Journal of Child Language (2026), **53**, 80–107
doi:10.1017/S0305000924000539

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

An empirical study on native Mandarin-speaking children's metonymy comprehension development

Songqiao Xie[1,2] and Chunyan He[3]

[1]Shanghai International Studies University; [2]University of Cambridge and [3]Key Laboratory of Brain-Machine Intelligence or Information Behaviour, Shanghai International Studies University, Shanghai
Corresponding author: Chunyan He; Email: xiaohe@shisu.edu.cn

(Received 23 April 2023; revised 10 September 2024; accepted 23 September 2024; first published online 13 December 2024)

Abstract
This study investigates Mandarin-speaking children's (age 3–7) comprehension development of novel and conventional metonymy, combining online and offline methods. Both online and offline data show significantly better performances from the oldest group (6-to-7-year-old) and a delayed acquisition of conventional metonymy compared with novel metonymy. However, part of offline data shows no significant difference between adjacent age groups, while the eye-tracking data show a chronological development from age 3–7. Furthermore, in offline tasks, the three-year-old group features a high choice randomness and the four-to-five-year-olds show the longest reaction time. Therefore, we argue that, not only age but also metonymy type can influence metonymy acquisition, and that a lack of socio-cultural experience can be a source of acquisition difficulty for children under six. Methodologically speaking, we believe that online methods should not be considered superior to offline ones as they investigate different aspects of implicit and explicit language comprehension.

Keywords: Mandarin-speaking children; metonymy comprehension; eye tracking

摘要
本研究采用在线与离线相结合的方法，探讨了汉语普通话儿童对新颖转喻和规约转喻的理解能力发展轨迹。在线与离线数据均显示，年龄最大的组别(6-7岁组)表现显著更优，且儿童对新颖转喻的理解早于其对规约转喻的理解。然而，部分离线数据显示相邻年龄组之间无显著差异，而线上数据(眼动追踪实验)则显示，汉语儿童转喻理解能力整体而言随年龄增长(从3岁到7岁)而提高。此外，在离线任务中，3岁组显示出较高的选择随机性，而4-5岁组的反应时最长。由此我们认为，年龄和转喻类型均会影响转喻习得，且6岁以下儿童对社会文化经验的缺乏可能是转喻习得困难的一个因素。在研究方法方面，我们认为，在线方法不应被视为优于离线方法，因为它们分别研究了隐性和显性语言理解的不同方面。

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (http://creativecommons.org/licenses/by-nc/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

1. Introduction

Figurative language frequently occurs not only in adults' but also in children's daily communication, where young interlocutors could show remarkable ability to use metaphorical, metonymical, and ironical expressions. As the understanding of figurative language usually requires a context-driven deviation from literal meanings and a construction of context-dependent meanings, figurative devices, representative of some fundamental ways of human cognition (Lakoff & Johnson, 1980), have served as test-beds for children's development of pragmatic and cognitive competence (Falkum et al., 2017; Köder & Falkum, 2020).

There is a growing consensus in the extant literature that children, on one hand, show an early onset of pragmatic competence in comprehending and even producing figurative language, and on the other, reach adult-like attainment of figurative language at an older age, compared with that of literal language (Cacciari & Padovani, 2012; Pouscoulous, 2011; Pouscoulous & Tomasello, 2020). It is also widely found that children acquire different figurative devices, showing different onsets and following different trajectories, as shown in Table 1. It is worth noting that among these figurative devices, idioms exhibit the latest onset in the acquisition process as a form of non-literal language. Meanwhile, idioms are characterised by a higher degree of lexicalisation and conventionalisation (Cacciari & Padovani, 2012), which might point to children's acquisition of increasing idiomaticity and conventionality of language (Tantucci & Wang, 2020).

However, in the current developmental literature on figurative language, there is an imbalance of interest. On the one hand, due to the higher usage frequency and test amenability, researchers pay primary attention to children's acquisition of metaphor and idiom, followed by irony and proverbs, leaving metonymy an inadequately studied field so far (Cacciari & Padovani, 2012). On the other hand, researchers are more interested in children's language development in relation to cognitive metrics (e.g., ToM & perspective taking), for which irony and novel metaphors serve as popular test beds (Caillies & Le Sourn-Bissaoui, 2008; Clark, 2019; Colston, 2007; Katsos, 2021; Kecskes & Zhang, 2009; Köder & Falkum, 2021; Martín-González et al., 2024).

Furthermore, within the handful of studies on children's metonymy comprehension development, researchers are still making preliminary attempts and are unable to reach an agreement in various aspects: onset of acquisition, developmental trajectory, measurement of comprehension, and categorisation of stimuli (Falkum et al., 2017; Jiang, 2019; Köder & Falkum, 2020; Nerlich, 1999; Rundblad & Annaz, 2010; Van Herwegen et al., 2013).

Aiming to better address the problems identified above, this study, employing a combination of online and offline measures, investigates children's comprehension development of novel and conventional metonymy, a figure of language in which one entity (linguistic/conceptual) is used to refer or provide access to another with which the former is somehow saliently related (Littlemore, 2015) in an associative or contiguous manner. Refer to examples (i), (ii), and (iii):

Table 1. Acquisition age windows for different figurative devices (Colston, 2020)

Hyperbole	Age 3–6
Metaphor	Age 4–8
Verbal irony	Age 3.5–8
Idioms and proverbs	Age 5.0 onward

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

(i) *The moustache* sits down first. (Falkum et al., 2017)
(ii) *The piano* is in a bad mood. (Panther & Thornburg, 2003)
(iii) *The ham sandwich* left without paying. (Nunberg, 1979)

In (i), "the moustache," as a salient (distinctive) feature of a certain person in the context, refers and provides access to the person who wears a moustache. The relation between the moustache and the referred person is from "part to whole." In (ii) and (iii), the sources and targets ("the piano" and "the pianist;" "the ham sandwich" and "the orderer") do not stand in a clear part-whole relation but in a relation that functions between two related and different (aspects of) entities.

1.1. Problems unsolved in extant research on children's metonymy comprehension

Concerning the developmental trajectory of children's metonymy comprehension, extant studies report different findings on whether there is a U-shape/non-linear development or not, with some finding a chronological metonymy comprehension development with age (eye tracking results in Köder & Falkum, 2020; van Herwegen et al., 2013) while others detecting a U-shaped development where the four-to-five-year-old children's performance is surprisingly poorer than the three-year-olds (Falkum et al., 2017; Jiang, 2019; picture selection results in Köder & Falkum, 2020). The U-shaped development in metonymy comprehension has been discussed with U-shaped curves found in the development of other cognitive modalities and also with the literal stage hypothesis (Billow, 1981; Gardner et al., 1975; Pouscoulous, 2011; Winner et al., 1976) in figurative language terms. However, there is to date no adequate and unified explanation for why the development of children's comprehension of metonymy could form a U-shaped curve.

