

ORIGINAL ARTICLE

The developmental trajectory of mandarin Chinese-speaking children's pure metonymy comprehension ability

Songqiao Xie*

University of Cambridge, Cambridge, UK

Abstract: This empirical study investigates Chinese children's developmental trajectory of pure metonymy comprehension. In the light of the experiment design in Jiang's (2019) and Köder and Falkum's (2020) studies, the present study, adopting a quantitative approach, employed a modified behavioral experiment and an eye-tracking experiment. Drawing on the experimental data, the study finds that: a) children's metonymy comprehension performance showed a tendency towards the U-shape in the behavioral experiment tasks; b) children's target (metonymy) fixation proportion, however, developed with age in the eye-tracking tasks; c) children's metonymy comprehension not only developed with age but also showed different features in different difficulty levels of metonymies. Thus, this study explains the U-shape by arguing that age-4 and -5 children's pure metonymy comprehension ability can be masked not only by a literal preference reported in Köder and Falkum's (2020) study but also by the high randomness of task results of the age-3 participants and the high level of difficulty of culture-related metonymies. Moreover, the study also argues that year six is a crucial stage for children's metonymy comprehension development, which provides implications for children's early figurative language education.

Keywords: children's metonymy; pure comprehension; developmental trajectory; eye-tracking experiment; behavioral experiment

*Corresponding author: Songqiao Xie, University of Cambridge, Cambridge; postal code: CB2 8PH, UK; sx264@cam.ac.uk

Received: April 22, 2022; **Accepted:** June 20, 2022; **Available online:** August 19, 2022

Citation: Xie S (2022) The developmental trajectory of mandarin Chinese-speaking children's pure metonymy comprehension ability. *Forum for Linguistic Studies* 4(1): 78–99. DOI: 10.18063/fls.v4i1.1457

1. Introduction

In recent years, the growing interest in figurative language use or use of expressions that require ad hoc concept constructions (Carston, 2002), doubled with the increasing interest in cognitive pragmatics, has given rise to a rich body of studies that informed the fields of human figurative language processing and production.

However, the existing figurative language studies have persistently focused on a narrow range of figurative language devices, among which the most frequently studied is metaphor (Rundblad and

Annaz, 2010), resulting in an inadequate concern for metonymy, a figurative device which expresses a real-world contiguity by referring to a target via a salient property. Metonymy is also claimed to rely on a cognitive pragmatic mechanism (Falkum, 2019) different from that of metaphor. While some studies were conducted on the cognitive mechanism of metonymy, few have approached this issue from a developmental perspective.

Even within the scarcity of literature on the trajectories of children's figurative language ability development, an imbalance of research efforts can be observed between children's metaphor development and their metonymy development. Existing studies (Özcaliskan, 2005; Peng and Zhang, 2009; Liu and Mi, 2008; Rundblad and Annaz, 2010) have found and confirmed children's early (at the age of four) ability, which improves chronologically with age, of understanding metaphorical uses of language, and have also realized that different categories of metaphors may be acquired at different stages (Zhou, 2003). However, when it comes to metonymy, children's metonymy development remained an uncharted territory until the recent decade.

Rundblad and Annaz (2010), Van Herwegen et al. (2013), Falkum et al. (2017) and Jiang (2019) have made valuable research efforts which paint a polychrome picture of this field. It is polychrome because, on the one hand, their studies serve as an intriguing guidance to a wide range of topics regarding children's metonymy, from metonymy comprehension to metonymy production (Falkum et al., 2017; Jiang, 2019), and from chronological development (Van Herwegen et al., 2013) to U-shaped development (Falkum et al., 2017); on the other hand, extant research also leaves contradictions (Van Herwegen et al., 2013; Falkum et al., 2017) and questions open to discussion. For example, one pending issue is how to, at the cognitive level, refine the existing experiments to further distinguish between children's metonymy comprehension and metonymy production. How to provide a more convincing and solid explanation for the U-shape observed in the previous research is another problem unsolved. Furthermore, whether the acquisition of different categories of metonymy corresponds with different age stages is also an interesting topic which few studies have explored.

In the light of the questions identified above, the present study aims to investigate Chinese children's pure metonymy comprehension development, employing eye-tracking and improved picture-selection activities as experiment instruments.

Köder and Falkum (2020) and Jiang (2019) managed to differentiate between children's metonymy production and comprehension, and Köder and Falkum (2020) concluded that "eye-tracking is a 'purer' and cognitively less demanding measure of figurative language comprehension." So, the present study employs an eye-tracking experiment, and also, with an intention to improve Jiang's (2019) experiment method (picture-selection, which touches upon the "production" level), the author of the present study will conduct the picture-selection activity followed by a retrospective interview with the children subjects, to minimize the influence of speaker intention in the instant context.

Since the subjects in the present study are Chinese children, this study will also serve as an attempt to test the translingual applicability of the U-shape observed in the previous studies, which were conducted in English and Norwegian contexts. Not confined to this, considering that the discussion (Köder and Falkum, 2020) of the U-shaped development of children's metonymy was backed up with solid evidence, this study attempts to provide a new perspective for the explanation of the U-shape, with the combined evidence from the eye-tracking and modified behavioral experi-

ment data.

Also, given that some metonymies are lexicalized (Hilpert, 2006), especially the culturally related ones, while some are ad-hoc metonymies whose understanding requires instant construction of relations, the present study will also probe into, though not so deeply, the question of the acquisition order of different categories of metonymies. To achieve this, my experiment will include a comparison between three levels of metonymies assigned with different difficulties.

2. Research on metonymy

In our daily communication, figurative language plays a ubiquitous part, making it possible for speakers and listeners to go beyond conventional lexical meanings and construct novel context-dependent meanings. One case of this is metonymy, a trope (Corbett and Connors, 1999) or a figurative device that expresses a real-world or mental contiguity by referring to a target via a salient property (Koder and Falkum, 2020). Refer to the examples in (i), (ii) and (iii):

- (i). The piano is in a bad mood. (Panther and Thornburg, 2005)
- (ii). The sax has the flu today and he will not be able to play tonight. (Zhang and Lu, 2010)
- (iii). The moustache sits down first. (Falkum et al., 2017)

In (i) and (ii), the pianist and the saxophonist are referred to via “piano” and “sax” respectively, the names of the instrument they play. In (iii), the “moustache” refers to a person wearing a noticeable moustache, which possibly differentiates him or her from others in the context. The piano-pianist, sax-saxophonist, and moustache-moustache wearer are in a contiguity relation, either conceptually or physically.

2.1. Metonymy from cognitive perspective

In figurative language literature, in spite of the burgeoning research interest in the cognitive process of figurative language, extant studies have primarily centered on a narrow range of figurative devices, mainly metaphor (Rundblad and Annaz, 2010), unfortunately resulting in an inadequate attention towards metonymy.

It was only after the end of the 1990s that metonymy started to become a research focus among cognitive linguists (Langacker, 1993; Radden and Kövecses, 1999; Panther and Radden, 1999; Barcelona, 2000; Alač and Coulson, 2004), whose works mainly focus on the mental conceptualization strategies and cognitive mechanisms of metonymy and interfaces and demarcations between metaphor and metonymy. Langacker (1993) categorized metonymy as a “reference point phenomenon”, based on which Alač and Coulson (2004) described the cognitive process of metonymy as a process where an item with relatively higher salience serves as the reference point, rendering other less salient items mentally accessible. Building on Lakoff and Johnson (1980), who proposed the idea of ICM (Idealized Cognitive Model) and argued that the human conceptual system is figurative in nature, Radden and Kövecses (1999) reemphasized cognitive nature of metonymy and defined metonymy as a cognitive process that functions within one ICM. All of these research efforts have deepened the understanding of the nature of metonymy and established a research trend and guidance for further cognitive linguistic studies on metonymy.

