとも〈モノ＋行為〉の未分化な総体としての〈出来事〉に対応しており，〈モノ〉だけに分節化していない，②ある発達時期を境に，語の意味は〈出来事〉から分節化（脱文脈化）して，名詞は〈行為〉に左右されない〈モノ〉だけを指示するようになる，という 2 つの下位仮説を立て，複数の研究を実施した。

本論文は 5 つの章からなる。まず，第 1 章では，子どもの初期の語彙および語意学習に関して，理論・観察・実験のそれぞれのアプローチに基づく先行研究を概観する。そして，ウェルナーとカプラン（1963）の理論的仮説を中心に，初期の語がどのような独自の特徴をもつのかを議論するとともに，間接的ではあるが，彼らの仮説を支持する最近の知見を

取り上げる。その上で，初期の語の未分化性・可塑性を実証的にとらえるために，新たに胚性詞仮説を提案する。

第 2 章では，胚性詞仮説について調べた最初の予備的研究について報告する。19〜35 ヶ月児 36 名を対象に，2 つの動画刺激を対提示し，「靴はどっち？」といった質問に指さしで答えてもらう 2 肢強制選択課題を実施した。その結果，21 ヶ月以降の幼児は新奇な行為を伴っていても靴がある方の動画を選択するという結果が得られ，語の意味は〈モノ〉に分節化していることが示唆された。一方，21 ヶ月未満の幼児の場合，《靴を履く》のように〈出来事〉全体がまとまっていれば靴がある方の動画を選択したが，《靴をこねる》対《カゴをあたかも靴であるかのように履く》というように，〈モノ〉と〈行為〉とが不一致となる動画刺激を提示すると，《靴》と《履く》のどちらを重視して語の意味を判断すれば良いかが分からなくなるという過程が見出された。したがって，予備的ではあるものの，初期の語の意味は〈モノ〉だけに分節化していないという胚性詞仮説を支持する結果が得られ，名詞的な語の意味分化は 21 ヶ月頃に生じる可能性が示された。

第 3 章では，前章で得られた結果の追試および新たな研究を実施するための準備として，2 肢強制選択課題における指差し指標と注視指標とがどの程度同等に扱えるのかを探索した研究を取り上げる。前章で用いた指差し課題は，幼児の選択を明確に反映し，かつ容易に解析できるという利点をもっていたが，特に 2 歳未満の子どもに用いる際には，指差しができる場合とできない場合とがあり，欠測データが生じやすいという問題点を抱えていた。低月齢の幼児に対する指差しの代替指標として，多くの研究では刺激への注視割合が用いられてきたが，注視指標がどのくらい幼児の選択を反映しているのかは明らかにされてこなかった。そこで，我々は視線の向きから指差しの方向（右・左）を予測するモデルを開発し，これらのモデルを使うことで，注視と指差しはある程度同等に扱える指標であるということを定量的に示した。本章の研究は，胚性詞仮説の検証において，方法論的な柔軟性を大幅に向上させるものである。特に，低月齢児を対象とした場合にも，データの除外率を下げながら 2 肢強制選択課題を実施することが可能になり，さらに，注視と指差しという異なる指標を同じ統計手法によって解析し，両者の結果を直接比較することができるようになった。

第 4 章では，前章までに得られた知見を統合し，胚性詞仮説を複数の視点からより深く探究する。まず，名詞的な語の意味分化が生じると考えられる 18〜23 ヶ月に対象月齢を絞り，初期の語の意味が〈出来事〉全体を指示する段階から〈モノ〉だけを指示する段階へと分化するという第 2 章の結果について，その再現可能性を調べた。横断研究（69 名）および縦断研究（16 名）のどちらにおいても，また，指差しと注視のどちらの指標においても，胚性詞仮説を支持する同様の結果が得られた。この研究では，名詞的な語の意味分化は予備的研究で見られた 21 ヶ月よりも 2 ヶ月ほど早く観察されたが，それでも，語の