Note also that most of these empirical studies on children's metonymy employ offline methods to measure children's comprehension, for example, forced choice tasks (Nerlich, 1999), picture selection tasks (Falkum et al., 2017; Rundblad & Annaz, 2010; van Herwegen et al., 2013), and verbal explanation tasks (Falkum et al., 2017; Jiang, 2019) (See Table 2). However, as children's conventional lexical knowledge and linguistic ability are still lacking and developing at a young age, it is possible that offline tasks, where children are required to perform verbal responses to examiners, can be so linguistically or pragmatically demanding that children's real comprehension competence gets masked. Furthermore, there is a possible gap between children's explicit responses and their online processing and comprehension. Therefore, studies using online measures (e.g., eye tracking experiment, ERP) for children's real-time metonymy processing are needed.

Table 2. Studies on children's metonymy development

Metonymy studies	Age of Investigation	Method/instrument
Nerlich (1999)	2–5 years old	Forced-choice + stories
Rundblad and Annaz (2010)	5 years old to adulthood	Picture selection
van Herwegen et al. (2013)	3–17 years old	Picture selection
Falkum et al. (2017)	3–5 years old	Picture selection + explanation
Jiang (2019)	3–5 years old	Picture selection + explanation
Köder and Falkum (2020)	3–8 years old	Picture selection + eye-track

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Finally, metonymy types need to be taken into consideration, as current research on children's metonymy either examines metonymy as a broad concept (Nerlich, 1999; Rundblad & Annaz, 2010) or only one specific type of metonymy relation (e.g., human part-whole metonymy) (Falkum et al., 2017; Jiang, 2019; Köder & Falkum, 2020).

In previous attempts to distinguish metonymy from metaphor, the proposal of "internal domain mapping" (Barcelona, 2003; Kövecses, 2002; Zhang & Lu, 2010) and "contiguity relationship" (Kövecses & Radden, 1998; Rundblad & Annaz, 2010) received the highest popularity. However, contiguity being a relatively broad and loose concept, metonymy is a more complex phenomenon with variations in terms of prototypicality (Peirsman & Geeraerts, 2006), conceptual relations (Littlemore, 2015; Radden & Kövecses, 1999), cross-linguistic familiarity (Brdar-Szabó & Brdar, 2003; Slabakova et al., 2016), and novelty or conventionality (Frisson & Pickering, 2007; Schumacher et al., 2023; Slabakova et al., 2013).

The standards above are not mutually independent but nested within each other. For example, as for novelty and conventionality, novel metonymy, which has metonymic readings which are unstable and flexible as context changes (Schumacher et al., 2023; Zheng et al., 2015), can be further categorised into different subtypes contingent on the strength of their link with the prototypical part-whole spatial metonymy according to Peirsman and Geeraerts' (2006) model (more detail in §2.3). Meanwhile, understanding conventional metonymy may require specific socio-cultural knowledge and experience, thus being expectedly more difficult than understanding novel metonymies for young children. Therefore, in order to have a fuller picture of children's metonymy comprehension, it is necessary to look into how children's metonymy comprehension ability develops across novel and conventional metonymy.

1.2. Research questions

To take stock, several researchable gaps and unsolved questions are identified from the current literature: the existence of U-shape metonymy development; an inadequacy of explanation for the U-shape; a need for real-time measures for metonymy comprehension; and a need for inclusion of metonymy type. Therefore, this study aims to investigate Mandarin Chinese-speaking children's metonymy comprehension development, employing both offline (picture selection task and explanation task) and online experimental methods (eye-tracking task), and address the following research questions:

1. How will the results from an eye-tracking experiment agree or disagree with the results from a picture selection experiment (including a picture selection accuracy and an explanation score)?
2. How will the results from the explanation task match with or differ from the results from the picture selection task?
3. Combining the results in both experiments, how can the U-shape attested in previous studies be further explained?

The present study employs a quantitative approach, conducting two phases of experiments (online and offline) where two different groups of children are recruited, and they see similar stimuli during the tasks. The two experiments were conducted to form a comparison that helps to better portray and explain children's metonymy developmental trajectory from a combination of online and offline perspectives.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

2. Experiment 1: Picture selection and explanation task

The picture selection experiment in this study partly follows but modifies the behavioural experiment designs in previous studies (Falkum et al., 2017; Jiang, 2019; Köder & Falkum, 2020) by adding a retrospective explanation task and adding a categorisation to different metonymy stimuli. The added explanation task is intended to provide modification to the offline picture-selection activity, where children's performance can be impacted or masked by the linguistically demanding task itself, which is a problem identified from some extant research on figurative language processing (Di Paola et al., 2020; Levorato & Cacciari, 1992; Pouscoulous & Tomasello, 2020; Rundblad & Annaz, 2010). The experiment was thus carried out as a bi-section activity including 14 trials (cf. more detail below). The metonymy stimuli are assigned into different categories of novelty and conventionality, aiming to investigate children's potentially different comprehension developmental trajectories for different metonymy types.

2.1. Participants

Sixty children, from a monolingual kindergarten in Zhengzhou, China, were invited to participate in the experiment. Written parental consent was obtained prior to the experiment. All the participants were screened to exclude health and intellectual impairments and speak mandarin Chinese as their mother tongue. The children were assigned to three age groups (See Table 3). All the participants managed to complete the trials; however, data from two participants in group 3 were excluded from the final analysis due to background noise and interruptions.

Prior to the experiment, a pilot study was conducted with two three-year-old children to make sure that children at this age and above can understand the words uttered in the recorded stories and are able to recognise and name the items that appear in the pictures.

2.2. Stimuli categorisation standard

Stimuli in the present study are initially classified based on the degree of novelty into two distinct categories: novel metonymy and conventional metonymy.

A novel metonymy (e.g., 'the ham sandwich left without paying') is constructed out of an accidental connection between concepts that have no established or stable metonymic reading yet (Schumacher et al., 2023; Zheng et al., 2015). For novel metonymy, although most types of those metonymy construction (e.g., body-part for person; property for person) are universal across language and culture (Slabakova et al., 2013), the intended reading of specific metonymy is largely contingent upon contextual information. In other words, the link between the "ham sandwich" and the person is salient but ephemeral,

Table 3. Participants

Age Group	Number	Gender (f/m)
Group 1 (Year 3)	15	8/7
Group 2 (Year 4–5)	25	10/15
Group 3 (Year 6–7)	18	10/8

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

susceptible to contextual changes (e.g., "I want the ham sandwich" where the "ham sandwich" refers to the meal set with ham sandwich as the main course).