2.2. Research on children's metonymy

Since associative relations, which render the comprehension and production of figurative language, are frequently exploited in our everyday communication, it is important to study when people acquire this ability and how this ability develops from childhood to adulthood. This problem has been studied since the 1960s, when researchers (Asch and Nerlove, 1960; Gardner, 1974; Mendelsohn et al., 1984) tended to address this issue from a semantic perspective. However, in the literature on figurative language cognition, it was not until the 2000s that children's uses of figurative language started interested researchers (Zhou, 2003; Özcaliskan, 2005; Peng and Zhang, 2009; Liu and Mi, 2008; Rundblad and Annaz, 2010) working on language acquisition and cognitive development. Liu and Mi's (2008) research found that children of four are equipped with the ability to understand metaphorical utterances in neutral context and are also able to make abstract reasoning about the metaphorical mappings; Rundblad and Annaz's (2010) result confirmed that the ability to comprehend metaphor resides in children early on and improves steadily throughout childhood to adulthood. There is growing consensus that children have an early competence of and affinity with figurative language.

As can be concluded from the existing studies on children's figurative language comprehension, much has been said about when and how children begin to cope with the uses of metaphor, while, to the best of my knowledge, less is known about children's ability of metonymy. Only a handful of research has investigated children's development in comprehension strategy and production purpose of metonymic expressions (Rundblad and Annaz, 2010; Van Herwegen et al., 2013; Falkum et al., 2017; Köder and Falkum, 2020). Though few, these research efforts have opened up a new area for the study of children's metonymy comprehension. Rundblad and Annaz (2010) and Van Herwegen et al. (2013) investigated the trajectory of development of metonymy comprehension from year 5 and 6, and discussed the result comparatively with the result of metaphor, and finally found in children an early metonymy comprehension competence which could chronologically improve with age. Falkum et al. (2017) conducted a research studying children's metonymy comprehension and production, and found a U-shape in metonymy comprehension competence variation, with 3-year-olds doing surprisingly better than 4- and 5-year-old children; Jiang (2019) found the similar result

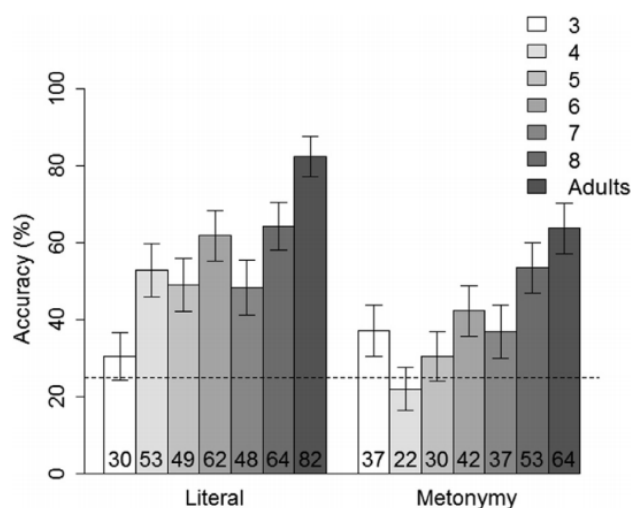


Figure 1. Percentage of correct picture choice in the literal and metonymy conditions (Köder and Falkum, 2020).

in the referential game she designed for the children tested. Köder and Falkum's (2020) conducted an empirical study on young children's (aged 3–8) comprehension of novel metonymy and they obtained both a similar and a different result from the two sections of the experiment they conducted, thus claiming that eye-tracking is “a purer and cognitively less demanding measure of figurative language comprehension” (Köder and Falkum, 2020).

2.3. Questions for further investigation

While providing valuable insights into children's abilities and affinity with metonymy, the extant studies are not flawless, thus leaving several contradictions and questions open to further investigation.

One thing problematic is a lack of effort to further distinguish between metonymy comprehension and production, especially in the experiment methods of previous research. Children's ability of metonymy consists of metonymy comprehension and metonymy production, as most of the studies on children's metonymy has differentiated between the two levels, focusing on either production on a pragmatic level or comprehension on a cognitive level. The researchers (Falkum et al., 2017; Jiang, 2019) have also designed and conducted their experiments accordingly. In Falkum et al.'s (2017) study and Jiang's (2019), researchers employed picture selection as their experiment instrument to examine and collect data concerning children's metonymy understanding, while using other methods to examine metonymy production. This seems to have made clear the boundary between the cognitive level and pragmatic level; however, the method itself is still problematic, as picture-selection inevitably involves interactions between the examiners and the children participants, whose perception of the examiners' speaker's intention would, to a large extent, affect their choices of the pictures. In other words, picture-selection as a method, though aiming to examine metonymy comprehension, still touches upon the production level, thus being a less pure method for metonymy cognition. Consequently, the present study, building on a reflection of the existing studies (Van Herwegen et al., 2013; Falkum et al., 2017; Jiang, 2019; Köder and Falkum, 2020), plans to employ picture-selection + interview and eye-tracking experiments, which was claimed to be a purer and less demanding measure of figurative language comprehension (Köder and Falkum, 2020), to collect data of children's pure metonymy comprehension.

Another pending issue is about the discussion of the U-shape developmental curve, which has been observed in several cognitive domains (Yan and Huang, 2009), not confined to the metonymy comprehension. One contradiction is that while Falkum et al. (2017) found a U-shape in her study, Van Herwegen et al.'s (2013) result found a linear steady improvement of children's metonymy comprehension. It might be due to a relatively small number of subjects in Van Herwegen et al.'s (2013) study, which might have concealed the U-shape. However, in Köder and Falkum's (2020) study, the researcher provided alternative explanations, claiming that the seemingly decreased competence of age-4 and age-5 children could be masked by a literal preference at that age, which means around the age of 4 and 5, children undergo a period when they show a general preference for literal interpretations and a rising awareness of speaker's intentions. In other words, it is possible that age-4 and -5 children are equipped with a better comprehension competence of metonymy; however, this competence is hidden under the cloak, and thus not manifested in the previous studies. Considering that U-shape is still a puzzling aspect of children's metonymy and that the previous study (Köder and Falkum, 2020) did not provide a thorough discussion over it, the present study

aims to provide some possible insights into this topic.

There is also a question that resides within “metonymy” itself, concerning the different categories of metonymic expressions. Few of extant studies on children’s metonymy comprehension paid attention to the possible difference of comprehension order resulted by different kinds of metonymies. In Zhou’s (2003) study on children’s developmental trend of metaphorical ability of time, the researcher realized that children’s comprehension ability of different categories of metaphorical expressions may be different with their age. For example, children’s comprehension of personification of time and space-time metaphors appears earlier than other categories of metaphors (Zhou, 2003). Accordingly, we tempt to ask a similar question about metonymy: are different categories of metonymies acquired by children in different time order? Which categories of metonymy are acquired earlier? Given that the time is limited, the present study does not address this issue in detail; however, the present study, in order to avoid stereotyping metonymy as “all metonymies”, sets the operational concept of metonymy by dividing metonymic expressions into different difficulty levels. In the behavioral experiment, the stimuli contain three difficulty levels of metonymies: the lowest difficulty-level of metonymy in this study is concerned with the relations of human-body-related physical contiguity, since children are early familiar with animated and human-like properties (Piaget, 1964), for example, the red cap and the cap wearer; the medium level is concerned with the relations of non-body-related contiguity; the highest level is added with cultural-related elements, for example, “lose your bowl” and “lose your job” (in the Chinese context).