意味が〈出来事〉全体から〈モノ〉へと分節化する点，およびその発達的変化が 1 歳後半に見られる点は，予備的研究と共通していた。また，この月齢の幼児は，〈行為〉だけから名詞的な語の意味を予測することはできなかった。したがって，名詞的な語の初期の意味は〈モノ〉と〈行為〉とが未分化に融け合っているものの，そのなかで〈行為〉に相対的な比重が置かれているわけではなく，名詞に対応する〈モノ〉それ自体が語の意味判断には必要であることが示唆された。さらに，本章では，胚性詞仮説を発展させ，語の意味分化と語彙数の成長との関係を調べた。具体的には，名詞的な語の意味が〈モノ〉と〈行為〉の両方を含んでいた段階から〈モノ〉だけに分化してしまうと，子どもは〈行為〉を表現する語彙を失ってしまうことになるため，そのことが特定の〈行為〉だけを指示する動詞的な語の形成を促すのではないかと仮説を立てた。その結果，予測したように，名詞的な語の意味分化の程度は，特に動詞において，同時期の語彙数やその後の語彙数の成長に正の関連をもっていることが明らかになった。

最後に，第 5 章では，得られた知見を要約し，それらをもとに胚性詞仮説を精緻化させることを試みる。各章で扱った研究から得られた結果は，胚性詞仮説を支持し，ウェルナーとカプラン（1963）の理論的仮説に初めて実験的な証拠を提供し，さらに子どもの語意学習に関する今後の研究に豊かな示唆を与えるものといえる。特に，胚性詞仮説は，相互に関連するにもかかわらず，それぞれ個別の文脈で研究が進められてきた領域（例えば，出来事の分節化や, 文脈情報が語意学習に与える影響, 通状況的統計学習など）を統合し，名詞か動詞かに関わらず，初期の語の発達を統一的に編み直す端緒になると期待される。加えて，胚性詞仮説は，子育てや保育・教育・療育などにおける実践にも示唆を提供できるだろう。語の意味分化は子どもの内部で生じるため，子どもの発語のみから，その分化の度合いを判断することは難しい。しかし，語の意味の未分化性と可塑的変化という胚性詞仮説の視点を大人の側がもつことで，子どもをより注意深く観察し，それに応じて子どもへの関わり方を柔軟に変化させることが可能になるかもしれない。このように，萌芽的段階ではあるものの, 本論文はこれまで見落とされてきた初期の語の特異性を胚性詞仮説の観点から新たに浮き彫りにし，発達の理解に対する学術的・実践的貢献を果たすものといえる。

List of publications

Hagihara, H., & Sakagami, M. (2019). Shoki-gengo ni okeru imi no zentaisei to kasoteki henka: Kodomo no kotoba ni hinshi-kouzou wa aruka? [Semantic pluripotency in early word learning: Do children have an explicit part-of-speech structure?]. *Baby Science*, *18*, 14–24. Retrieved from https://www2.jsbs.gr.jp/LEARNED/18/2018-18_2_Hagihara.pdf

Hagihara, H., & Sakagami, M. (2020). Initial noun meanings do not differentiate into object categories: An experimental approach to Werner and Kaplan's hypothesis. *Journal of Experimental Child Psychology*, *190*, 104710. https://doi.org/10.1016/j.jecp.2019.104710

Hagihara, H., Ienaga, N., Terayama, K., Moriguchi, Y., & & Sakagami, M. (2021). Looking represents choosing in toddlers: Exploring the equivalence between multimodal measures in forced-choice tasks. *Infancy*, *26*(1), 148–167. https://doi.org/10.1111/infa.12377

Hagihara, H., Yamamoto, H., Moriguchi, Y., & Sakagami, M. (2022). When "shoe" becomes free from "putting on": The link between early meanings of object words and object-specific actions. *Cognition*, *226*, 105177. https://doi.org/10.1016/j.cognition.2022.105177