Conversely, conventional metonymy (e.g., "Dickens" for Dickens' works; "maoyeye" /"Chairman Mao" for banknotes/money in mandarin Chinese contexts) has undergone a process of conventionalisation and has established a stable association between concepts in commonly used language. The figurative meaning of conventional metonymy is experience-based and is largely retrieved from linguistic and world knowledge (e.g., long-term social and cultural memory), but not as largely cognitive-based as novel metonymy, as predictors such as semantic linguistic ability, compared with cognitive-based measures, are found to better predict participants' performance and acquisition of conventional metonymy (Vicente & Falkum, 2023; Zheng et al., 2015).

The conventionality and novelty of metonymy have been measured using various standards and are found to be related to acceptability and effort demanded for processing, with conventional metonymy being less demanding than novel metonymy. Slabakova et al. (2013) distinguished conventional (regular) from novel metonymy based on the "noteworthiness and stability" of the link between concepts and rated place-for-event (e.g., "Chernobyl" for the nuclear accident) and producer-for-product (e.g., "Dickens" for Dickens' works) metonymy as more regular (conventional) than instrument-for-player (e.g., "clarinet" for clarinet player) and loose association metonymy (e.g., the ham sandwich example). Frisson and Pickering (2007) and Schumacher et al. (2023) proposed familiarity as a key proxy for determining the conventionality and thus the processing effort required for metonymy. For example, within the construction of producer-for-product metonymy, "I never read Dickens" would be more conventionally metonymic and more readily accessible than "I never read Needham" where "Needham" is not a familiar name for an author (Frisson & Pickering, 2007).

To take stock, there is more of a continuum than of a dichotomy from novelty to conventionality. Highly novel metonymy, found towards one end of the continuum, typically features a contextually contingent (unstable) link, lacks idiomaticity or lexicalisation, and lacks central representativeness.[1] Moving towards the conventional end, metonymies exhibit figurative readings that are stabler across contexts, especially, for example, the language- or culture-specific idiomatic metonymy constructions (Nunberg, 1995) (e.g., "one mouth" in a family referring to a member who lives and consumes food in the family).

As "stability of metonymy link" (Slabakova et al., 2013) is difficult to measure, the present study distinguishes between novel and conventional metonymy by labelling the idiomaticity of metonymy and checking the frequency of metonymy readings of expressions in the CCL (Center for Chinese Linguistics Peking University) corpus (See Table 4). For example, for conventional metonymy, we include idiomatic and cultural-specific metonymy expressions in the Mandarin Chinese context and expressions that are frequently taken metonymically.

Taking one step forward, the present study further subcategorizes novel metonymy according to Peirsman and Geeraerts (2006), who provided the prototypical classification model for metonymy, based on their conceptualisation of the relation of contiguity. In their model, the strength of contact (the distance between the salient property and the

[1]By "central representativeness," we mean that the ad-hoc link between concepts in highly novel metonymy does not usually touch upon the defining feature of the referent, as many (but not all) conventional metonymies (and some not-so-highly novel metonymies) do. For example, (producing) books represent(s) the core identity of the author as a profession, while the ham sandwich would not.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Table 4. Frequency of metonymical/literal readings of conventional metonymy in CCL

Conventional metonymy	Metonymical reading	Literal reading	Total
喝白的 Drink white	26	0	26
露两手 Display two hands	72	0	72
一张嘴 One mouth	1245	286	1531
毛爷爷 Grandpa Mao	279	417	696

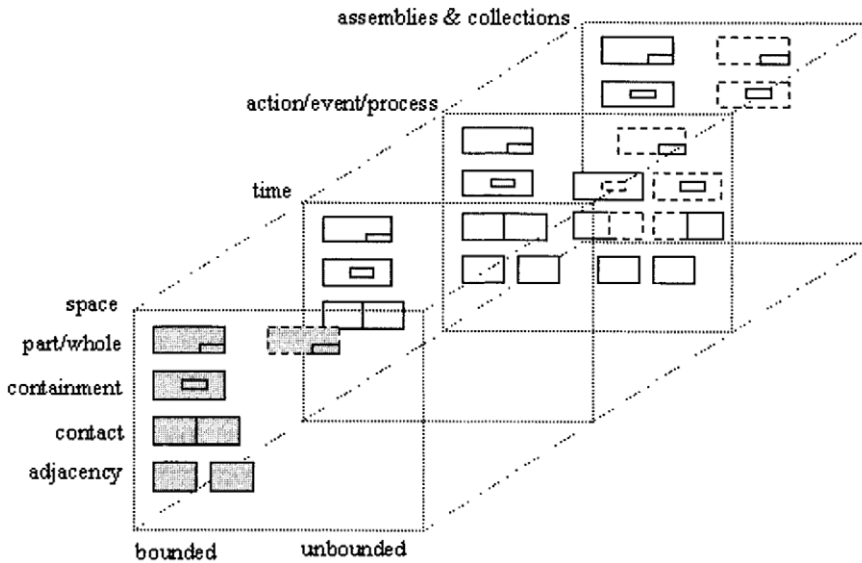


Figure 1. Prototypical category of metonymical patterns. (Peirsman & Geeraerts, 2006, p. 310)

referred target), boundedness (whether there is an ontological border between the salient property and the referred target), and domain type (spatial, temporal or categorical) together form the metonymy category, where a rich set of metonymies are categorised and linked in terms of the type of contiguity they are motivated by. Refer to Peirsman and Geeraerts' (2006) model in Figure 1:

In Peirsman and Geeraerts' (2006) model, there are three dimensions indicating three standards by which various metonymies can be linked with the core category, spatial part-whole metonymy. In the utterance "the big beard walked towards him," the "big beard" itself is a physical part of and spatially linked to the person who wears it.

Along each dimension line, the prototypicality of metonymy becomes weaker. In a sense, various types of metonymy are able to find their places on the proposed three-dimension model, which classifies them by their degrees of prototypicality.

Therefore, the stimuli of this study consist of novel and conventional metonymy; and within the novel type, metonymy stimuli are further categorised into three subtypes based on children's animacy preference (Piaget, 1964) and Peirsman and Geeraerts' (2006) model. As we hypothesise that (1) human-related metonymies are easier for children and that (2) prototypical metonymies are easier than non-prototypical ones in metonymy-

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

motivating contexts, the hypothesised relative difficulty of novel metonymy subtypes is: human-related part-whole (e.g., big ears – the person with big ears) < human-related adjacency (e.g., white trousers – the person wearing white trousers) < non-human-related feature (e.g., watermelon – a cup of red juice made from watermelon).

2.3. Stimuli and procedure

The fourteen trials in this experiment consist of one training trial, ten trials with metonymy target utterances, two trials with literal target utterances, and one distracting trial.

The first training trial helps children to familiarise themselves with the experimental procedure. Data collected in this trial are excluded from the final analysis.