As for the stimuli in the eye-tracking experiment, the present study modifies Peirsman and Geeraerts’ metonymy classification model and constructs a three-dimensional model where metonymies, especially for children, can be classified into different types and difficulty levels.

Peirsman and Geeraerts (2006) provided a classification model for metonymies, where a large number of typical metonymies are differentiated and categorized in terms of the type of contiguity they are motivated by. Their model is three-dimensional (**Figure 2**).

As positive shifts are made on each dimension line (domain, boundedness or strength of contact), the difficulty of processing and understanding the certain metonymies increases. For example, along

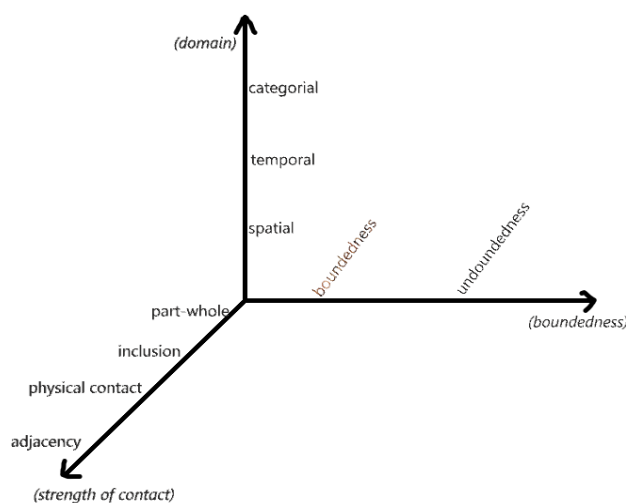


Figure 2. Metonymical patterns in Peirsman and Geeraerts (2006) (adapted).

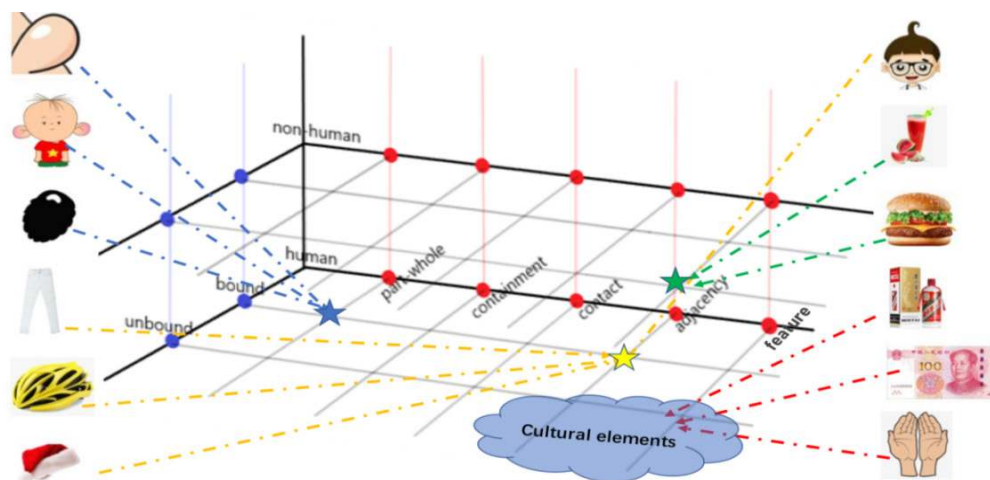


Figure 3. A revised model of metonymy taxonomy.

the line of strength of contact, the contact between the beard and the beard wearer (a part-whole contiguity) is stronger than the contact between the clothes and the clothes wearer (an adjacency contiguity), and thus the difficulty of using beard to refer to a man who wears a beard is lower than the difficulty of using a pair of shoes to refer to the person in them.

The present study, on the basis of Peirsman and Geeraerts' model, has made realistic adjustments in setting the difficulty of the stimuli in the experiment. Since children are early familiarized with animated and human-like properties (Piaget, 1964), human-body-related metonymies are considered less difficult than non-body-related metonymies, which is, in the present study, the primary standard for metonymy difficulty classification. Then, under this primary standard, the classification is subject to the above-mentioned model.

As a result, in **Figure 3**, the three dimensions are changed, compared with Peirsman and Geeraerts' model, into "human or not", "boundedness" and "strength of contact". Human related metonymies are attributed to the first layer, namely the "human platform", and are regarded less difficult to comprehend for children than the non-human-related metonymies, which belong to the second layer. For example, the clothes-wearer metonymy is easier than the bottle-water metonymy, although it reverses the rank between containment contiguity and adjacent contiguity.

The U-shape developmental curve in children's metonymy comprehension lacks test in the Chinese speaking context. Falkum et al.'s (2017) experiment exploring children's acquisition of metonymy was conducted in English, while the experiment of Köder and Falkum's (2020) another study was conducted in Norwegian, which marks an inter-language attempt. However, in view of the fact that English has, in history, received heavy influence from Scandinavian languages (Lutz, 2017), whose descendants include Norwegian, I argue that this attempt or extension is not far enough. To further test the universality of the conclusion made by the study, the present study is conducted to investigate the metonymy comprehension ability of children whose mother tongue is Chinese, a language that does not belong to the Indo-European language family.

Aware of the problems identified above, the author of the present study intends to address the following research questions:

- Can the U-shape be replicated in the Chinese speaking context?
- How can the explanation of the U-shape be further improved concerning the metonymy competence of children of different age groups?
- Can higher difficulty levels of metonymies mask children's pure metonymy comprehension?

3. Methodology

The present study employs a quantitative approach, conducting a three-section behavioral experiment (picture-selection plus retrospective interview + story Q&A) and an eye-tracking laboratory experiment. The behavioral experiment in this study partly follows and modifies the behavioral experiment designs in previous studies (Jiang, 2019; Falkum, 2020) by adding retrospective interview sections and adding stimuli of different difficulty levels. Eye-tracking is increasingly applied as an experiment instrument in the fields of human cognitive studies (Dong, 2014) and the method itself is supported by the Eye-brain Unity Hypothesis, which states that there is no appreciable lag between what is fixed and what is processed (Just and Carpenter, 1980). Thus, the present study employs an eye-tracking experiment to explore children's inner comprehension process of metonymy expressions.

3.1. Behavioral experiment

In the behavioral experiment, which is divided into three sections (each section contains 5 questions), we employed two experiment methods, including picture-selection plus retrospective interview (for section 1) and story Q&A (for sections 2 and 3). The three sections of the experiment are assigned with different levels of difficulty, aiming to investigate the children's metonymy comprehension development in different types or difficulty-levels of metonymies. All the participants are asked to complete all the three sections. In this experiment, the experimenters collect the data of error rate in all the three sections, the data of reaction time in sections 1 and 2.

3.1.1. Participants

We invited 32 children, all of whom are screened to exclude health and intellectual impairments and speak Chinese as their mother tongue, to participate in our behavioral experiment. The age of the children ranges from 3-year to 8-year, and all the participants were distributed into three age groups: group 1 (3 years old), group 2 (4–5 years old) and group 3 (over 6 years old). The context of this experiment is an international kindergarten located in Gaoxin District, Zhengzhou. Written parental consent was obtained prior to the experiment.

Prior to the experiment, we conducted a pilot study with two 3-year-old children to make sure that children older than 3-year can understand the words uttered by the experimenters and can recognize the different components that appear in the pictures.

3.1.2. Design

This was a 3 (age) \times 2 (gender) \times 3 (task) in-subject design, with age and gender as between-groups factors and difficulty level as an in-group factor.

The three independent variables are: (1) age (year 3, year 4–5, year 6–8); (2) gender (female vs.



Figure 4. Example stimulus.

male); (3) task (picture selection + retrospective interview of human body-related metonymy, story Q&A of non-body related metonymy, story Q&A of culture-related metonymy).