The ten metonymy trials are divided into two categories (novel and conventional), with four subcategories (1 sub-category under conventional, 3 sub-categories under novel) in total. Among the 10 trials, 4 trials contain conventional metonymy stimuli, while 6 contain novel stimuli (2 trials under each sub-category).

The two literal trials serve to measure children's ability to understand the literal meaning of the stories. As it is assumed that three-year-olds should grasp the meaning of these items, they can also signal whether children might select a picture at chance, especially for three-year-olds. The distractor trial presents an utterance that makes no sense. It is put in the middle of the trials to avoid the learning effect.

All the 14 trials are identical in structure and similar in length. In each trial, the participant listens to a pre-recorded story lasting for about 20 seconds, while looking at a set of four pictures (as in Figure 2 below). The background picture on the top of the screen gives participants a general idea of "what" with regards to "whom". The three option pictures in metonymy trials contain one metonymy picture (the right answer,

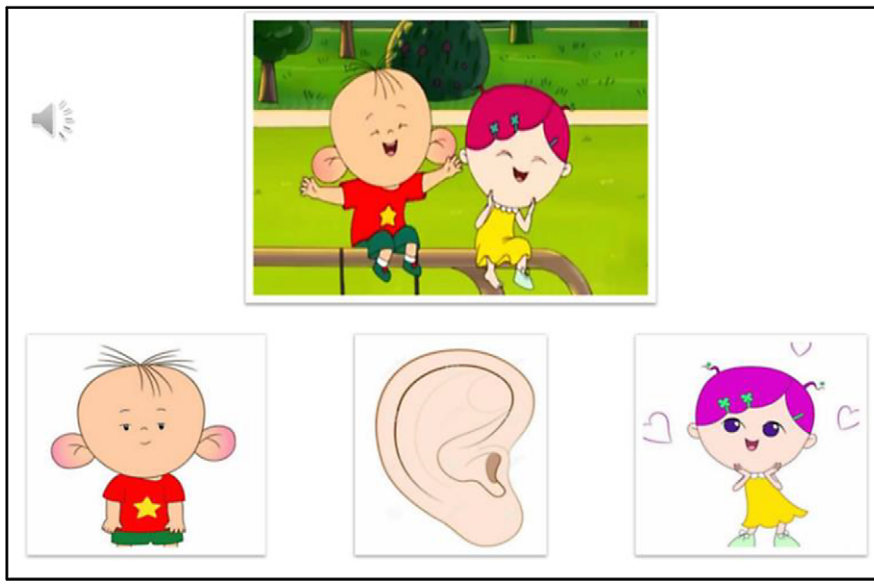


Figure 2. Example stimulus

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

i.e., protagonist with big ears), one literal picture (the object, i.e., the "big ear" itself), and the irrelevant picture (the other person from the background picture). Each story is composed of one context utterance and one target utterance. Each target utterance contains a referential expression of the metonymy target ("big ear") and a descriptive expression ("always gets criticized by teachers…") which helps with the question that follows.

Following the content of the story, a question for instruction unfolds. The questions in all the trials would use "*shenme*" (which) instead of "*shei*" (who) at the beginning, to avoid causing biases to participants' choices. One example of the whole 20-second audio recording in one trial is shown below.

Example 1:

上面这幅图的两个小朋友在同一所幼儿园上学。可是，**大耳朵**因为淘气经常被老师批评。 请问小朋友, 下面三幅图, 是哪个经常挨批评?

"These two children go to the same school. However, **big ear** always gets criticized by teachers for being naughty. So, young boy/girl, can you please tell me, which of the following three gets criticized very often at school?"

Then, one follow-up question ("Could you tell me why you chose this?") is asked to collect data concerning the comprehension process of the participant. Children's feedback is recorded and transcribed for analysis of the explanation task.

2.4. Coding

The participants' answers to the forced-choice questions are treated as a categorical variable. All correct selections are coded as "1;" all incorrect selections are coded as "0".

Concerning the "explanation score" in metonymy trials, children's explanations of their correct selections fall into three categories (refer to Table 5 and Example 1):

To check inter-rater agreement, the two authors coded 100 explanations randomly selected from the children's explanations. They were unaware of the child's age when

Table 5. Three categories of children's explanations

Categories	Possible cases	Sample explanations	Coding
Relevant	Making the association between person/item and the salient feature explicit	*Because the boy has big ears*	2
	Substituting the metonymical for the literal referent	*Because the boy is naughty*	
Semi-relevant	Repetition of the metonym	*Because that's the big ear*	1
	Paraphrase of the target sentence		
Irrelevant	Uninformative response	*Because it's big*	0
	No response		

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

coding the answers (κ = .88). According to Landis and Koch (1977), a κ value above .81 indicates almost perfect agreement.

Concerning the "reaction time," it starts counting once the target expression (e.g., "big ear") unfolds and stops when the participant presses the button to give feedback. To guarantee the validity and accuracy of the timing, the time span of audio recording (8000 ms) from the target expression to the end of the question is strictly controlled, which means all the figures taken down should be greater than 8000ms, except for cases of false triggering or rushing to answer. Also, data analysis only contains the reaction time from metonymy trials where participants select the correct pictures.

2.5. Results

Data analyses were conducted using the SPSS version 26 statistical package (SPSS 26.0). We used three performance measures in the study: selection accuracy, explanation score, and reaction time. We ran generalised linear mixed models (GLMMs) on metonymy trial accuracy and explanation score, and one-way ANOVAs on literal trial accuracy and reaction time, including age (3 age groups) as the between-subject variable and condition (metonymy types) as the within-subject variable. All models were fitted with fixed effects structure and random effects structure including by-subject and by-item random intercepts. We report relevant F and *p*-values below, as well as standard errors (*SE*) and *t*-statistics for pairwise comparison tests.

The data were analysed as follows.

First, we examined metonymy trial accuracy for each age group in the selection task. Then, in order to avoid the possible over-estimation of metonymy trial performance, three more analyses were carried out: participants' performance in literal trials; times of unrelated choices in metonymy trials; and participants' reaction time, which is a relative measuring of time for processing the target metonymy expression.

Second, we investigated the explanation score for each age group in the selection task. To help with solving and explaining the discrepancy between metonymy trial accuracy and explanation score, an ANOVA analysis of explanation-to-choice ratio was conducted.

Third, the within-subjects variable, metonymy type, was added to the analysis. As the most revealing measure between different age groups, the explanation score was chosen to be the measure in the analysis of the influence of metonymy types on children's metonymy comprehension.

2.5.1. Selection task

The table below shows the mean selection scores in literal trials, selection accuracy, and explanation scores in both novel and conventional metonymy trials.

Table 6. Mean accuracy/score in different types of trials in 3 age groups

Age	Number	Literal	Selection novel	Selection conventional	Expl score novel	Expl score conventional
Age 3	15	1.4/2	53.30%	40%	2.53/12	1.33/8
Age 4–5	25	1.92/2	59.30%	49%	5.8/12	2.44/8
Age 6–7	18	1.97/2	81.50%	51.50%	7.94/12	3.61/8

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

2.5.1.1. Metonymy trial accuracy. We analysed the correct responses in metonymy trials using mixed logit modelling with the function of GLMM in SPSS 26.0. Our final model included as fixed effect Age (as categorical variable) and as random effects the random intercepts for subjects.

The overall model thus reveals differences in selection accuracy in Age, $F(2,577) = 6.472$, $p = .002$, which suggests that metonymy comprehension ability in picture selection grows with age. A pairwise test using Bonferroni contrasts reveals that this effect is due to significant differences between the youngest group (age-3 group) and the oldest group (age 6–7 group) ($SE = .14$, $t = 2.07$, $p < .05$). No significant differences are observed between age-3 group and age 4–5 group ($SE = .09$, $t = 1.00$, $p = .319$) or between the two older groups ($SE = .13$, $t = 1.56$, $p = .119$), namely age 4–5 and age 6–7.

2.5.1.2. Literal performance and unrelated choices. Choosing the wrong answers in literal trials and choosing unrelated choices in metonymy trials can be interpreted not only as failure to comprehend but also as selection at chance, which would cause the metonymy trial performance to be over-estimated in the selection task. Thus, to test this hypothesis, one-way ANOVA between groups was carried out.

It is clear from the first plot that participants in the youngest age group often face difficulty in choosing the correct pictures and explaining their choices in literal trials (full score = 2), while the other two age groups almost perform at ceiling level. From the second plot, it could be almost all the age-3 participants selected unrelated options in some trials in the selection task, which, together with their relatively poor performance in literal trials, further suggests that the metonymy trial performance of the youngest group might be largely over-estimated.

2.5.1.3. Reaction time. In terms of reaction time, the oldest age group is able to give correct answers in the quickest manner in selection tasks; however, between the two younger groups, the reaction time does not become shorter with the increase of age – instead, the youngest group even tends to give slightly quicker feedback than the age 4–5 group, although this difference is not statistically significant ($p = .605$). The reaction time difference between the two older groups is statistically significant ($p < .05$). Among

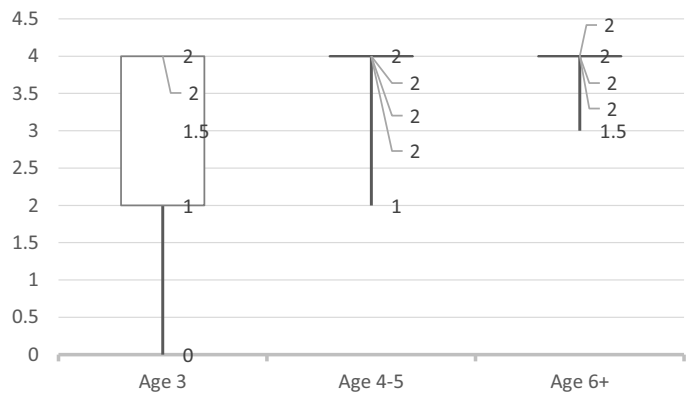


Figure 3. Literal trial performance between age groups.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

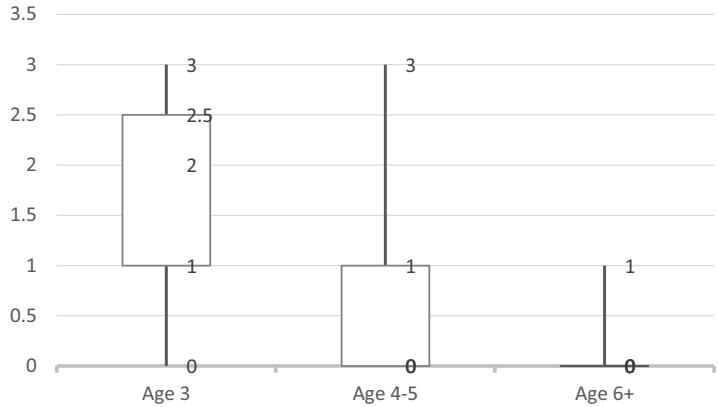


Figure 4. Times of unrelated choice between age groups.

Table 7. Average reaction time in 3 age groups

Group	Number	Reaction time	
		Mean	Variance
Age 3	15	12.14	10.04
Age 4–5	25	12.46	17.91
Age 6–7	18	10.94	9.51

the three age groups, the age 4–5 group provides reaction time data with the greatest variance, suggesting a more discrete distribution of data (See Table 7).

2.5.2 Explanation task

Table 6 and Figure 5 present the data of explanation score (full score = 20) in different age groups. A GLMM analysis on explanation score has been done. We were left with a model including the main effect of Age (as categorical variable).

The overall model thus reveals significant differences in explanation score in Age, $F(2,577) = 24.248$, $p < .001$. A pairwise test reveals that this effect is due to significant differences between age-3 group and age 4–5 group on the one hand ($SE = .10$, $t = -3.87$, $p < .001$), and between age 4–5 group and age 6–7 group on the other ($SE = .10$, $t = -3.78$, $p < .001$), which forms a discrepancy with the selection results.

To help with solving and explaining the discrepancy between metonymy trial selection accuracy and explanation score, an ANOVA analysis of explanation-to-choice ratio was conducted, with selection accuracy temporarily transformed from categorical data to numerical data. The explanation-to-choice ratio can be revealing for the following reasons: Ideally, if a participant gets all the trials correct in both tasks, the ratio would be 2 (20/10); however, in real situations, if the ratio calculated to be greater than 2, it would mean that there must be cases where the participant selects the incorrect picture in the

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

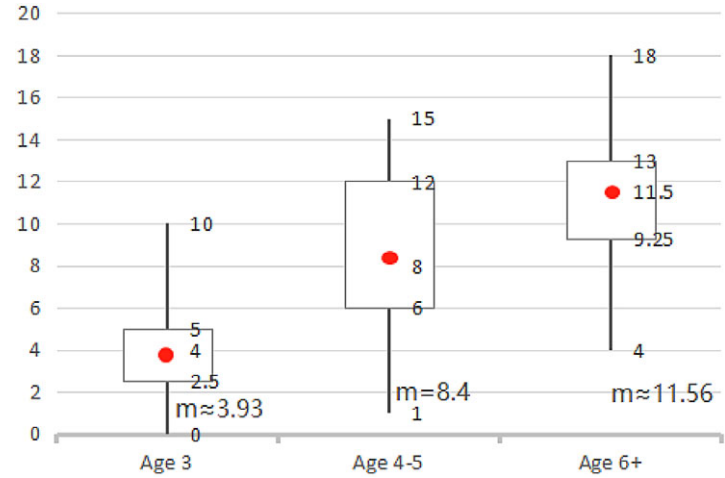


Figure 5. Metonymy explanation score in different age groups.