There were also three dependent variables: error rate, experiment score, reaction time.

3.1.3. Stimuli and procedure

In the first section, aiming to test children's metonymy comprehension competence concerning the relations of physical human-body-related contiguity, the experimenters would show the participants 5 sets of pictures. Each set consists of 4 pictures: one background picture and three option pictures. The background picture gives participants a general idea of "what" is going on with "whom". The three option pictures contain one literal picture (the object, for example the "big nose" itself), one metonymy picture (the right answer) and the irrelevant picture (the other person from the background picture).

The participants are tested one by one. At each time, one participant is shown one set of the pictures while being told a story which narrates the scene and gives the instruction to choose one picture. Each story is composed with one context utterance ("These two cooks work in one restaurant..."), which helps the children to establish a focus on both human-components in the background picture, and one target utterance, which helps the participants to differentiate between the literal and metonymic pictures. Each target utterance contains a referential expression of the metonymic target ("the big nose...") and a describing expression ("cooks the better") which helps with the instruction.

One example of the story is shown below. (Since the whole experiment is conducted in Chinese, the lines below are the translated version of the story told in the experiment.)

"These two cooks work in one restaurant. However, most of the customers of the restaurant comment that the big nose cooks better meal. So, please tell me who cooks the better meal?"

After the instruction is given, the reaction timing starts, and stops once the participant gives his or her choice. Then, after the participant's choice is given, one follow-up question ("Could you tell

me why you choose this?") is asked to collect data concerning the retrospective comprehension process of the participant.

In the second section, the experimenter tells the participant short stories (5 in total) concerning non-body-related metonymies. Each story contains the background information of the target and the metonymic expression. Following the story, two choices are given to the participant who is instructed to pick one according to the context of the story. The reaction of each participant is timed.

One example is shown below.

"Lu Xun is a famous thinker and writer in China. As a writer, he has many famous works, however, many people think that Lu Xun is really hard to read. So, please tell me what is hard to read, Lu Xun the person himself or the works he wrote?"

In the third section, the experimenter tells the participant short stories (5 in total) concerning metonymies with cultural-related elements. Each story contains the background information of the target and the metonymic expression; however, this time no choices would be given to the participant—the experimenter would ask directly for an explanation of the meaning of the metonymic expression. In this section, the experimenter does not time the reaction of each participant, considering that the difficulty of the metonymies in this section is relatively higher than the previous, and that, therefore, "no response" from the participant may occur frequently.

One example is shown below.

"Uncle Wang has messed things up in his work. He does not dare tell this to his leader because he is afraid of 'losing his bowl' (understandable within the Chinese cultural background). So, please tell me what 'losing his bowl' means here?"

3.2. Lab experiment

This eye-tracking experiment was conducted in the cognitive laboratory of Shanghai International Studies University. The participants' eye movements were tracked by Eyelink-1000 plus, an eye-tracking device with a sampling rate of 1,000 hz. The whole experiment procedure was designed and programmed on Experiment Builder, an experiment design software intended for Eyelink devices. Data collected in this experiment were used to analyze children's eye movements during the process of hearing the audio story while looking at the picture stimuli on the display screen. After collection, the data were generated on Data Viewer, and analyzed on Python.

3.2.1. Participants

Six children ranged from 3 to 8 years old were invited and tested in this experiment. **Table 1** lists the mean age, number and gender distribution in the different age groups. All participants were native Chinese speakers screened to exclude health, vision and intellectual impairments. Children were recruited from kindergartens and schools in Shanghai. Written parental consent was obtained prior to the experiment. Each participant was tested individually in the eye-tracking lab.

3.2.2. Design

This was a 3 (age) × 2 (gender) × 2 (difficulty level) in-subject design, with age and gender as between-groups factors and difficulty level as an in-group factor.

Table 1. Participants

Age group	Mean age	Number	Gender (f/m)
Year 3	3.46	2	2/0
Year 4–5	5.04	2	1/1
Year 6–8	7.34	2	2/0

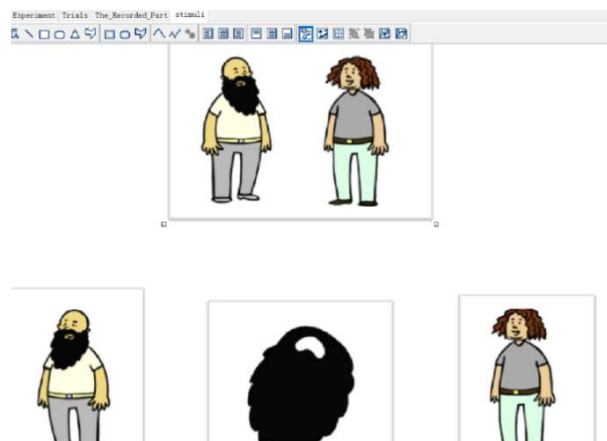


Figure 5. Example stimulus.

The three independent variables are: (1) age (year 3, year 4–5, year 6–8); (2) gender (female vs. male); (3) difficulty level (lower difficulty levels vs. higher difficulty levels).

The eye movement measure is children's fixation duration on the target areas.

3.2.3. Stimuli and procedure

The stimuli contain 12 sets of pictures, the first two sets being training stimuli. The rest 10 sets of picture stimuli are divided into four difficulty levels (two of lower levels and two of higher levels), following the classification provided in §2.4 (Figure 3). Every participant was asked to complete all the 12 trials.

Each set of pictures contains four pictures placed on four areas of the screen. The largest picture on the upper half of the screen is the background picture, which helps the audio listener to form the context of the story; the three pictures on the lower half are option pictures: one literal picture (the object, for example the “big beard” itself), one metonymy picture (the right answer) and the irrelevant picture (the other person from the background picture). The participants are asked to choose one, by pointing with finger, according to the story, which lasts around 20,000 ms, they hear.

Each audio stimulus, similar to the stories in the behavioral experiment, consists of one context utterance and one target utterance. Each target utterance contains a referential expression of the metonymic target and a describing expression, which helps to instruct the participants to make the choice.

The set of pictures are shown to the participant 4,000 ms earlier than the audio is displayed, which allows the participants enough time to recognize and be familiar with the elements in the pictures.

Eye-movements of participants are tracked and recorded during the whole process of each task. Viewing was binocular, but only the movement of the left eye was tracked and recorded¹.

Participants were positioned for individual comfort in an adjustable chair with a chin rest to minimize head movements for an increase in measurement accuracy.

4. Findings and discussion

4.1. Behavioral experiment results

To examine the age differences in metonymy comprehension and the influence of the task differences on performance, we used Excel to analyze the data collected from the behavioral experiment, treating “age” and “task” as independent variables, while “error rate”, “experiment score” and “reaction time” as dependent variables.

Concerning the “error rate”, all the wrong responses in picture-selection, including choosing the literal pictures and the irrelevant pictures, were all coded as “error”. Concerning the “experiment score”, all the right answers were coded as “1”; all the literal choices in picture-selection, all the irrelevant responses in retrospective explanation and all the wrong answers in other tasks were coded as “0”; all irrelevant choices in picture-selection were coded as “X” (not shown in the tables); all the half-relevant responses in retrospective explanation were coded as “0.5”.

Concerning the “reaction time”, considering the relatively high difficulty of task three and the ages of the children, we did not time the reaction of the participants in the third task, the story Q&A on metonymies with cultural elements.