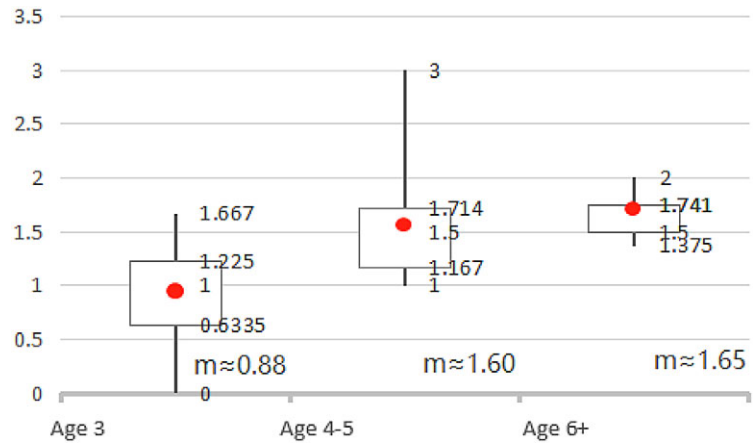


Figure 6. Explanation/choice score in different age groups.

selection task while gives the correct response in the explanation task, which would suggest that the child might actually understand the metonymy but just does not choose the correct picture for certain reasons; if the ratio is lower than 1, it would mean that there must be cases where the participant selects the correct picture while gives no response or zero-point response in the explanation task, which would suggest that the child might be unable to explain his or her choice or that the child chooses the right answer only by chance.

The boxplot above presents the explanation-to-choice score in different age groups. The mean ratio grows as age increases, but the difference is only significant between two younger groups ($F = 16.669$, $p < .001$). As is noteworthy, the ratio of the youngest group is below 1 (the median), which suggests that it might be a common practice for the three-

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

year-old participants to gain points in selection tasks by selecting pictures at chance. Among the data of age 4–5 group, there are figures greater than 2, which suggests that there might be factors masking these children's metonymy comprehension ability and preventing them from choosing the metonymy option they actually understand in the selection task.

2.5.3. Influence of metonymy type

Since literal trial accuracy, times of unrelated choices, reaction time, and explanation-to-choice ratio jointly suggested that explanation score was the most reliable measure between different age groups, it was chosen in the analysis of the influence of metonymy types on children's metonymy comprehension. We conducted two analyses of GLMM on explanation score, one for 2-type categorisation of difficulty and one for 4-type categorisation of difficulty. In each analysis, Age was the between-subject variable, and Condition (metonymy types) was the within-subject variable.

Our first analysis with Age (3 age groups) and Condition (novel, conventional) yields a main effect of Age, such that the explanation score increases with age, $F(2,576) = 22.173$, $p < .001$; and a main effect of Condition, such that there is a clear drop in explanation score from novel metonymies to conventional metonymies, $F(1,576) = 25.190$, $p < .001$ (See Table 8).

To have a direct combining view of the influence from age and the influence from metonymy type, a pairwise test was conducted concerning the explanation score between different age groups.

In terms of novel metonymy, the explanation score difference by age is more significant between the two younger groups ($SE = .14$, $t = -3.86$, $p < .001$) than between the two older groups ($SE = .13$, $t = -2.73$, $p < .05$); while in terms of conventional metonymy, the difference is more significant between the two older groups ($SE = .13$, $t = -2.33$, $p = .021 < .05$) than between the two younger groups ($SE = .13$, $t = -2.08$, $p = .039 < .05$).

In the second analysis, metonymy stimuli were divided into four types according to different pre-set difficulty levels. Our final model included as fixed effects Age (3 age groups) and Condition (4 metonymy types) and as random effects the random intercepts for subjects.

The overall model thus reveals differences in explanation score in Age, $F(2,574) = 22.173$, $p < .001$, and Condition, $F(3,574) = 18.711$, $p < .001$. The interaction of Age by Condition is non-significant, $F(6,574) = 1.048$, $p = .393$, which indicates that children of different age groups exhibit similar developmental trajectory in comprehension of four types of metonymy. Explanation scores are calculated for each participant and the mean scores of the four metonymy types are shown in Table 9.

Table 8. Explanation score of different metonymy types in three age groups

Group	Number	Novel metonymy		Conventional metonymy	
		Mean	SD	Mean	SD
Age 3	15	.43	.56	.33	.55
Age 4–5	25	.97	.55	.61	.55
Age 6–7	18	1.32	.55	.90	.55

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Table 9. Explanation score of different metonymy types in three age groups

			95% Confidence interval	
Types	Mean	Std. error	Lower	Upper
1	.691	.538	−.366	1.748
2	.863	.538	−.194	1.921
3	1.191	.538	.134	2.248
4	.605	.536	−.449	1.658

At this stage, the previously hypothesised relative difficulty of novel metonymy subtypes turns out to be wrong. These results suggest that the difficulty ranking of the 4 metonymy types should be: novel non-human-related feature metonymy (type 3) < novel human-related adjacency metonymy (type 2) < novel human-related part-hole metonymy (type 1) < conventional metonymy (type 4). Pairwise comparison reveals that the main effect of Condition is due to significant differences between type 1 and 3, type 2 and 3, type 2 and 4, type 3 and 4; $p < .001$.

3. Experiment 2: Eye-tracking

The eye-tracking experiment was conducted in the cognitive laboratory of Shanghai International Studies University. The participants' eye movements were tracked by Eyelink-1000 plus. Data collected in this experiment were used to analyse 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 analysed on SPSS 26.0.

3.1. Participants

Thirty-six children ranged from 3 to 7 years old were tested in this experiment (See Table 10). All participants were native mandarin Chinese speakers screened to exclude health, vision and intellectual impairments. Children were recruited from kindergartens and primary schools in Songjiang District, Shanghai. Written parental consent was obtained prior to the experiment.