4.1.1. Error rate

Table 2 presents the error rate data of the children’s performance in the 3 experimental tasks. One-way analysis of variance (ANOVA) between groups is carried out on children of different ages in different experimental tasks. In the picture-selection task (the retrospective explanation task is not included), we find no difference between the performance of the age-3 group and the age-4-to-5 group ($F = 0.306$, $P = 0.587$); in comparison to the age-4-to-5 group, age-6-to-8 group demonstrate a lower error rate, however, this difference is not statistically significant ($F = 2.673$, $P = 0.119$). In the story Q&A concerning non-body related metonymies, no difference is found between the performance of the groups of lower ages (age-3 and age-4-to-5) ($F = 0$, $P = 1$); however, the error rate of the age-6-to-8 group shows a significant drop ($M = 0.12$, $F = 4.314$, $P = 0.052$). In the third task (story Q&A concerning metonymies with cultural elements), significant difference can only be detected between age-3 group and age-4-to-5 group ($F = 5.018$, $P = 0.038$).

4.1.2. Experiment score

We analyzed the data of experiment scores by conducting one-way ANOVA between groups. **Table 3** mainly shows how the mean scores of picture-explanations and the mean total scores vary between the three groups.

A significant trend can be found in picture explanation that children’s ability to explain their

¹ Kliegl et al. (2006) has shown high correlations between right-eye and left-eye fixation durations ($r = 0.98$).

Table 2. Error rate of 3 age groups in 3 tasks

Age group	Picture selection		Non-body Q&A		Cultural-related Q&A	
	M	SD	M	SD	M	SD
Year 3	0.24	0.2459	0.3	0.2539	0.86	0.1350
Year 4–5	0.18	0.2394	0.3	0.2357	0.64	0.2797
Year 6–8	0.04	0.1265	0.12	0.1398	0.52	0.2700

Table 3. Score of picture explanation and total score

Age group	Picture explanation		Total score	
	M	SD	M	SD
Year 3	1.6	1.1972	9.6	2.2828
Year 4–5	2.4	1.3703	11.8	3.2335
Year 6–8	4.5	0.7454	16.1	2.3428

Table 4. Reaction time of different groups in tasks 1 & 2

Age group	Picture-selection		Non-body related Q&A	
	M (ms)	SD	M (ms)	SD
Year 3	3,559	3,183.6	2,203	1,627.4
Year 4–5	2,517	1,720.2	1,995	431.8
Year 6–8	2,101	2,479.1	1,239	620.5

choice of metonymic pictures improves with age. This improvement is most obvious between the group of year 4–5 and the group of year 6–8 ($F = 18.123$, $P = 0.000$).

In the analysis of the data “total score”, which consists of four parts (picture-selection score, picture-explanation score, task 2 score and task 3 score), we find an improvement, not very significant though, from the group of year 3 to the group of year 4–5 ($F = 3.089$, $P = 0.096$). However, from year 4–5 to year 6–8, a significant development can be observed in children’s metonymy comprehension ($F = 11.597$, $P = 0.003$).

4.1.3. Reaction time

Table 4 presents the one-way ANOVA result of the reaction time data of the three age groups in picture-selection task and task 2 (story Q&A concerning non-body related metonymies). The participants’ reaction time in picture-selection task does not have significant difference, either between year 3 and year 4–5 ($F = 0.829$, $P = 0.375$), or between year 4–5 and year 6–8 ($F = 0.109$, $P = 0.668$). In task 2, the age-6-to-8 group shows a significant improvement in reaction rate compared to the group of year 4–5 ($F = 10.002$, $P = 0.005$).

4.2. Eye-tracking results

Heatmaps in §4.2.1 are generated by Data Viewer (64 bit 4.1.1) and the figures in §4.2.2 are generated by Numerical Python (NUMPY) packages in Python.

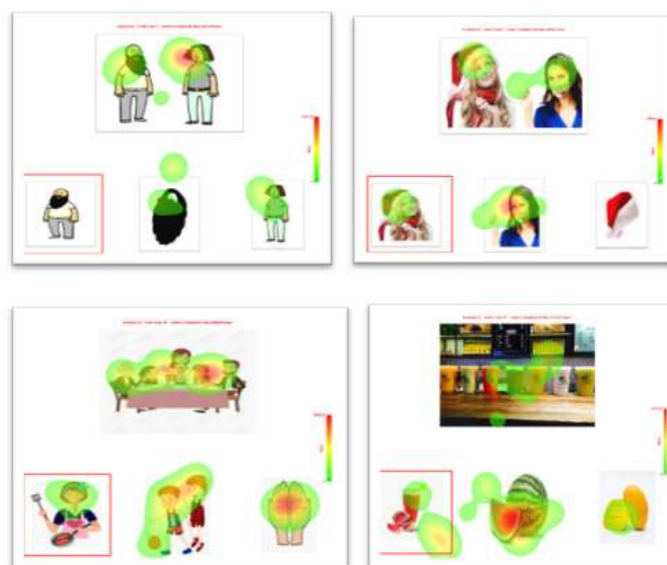


Figure 6. 3-year-olds' fixation area.



Figure 7. 4-to-5-year-olds' fixation area.

4.2.1. Gaze heatmap

The four heatmap examples in **Figure 6** show the 3-year-olds' fixation area during the 10,000 ms after the target utterance is heard. The four heatmaps cover all the four difficulty levels (from low to high clockwise) according to Figure 3. The pictures surrounded by red squares are the metonymic pictures, namely the target areas.

As can be concluded from the heatmaps, the fixation of age-3 participants on the different areas of the pictures are, to a large extent, at chance—there is no obvious preference between the three option pictures, and there is also no noticeable competition between the target (metonymic) picture and the literal picture.

The four heatmap examples in **Figure 7** show the 4-to-5-year-olds' fixation area during the

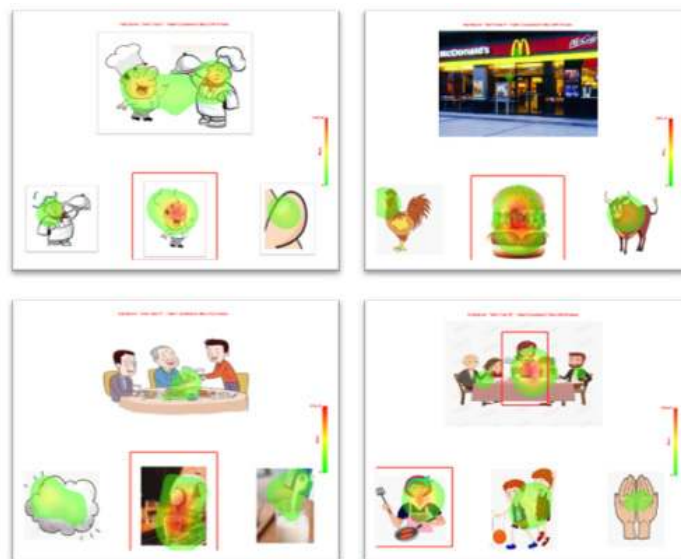


Figure 8. Over-6-year children's fixation area.

10,000 ms after the target utterance is heard. As can be concluded from the four heatmaps, the age-4 and age-5 children show a high preference to the literal pictures, whose fixation competes with fixation on the target pictures.

Concerning the pictures of lower difficulty levels of metonymies (the two in the first line), children of age-4 and -5 fix their eyes on irrelevant pictures significantly less, compared with the random fixation of 3-year-olds. However, as for higher difficulty levels, the gaze heatmaps of age-4-to-5 children are still largely disordered and random.

The four heatmap examples in **Figure 8** show the over-6-year children's fixation area during the 10,000 ms after the target utterance is heard. As can be concluded from the heatmaps, over-6-year participants show a significantly different and improved fixation heatmap compared with the other two age groups.