3.2. Stimuli and procedure

The stimuli contain twelve trails, the first two being training trials. The rest consist of 7 metonymy trials, 2 literal trials, and 1 distracting trial. The reason why the

Table 10. Participants

Age group	Mean age	Number
Year 3	3.48	9
Year 6–7	5.08	14

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

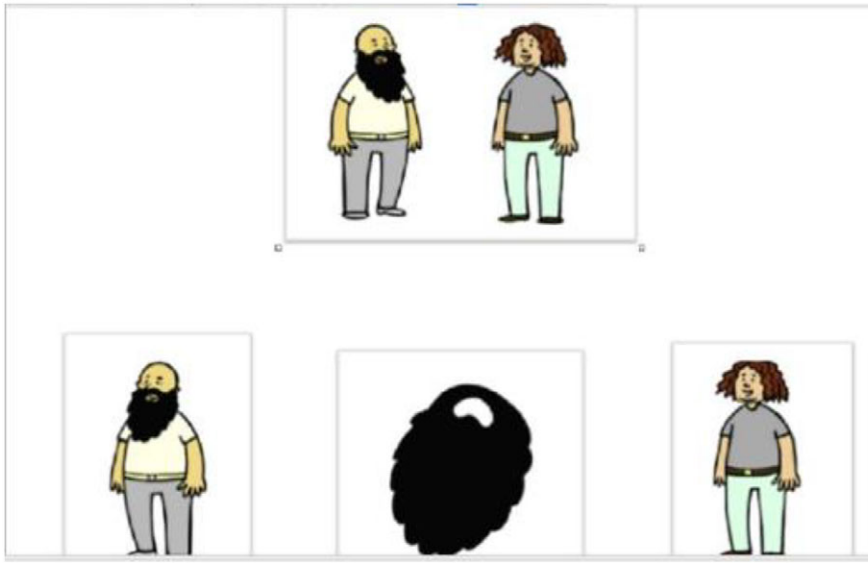


Figure 7. Example stimulus.

metonymy trials (7 trials) in eye tracking experiment are fewer than those (10 trials) in the picture selection experiment is that, on the one hand, child participants of the eye tracking experiment need to complete the task alone in the lab and thus need more training trials; on the other hand, children doing the task alone easily get distracted or feel bored, so, the time span of the task needs to be shorter. To ensure the comparability of the data from two experiments, we make sure that the 7 metonymy trials in the eye tracking experiment come from the 10 metonymy trials in the picture selection experiment.

Similar to the picture selection experiment, the seven metonymy trials are divided into two categories, novel metonymy (4 trials) and conventional metonymy (3 trials).

In each trial, the stimuli contain four pictures[2] placed on four areas of the screen (See Figure 7). The largest picture on the upper half of the screen is the background picture; the three pictures on the lower half are option pictures. The participants are asked to choose one, by pointing with finger, according to the 20-second audio story.

3.3. Coding

Proportion of fixation is calculated by the number of fixations on target areas divided by the number of all fixations that take place during the selected period of time. The proportions of fixation before and after the target utterance are calculated separately so

[2]The set of pictures are shown to the participant 4000ms earlier than the audio is displayed, which allows the participants enough time to recognize and be familiar with the elements in the pictures.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

that they can form a comparison which may show the changes in participants' fixation patterns triggered by the target utterance.

3.4. Results

At the first step, GLMM was carried out between the three age groups concerning the proportion of fixation on metonymy target areas after the target utterance. The fixation proportion after the target utterance can directly reflect children's online processing of the target metonymy.

Following this, the variable of metonymy type was entered into the analysis. Two GLMMs were done in each age group concerning the change of fixation proportion before and after the target utterance in two metonymy types (novel vs. conventional). The difference between pre-target fixation proportion and post-target proportion shows how the target utterance works to trigger the changes in participants' pattern of looks. If there is not an increase in the fixation proportion after the target utterance, it is reasonable to say that the participant's eye movement is not clearly triggered by the utterance, and thus he or she might not understand the metonymic meaning the utterance carries.

The final analysis took a more detailed look at the data in seven individual trials so as to better explore how metonymy type influences children's fixation pattern. One-way ANOVA analysis was conducted in each age group concerning the comparison between pre-target and post-target fixation proportion in different trials.

3.4.1. Fixation proportion between age groups

In each trial, as the three option pictures on the screen are equal in area and the background picture is 3 times as large as the option picture, the fixation proportion of random looks should be the ratio of the area of one option picture to the area of all pictures, which is around 16% (See Table 11).

No main effect of Age is observed in the fixation proportions before target utterance, $F(2,249) = 2.198$, $p = .113$. The means are around 16% in all age groups, which indicates that the participants scanned the pictures on the screen in a random manner before the target utterance unfolded.

However, significant group differences are found in the fixation proportions after target utterance, $F(2,249) = 6.029$, $p < .05$. The means in all age groups are higher than 16%, which suggests that all groups show a metonymy bias triggered by the metonymy target utterance. Pairwise comparison reveals that the difference between age groups is statistically significant between the two younger groups ($SE = .02$, $t = -2.67$,

Table 11. Fixation proportion before and after target expression in three age groups

Group	Number	Fixation proportion before TE		Fixation proportion after TE	
		Mean	SD	Mean	SD
Age 3	9	.15	.07	.25	.09
Age 4–5	13	.16	.07	.32	.09
Age 6–7	14	.18	.07	.33	.09

TE: Target expression

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

$p < .05$), and even greater between the youngest group and the oldest group ($SE = .02$, $t = -3.38$, $p < .001$). As for the two older groups, the difference is not significant ($SE = .02$, $t = -.74$, $p = .458$).

3.4.2. Influence of metonymy type

In this section, we focus on the difference between pre-target fixation proportion and post-target proportion. If the difference is positive (post-target > pre-target) and significant, it suggests that children's fixation is stimulated by their understanding of the target expression.

Two analyses of GLMM were done, one for novel metonymy and one for conventional metonymy. In each analysis, Age was the between-subject variable. No main effect of Age is found in novel metonymy, $F(2,141) = 1.101$, $p = .335$; while in conventional metonymy, age difference is significant, $F(2,105) = 3.347$, $p < .05$. Pairwise comparison reveals that the main effect of Age in conventional metonymy comprehension is due to significant differences between the youngest group and the oldest group ($SE = .05$, $t = -2.53$, $p < .05$) and marginally significant differences between the two youngest groups ($SE = .05$, $t = -1.95$, $p = .054$).

However, to better look at how metonymy type influences children's fixation pattern, we need to take a closer look at the data in different trials (See Table 12).

In age-3 group, participants show an increase of fixation proportion after target utterance in all of the four novel metonymy trials and the increases are statistically significant in three of the four trials. On the other hand, in all the conventional metonymy trials, the fixation pattern of age-3 participants does not show a significant difference between pre-target and post-target fixation proportion.

In age 4–5 group, participants show an even clearer and more significant increase of fixation proportion in all the novel metonymy trials. In terms of conventional metonymy trials, the fixation pattern of participants at this age begins to show significant increases in two of the three trials.