Compared with the heatmaps of age-4-to-5 children, over-age-6 participants' heatmaps in low difficulty level pictures show a clear preference towards target pictures and thus a less competition between target pictures and literal pictures. Furthermore, concerning the metonymies of higher difficulty levels (the second line), the heatmaps of over-6-year children show a lower randomness and the target areas are more heated.

4.2.2. Fixation proportion

Figure 9 presents the proportion of fixation on the target areas (in the trials containing lower difficulty levels of metonymies) during the 10,000 ms after the audio target utterance is played. The three lines represent the fixation performance of the three age groups respectively.

Fixation is detected and recorded by the eye-tracker when an eye movement lasts over 50 ms within the range from 0.5° to 2.0° (EyeLink Data Viewer 4.1.1, 2019). Thus, target fixations are hereby defined as the fixations whose coordinates fall within the area of target metonymy pictures.

Since the target area in each trial covers approximately 15% of all the picture areas, the fixation

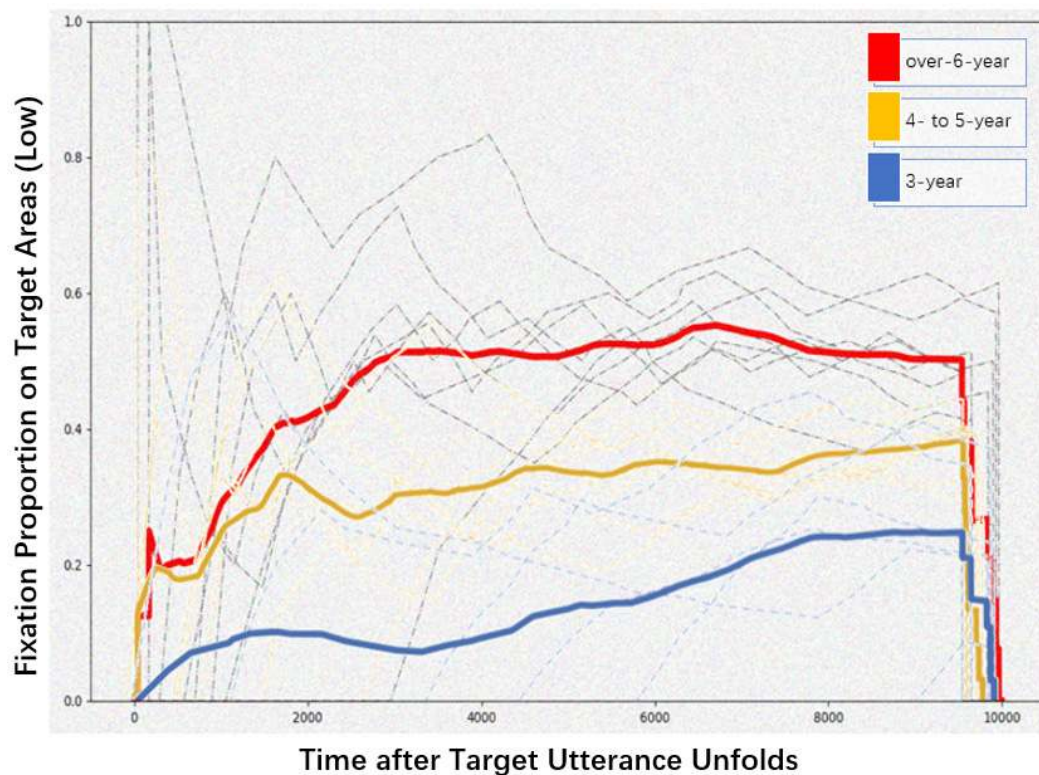


Figure 9. Fixation proportion on target areas (low difficulty levels).

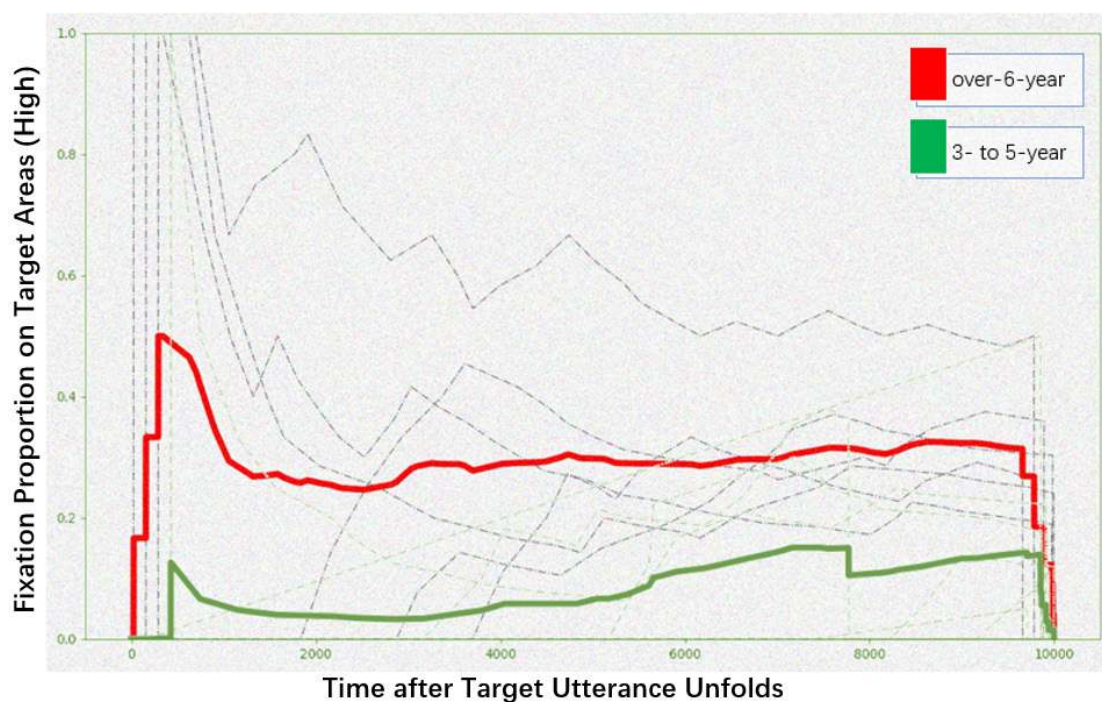


Figure 10. Fixation proportion on target areas (high difficulty levels).

of the 3-year-olds, though with a rise after the target utterance unfolds, can be regarded as largely at random. For the older age groups, the proportion of target fixations witnesses a rise as the audio target utterance is played. However, the target fixation proportion of over-6-year participants shows an

even more significant rise and levels out at around 0.55 as time passes.

Figure 10 presents the proportion of fixation on the target areas (in the trials containing higher difficulty levels of metonymies) during the 10,000 ms after the audio target utterance is played. The red line represents the fixation performance of the over-6-year participants and the green line, the performance of the other two age groups.

When being exposed to metonymic expressions of higher difficulty levels, both the age-3 children and the age-4-to-5 children produced a relatively random fixation pattern between different areas, with the target fixation consistently lower than 0.2 along the 10,000 ms.

However, target fixation proportion of the oldest group shows a different and improved feature. As can be concluded from Figure 10, their target fixation proportion, though being lower compared with Figure 9, is significantly above random, which means that their comprehension ability of higher difficulty levels of metonymies has developed at this stage.

Between Figure 9 and Figure 10, there is another comparison noted between lower and higher difficulty levels of metonymies. In different age groups, children's target fixation proportion in trials of higher difficulty levels is noticeably lower than that in trials of lower difficulty levels, except for the 3-year-olds whose fixation appears largely random in each level.

4.3. Discussion

4.3.1. U-shape tendency in picture selection

A U-shape demonstrating the trajectory of children's metonymy comprehension development was found by Falkum's (2017) study and an explanation for this phenomenon was provided in her following study (Falkum, 2020) that the inferior performance of group of age-4-to-5 can be explained as the result of a literal preference of the children at year 4 and 5.

However, the data collected and analyzed in the present study indicate some different characteristics, compared with the U-shape, concerning children's metonymy comprehension development.

The present study has not found a direct U-shape in children's performance in the picture selection task. However, the author has found a tendency towards the U-shape. Combining the data in Table 2, which show limited difference between age-3 and age-4-to-5 in picture selection, with the data in Table 3, which show a significance difference between the two groups in picture explanation, we can find that children's ability of picture explanation, compared with the accuracy of picture selection, marks a more noticeable improvement with age. Although the present study has not found a U-shape, the author can still argue that the method of picture selection, to a large extent, masks children's ability of pure metonymy comprehension, especially at the stage of year 4 and 5, and that the method of retrospective interview can alleviate this masking influence. In addition, the author could also argue that the alleviation of the masking influence is also due to the fact that the participants knew in advance that they would be asked to explain their choices of pictures, so that they were influenced less by the literal preference in the process of selection.

Furthermore, in the process of the behavioral experiment, the experimenter noticed some phenomena which may stand for the age-4 and -5's "literal preference". For example, in the second

task, when asked “What was boiling, the stadium itself or the spectators?²”, one 4-year-old participant replied: “One player robbed the ball off the defender and ran, and the opposing players wanted to stop him and make his team lose the game.” This is an irrelevant response to the question; however, it denotes a literal preference of the child—though possibly unable to understand the question, he still wanted to say something to show that he is in the conversation. Another example is that when asked “what do people prefer, the food they make or the name McDonald’s itself?³”, one 4-year-old children answered: “My mother took me to Pizza-hut.” The literal preference of the 4-to-5-year-olds may be a significant contributor to the U-shape; however, it is not the only factor.

Another factor that may contribute to the U-shape in previous studies and the tendency to U-shape in the present study is a fake superiority of the performance of the 3-year-olds. In the data collection process of picture selection, the experimenter coded irrelevant choices as “X”s; however, in the data analysis process, these “X”s, together with literal choices, were all entered as wrong answers, which might have been a problem. The author has found that the wrong answers of 3-year-olds contain far more “X”s than other groups’, indicating that the 3-year-olds made their choices largely at chance. Accordingly, it is reasonable that we suspect that the right answers of the 3-year-olds also have a high randomness. In this way, the positive performance of the 3-year-olds might have been faked up by this randomness, and thus the U-shape or the tendency to the U-shape can be observed.

4.3.2. *U-shape explained by eye-tracking*

Slight difference was detected between the 3-year group and the 4- to 5-year group; however, in the eye-tracking experiment, both the heatmaps and the fixation proportion indicate that, from age-3 to age-4-to-5, there is a significant development in children’s metonymy comprehension ability.

The “fake superiority” and the high randomness of the 3-year-olds’ metonymy comprehension can also be explained and thus supported by the data of the eye-tracking experiment. The heatmaps of the 3-year-olds show that the participants’ fixation is highly random between different areas. Although, during the process of the eye-tracking experiment, the experimenter asked the children to choose one picture according to the story they heard and some of them gave correct feedbacks by picking the target pictures, it is still reasonable to argue that their answers were, to a large extent, by chance, which can be explained by their “disordered” fixation heatmaps.

Compared with the heatmaps of the 3-year-olds, age-4 and age-5 participants’ heatmaps show that the irrelevant picture areas are significantly less heated than the target areas and the literal areas, which means that they no more look at the pictures at random and that their metonymy comprehension has started to develop, though possibly masked by a literal preference.

4.3.3. *Age-6 as a crucial stage*

The present study, supported by the data collected, argues that age-6 might be a crucial stage for children’s metonymy comprehension development.

² The whole story and the question were: “There is a football game between China and Japan. In the last minute, the Chinese team scored a winning goal, and the whole stadium was boiling (a metonymy that only makes sense in the Chinese context). So, please tell me, what was boiling, the stadium itself or the spectators?”

³ The whole story and the question were: “McDonald’s and Dicos are two famous fast-food companies. Between them, more people prefer McDonald’s. So, please tell me, what do people prefer, the food McDonald’s make or the name McDonald’s itself?”

The error rate data, experiment score data, and the reaction time data all indicate a significant improvement of the oldest group in metonymy comprehension competence, compared with other groups. In detail, during the picture explanation task, nearly all the participants of the oldest group could clearly point out the relation of contiguity between part and whole, for example, “you are asking about the beard, not the beard itself! Beard is not human being!”. More surprisingly, we attempted to ask the participants of the oldest group to explain their answer in task 2, and some of them replied “The kettle cannot be inside itself. Only the water can, so it is the water that is boiling.” and “Van Gogh would die if you put him on the wall!⁴”, which accurately tell the difference between metonymic meaning and literal meaning.

The eye-tracking heatmaps can also support the argument. Compared with the data of the other two age groups, the data of over-age-6 participants in the eye-tracking experiment generate a more ordered heatmap, in which target pictures received significantly more fixation than other areas. Even concerning metonymies of higher difficulty levels, the heatmaps of the oldest group also show a clear difference and regularity compared with the other two groups.

The factors contributing to the cruciality of the stage are manifold. Children at the age of four are able to construct the relation of similarity between concepts, thus able to comprehend metaphorical expressions (Zheng, 2008). Children's comprehension of metonymy, as has been found in the present study and previous studies (Falkum, 2017; Jiang, 2019; Köder and Falkum, 2020), develops earlier from the age of three. However, due to the poor vocabulary and a lack of social-cultural experience, the metonymy comprehension ability at the age of three is a rather incomplete one. It is after the kindergarten stage (6 years old) that children's vocabulary gets enriched due to primary school education or preschool education (for children ready to attend school in the next year) where children receive more systematic trainings on language. This, to a large extent, activates children's contiguity-comprehending ability, which already exists at the age of 3, and it also helps this metonymy comprehension mechanism to function between a wider range of concepts.

The cruciality of this stage may be also due to the procedural features of children's cognition development. Children's cognition of language, which develops correlatedly with their cognition of the physical objects (Liu and Hong, 2000), shows a tendency from concrete to abstract and goes through a mixed trajectory of continuity and discontinuity (Hudspeth and Pribram, 1990). “Discontinuity” here means that although children's language cognition improves with age, there are still clear phases and stages. Children at the age of 6 finish the stage of literal language development and enters a phase during which children's abstract thinking develops rapidly, which may help to explain the superior performance of the oldest group, and also may shed light on language education for preschoolers.

4.3.4. Comprehension development in different metonymies

Previous studies on children's metonymy comprehension development either investigated children's metonymy competence of one certain type of metonymy (Köder and Falkum, 2020), or stereotyped different types of metonymies as “metonymy” as a whole when designing their experiments (Falkum, 2017; Jiang, 2019). However, against this stereotype, the present study has found that children's competence of comprehending metonymies of different difficulty-levels shows different

⁴ These two examples are about the stories and questions in task 2.

developmental features.

From metonymies concerning human-body part-whole relations, to metonymies of non-body relations, and finally to the cultural-related metonymies, children's error rate, according to Table 2, rises with the increasing difficulty of the metonymies. Furthermore, the third task was where all the age groups answered over half of the questions wrong (error rate: year 6–8, 0.52; year 4–5, 0.64; year 3, 0.86), which forms a contrast with the error rates in other tasks.

Furthermore, heatmaps and figures generated from the eye-tracking data also tell the difference of children's comprehension development in different difficulty levels of metonymies. As for lower difficulty levels, participants' target fixation proportion demonstrates a steady improvement with age from age-3 to over-age-6; however, higher difficulty levels of stimuli give rise to a lower target fixation proportion of all age groups as compared to their performance in less difficult trials, even with the age-3 and age-4-to-5 children giving random fixations in higher difficulty levels of trials.

This result indicates that, even though there is development with age in children's cultural-related metonymy comprehension, children's metonymy comprehension competence not only improves with age, but also can be different between different types of metonymies. As the present study aims to investigate children's pure comprehension development of metonymies, the author hereby argues that children's inferior performance in the third tasks does not imply a low metonymic ability of children, but suggests that children's metonymy comprehension ability can be heavily masked by cultural elements.

5. Conclusion

The goal of the present study was to investigate the developmental trajectory of children's pure metonymy comprehension ability. The focus of this study was on children's metonymy comprehension development with age and children's metonymy comprehension development in different types or difficulty levels of metonymic expressions. Contributive points mainly include the modified behavioral + eye-tracking experiment method and the anti-stereotypical exploration into different types of metonymies when studying children's comprehension.

Data from the modified multi-task behavioral experiment, together with the eye-tracking data, provided innovative insights into how to further explain the U-shape developmental feature as reported in previous research. Concerning the U-shape tendency, the author argues that behavioral experiment design, the literal preference of the age-4 and -5 children, and the randomness of age-3 children's feedbacks jointly contribute to the U-shape tendency observed in the behavioral experiment. The findings of this study also suggested that, when studying children's metonymy comprehension development, the concept of metonymy could not be generalized, since the present study found that the different types and difficulty levels of metonymic expressions may render different stages of comprehension development. Although there are masking effects from cultural elements in higher difficulty levels of metonymic expressions, when children enter the stage of six years old, their metonymy comprehension ability, in both lower and higher difficult levels, shows an obvious improvement compared with other younger age groups, which suggests the cruciality of the stage of age six.

For future studies on children's figurative language comprehension development, this empirical

study would suggest that researchers avoid stereotypical operation of concepts, since different types of a certain figurative device may differ in difficulty, and thus in developmental stages. Furthermore, as the present study has not yet made a thorough enquiry into how different types of metonymies can influence the developmental trajectory of children's metonymy comprehension, the author recommends conducting more detailed investigation into different types of metonymies. Furthermore, researchers who are interested in children's metonymy comprehension are advised to set a wider range of variables in both behavioral and eye-tracking experiments, including, if possible, family backgrounds, personality, and (in eye-tracking) saccade events analysis.

Furthermore, although this study provides insights and implications for further studies, it comes with certain limitations. Due to the pandemic and the relatively short time span of the research, the author did not manage to recruit as many children as expected to participate in the experiments, which might have affected the experiment results. In the next stage, the author still has an ambition to expand or do more on the basis of the present results. The author plans to further improve the experiment instruments and hopefully, to invite more children to participate.

Conflict of interest

No conflict of interest was declared by the author.

References

- Alač M and Coulson S (2004) The man, the key, or the car: Who or what is parked out back. *Cognitive Science Online* 2: 21–34.
- Asch SE and Nerlove H (1960) The development of double function terms in children: An exploratory investigation. In: Kaplan B and Wapner S (eds.) *Perspectives in Psychological Theory: Essays in Honor of Heinz Werner*. New York: International Universities Press, pp.47–60.
- Barcelona A (2000) On the plausibility of claiming a metonymic motivation for conceptual metaphor. In: Barcelona A (ed.) *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin: Mouton de Gruyter, pp.31–58.
- Carston R (2002) *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Corbett EPJ and Connors RJ (1999) *Classical Rhetoric for the Modern Students*. New York: Oxford University Press.
- Dong W (2014) *Metaphor Cognition and Its Educational Application in Children Aged from 3 to 6*. PhD Thesis, Zhejiang University, China.
- EyeLink Data Viewer 4.1.1 [Computer software] (2019) Mississauga, Ontario, Canada: SR Research Ltd.
- Falkum IL (2019) *Metaphor and Metonymy in Acquisition: A Relevance-theoretic Perspective. Relevance, Pragmatics and Interpretation*. Cambridge: Cambridge University Press.
- Falkum IL, Recasens M, and Clark EV (2017) “The moustache sits down first”: On the acquisition of metonymy. *Journal of Child Language* 44(1): 87–119.
- Gardner H (1974) Metaphors and modalities: How children project polar adjectives onto diverse domains. *Child Development* 84–91.
- Hilpert M (2006) *Keep an Eye on the Data: Metonymies and Their Patterns*. New York: Mouton de Gruyter.
- Hudspeth WJ and Pribram KH (1990) Stages of brain and cognitive maturation. *Journal of Educational Psychology* 82(4): 881–884.
- Jiang X (2019) An empirical study of preschoolers' metonymic ability development. *Modern Foreign Languages* (4): 487–500.

- Just MA and Carpenter PA (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4): 329–354.
- Kliegl R, Nuthmann A, and Engbert R (2006) Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General* 135(1): 12–35.
- Köder F and Falkum IL (2020) Children’s metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics* 156: 191–205.
- Lakoff G and Johnson M (1980) *Metaphor We Live by*. Chicago: University of Chicago Press.
- Langacker RW (1993) Reference-point constructions. *Cognitive Linguistics (Includes Cognitive Linguistic Bibliography)* 4(1): 1–38.
- Liu D and Hong X (2000) Study major factors influencing early-childhood language acquisition. *Journal of Chongqing University* 6(1): 103–106.
- Liu Z and Mi X (2008) A study of preschool children’s comprehension of metaphor. *Journal of Sichuan International Studies University* 6: 40–44.
- Lutz A (2017) Norse loans in middle English and their influence on late medieval London English. *Anglia* 135(2): 317–357.
- Mendelsohn E, Robinson S, Gardner H, et al. (1984) Are preschoolers’ renamings intentional category violations? *Developmental Psychology* 20(2): 187.
- Panther KU and Radden G (eds.) (1999) *Metonymy in Language and Thought* (Vol. 4). Amsterdam/Philadelphia: John Benjamins Publishing.
- Panther KU and Thornburg L (2005) The role of conceptual metonymy in meaning construction. In: Ruiz de Mendoza Ibanez FJ and Pena Cervel MS (eds.) *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*. Berlin: Walter de Gruyter GmbH & Co., KG, pp.353–386.
- Peirsman Y and Geeraerts D (2006) Metonymy as a prototypical category. *Cognitive Linguistics* 17(3): 269–316.
- Peng X and Zhang S (2009) Trope ability development: An empirical study of preschool children. *Foreign Language Teaching and Research* 5: 359–364.
- Piaget J (1964) Part I: Cognitive development in children—Piaget development and learning. *Journal of Research in Science Teaching* 40.
- Radden G and Kövecses Z (1999) Towards a theory of metonymy. *Metonymy in Language and Thought* 4: 17–60.
- Rundblad G and Annaz D (2010) Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology* 28(3): 547–563.
- Özcaliskan S (2005) On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity? *Journal of Child Language* 32(2): 291–318.
- Van Herwegen J, Dimitriou D, and Rundblad G (2013) Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities* 34(4): 1300–1311.
- Yan C and Huang M. (2009) On children’s U-shaped development and its inspirations on children’s development assessment. *Early Childhood Education. Educational Sciences General* 33: 1–4.
- Zhang H and Lu W (2010) *Cognitive Metonymy*. Shanghai: Shanghai Foreign Language Education Press.
- Zheng L (2008) *A Study of the Rhetoric Features of Preschool Children’s Language*. PhD Thesis, Nanjing Normal University, China.
- Zhou R (2003) A preliminary study of children’s developmental trend of metaphorical ability of time. *Modern Foreign Languages* 3: 222–231.