Table 12. P-value in all trials in three age groups

Age groups	Trial type	Novel metonymy				Conventional metonymy		
Age 3	Trial number	5	8	9	12	7	10	11
	Mean difference	0.10	0.14	0.13	0.10	0.09	0.00	0.10
	P-value	0.0005	0.006	0.005	0.0761	0.11	0.97	0.06
Age 4–5	Trial number	5	8	9	12	7	10	11
	Mean difference	0.19	0.19	0.13	0.15	0.18	0.21	0.07
	P-value	3.13E−05	4.09E−05	0.001	0.004	0.001	3.55E−05	0.270
Age 6–7	Trial number	5	8	9	12	7	10	11
	Mean difference	0.12	0.13	0.14	0.17	0.20	0.19	0.14
	P-value	2.5E−05	0.0001	8.89E−08	2.24E−05	1.3E−05	6.66E−05	0.0003

*Mean Difference = mean post-target proportion − mean pre-target proportion
*The underlined figures are those greater than 0.05 (based on the threshold $p<0.05$ for significance)

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

In age 6–7 group, participants show significant increase in both types of trials. Compared with the two younger groups, participants of the age 6–7 group show a clear metonymy bias when faced with tasks of conventional metonymy in all the three trials, which might suggest that it is at the age of six that children really begin to show a good command of conventional metonymies.

In general, in novel metonymy trials, although all the three age groups show significant or marginally significant increase in fixation proportion after target expression, there is a difference in the degree of increase between different age groups: the increase is clearer between the two younger groups. On the one hand, the age-3 participants still have problems dealing with novel metonymy stimuli, receiving less triggering influence from target utterances, and thus showing smaller mean difference of fixation proportion and greater *p*-values; on the other, the age 4–5 children show a clearer increase in all the four trials, which suggests their good comprehension of novel metonymy stimuli, and their fixation pattern is similar to that of the oldest group.

As for the conventional metonymy trials, the clearer increase occurs between the two older groups. The fixation pattern of the age 4–5 participants shows that they still have difficulty comprehending some conventional metonymies, as the youngest group does in all the three trials; while the age 6–7 group's fixation proportion shows that they have a good command of understanding conventional metonymies.

4. General discussion

4.1. Summary of results

In terms of children's developmental trajectory of metonymy comprehension, the three tasks yield different but mutually explainable results. Our results, regardless of the influence of metonymy type, do not show a U-shape in any clear form or shape as some previous studies did, but show that (1) age-3 children's performance in the selection task is more random than the other two age groups, selecting unrelated choices and making mistakes even in literal trials; (2) age 4–5 participants' selection accuracy is not significantly higher than their age-3 counterparts but (3) their performance in the explanation task and eye tracking task is significantly better; (4) in terms of reaction time, age 4–5 children need the most time for processing before choosing a picture in the selection task, which forms a slight U-shaped developmental trend; (5) children older than 6 years perform significantly better under almost all the measurements, compared with the other two groups.

The present study takes into account not only the factor of age but also the possible influence from metonymy type on children's acquisition of metonymy, as most of the extant empirical studies on metonymy development only focus on the time course (age) of development, using novel metonymy as stimuli. Experimental stimuli in the present study consist of novel metonymy and conventional metonymy, and novel type is further divided into three sub-categories with different difficulty levels. However, in terms of the influence of metonymy type, our results of the three-sub-type metonymy categorisation turn out to go against the hypothesis, as children's performance in the offline experiment (selection task and explanation task) does not decline as difficulty increases.

We then merged novel sub-types and compared novel metonymy with conventional metonymy, adopting a two-type categorisation. The results from both experiments show that children begin to understand conventional metonymy significantly later than they show a decent comprehension capacity with novel metonymy, and the results from

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

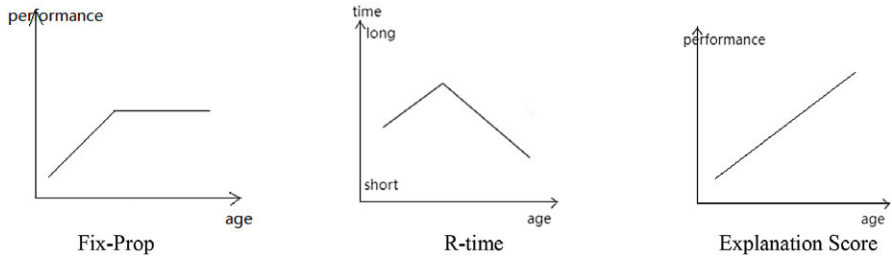


Figure 8. Developmental trajectory found under different measurements.

explanation task and eye tracking task show that it is only after six years old that children show a good command of conventional metonymy comprehension.

In combination, the three tasks in the two experiments yield different results in terms of age trajectory, but similar and mutually explainable results in terms of metonymy type influence. For example, the developmental trajectory yielded from the three tasks are different in shape, as is shown in Figure 8 (e.g., chronological development suggested by explanation score; U-shaped (reversed) trend suggested by reaction time; rapid development in metonymy comprehension at age 4–5 suggested by eye-tracking task); children are able to understand and explain novel metonymy at a younger stage than conventional metonymy; conventional metonymy acquisition starts at around school age (6–7 years old).

In relation to some extant developmental studies on metonymy (Falkum et al., 2017; Köder & Falkum, 2020), the present study only finds a U-shape in reaction time (by mean) – there is no U-shape detected in other measures in selection, explanation and eye-tracking tasks. In general, most of the measures in this study presents children's chronological development with age in metonymy comprehension, but with different rate of development at different stages. In response to the on-going debate between the early onset hypothesis versus the literal stage hypothesis, the results of our study suggest an early onset of novel metonymy before the age of 4–5, and also show that there is no literal stage in children's metonymy comprehension development. As for conventional metonymy, children's onset of comprehension is delayed until 6 years old.

Although the present study does not find a U-shaped development with age in the picture selection score, our results can provide further explanation for the U-shape developmental trajectory which was detected in previous studies (Falkum et al., 2017; Köder & Falkum, 2020).

4.2. High selection randomness of age-3 group

Combining the findings from selection task with explanation task, and also taking into account the three-year-olds' performance in literal trials, the present study argues that three-year-old children's performance can be highly random and thus be overestimated in picture selection tasks which were frequently conducted in previous studies.

As is shown in Figures 3–4, the youngest group still make mistakes in literal trials and choose unrelated choices significantly more often than the other two groups, suggesting that, although their selection accuracy is significantly above chance, their choices are largely random. In this way, the picture selection performance of the three-year-olds could be overestimated.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Furthermore, in Figure 6 where the ratio of explanation score to selection score is calculated, the mean (0.8832) and the distribution (most being lower than 1) of the ratio of the youngest group mean that there must be cases where three-year-old children choose the correct picture but are unable to explain their choices. This can be caused by two reasons: (1) the three-year-old children get some of their choices right by chance; (2) the three-year-old children's linguistic ability is still at a relatively low stage of development and thus they are unable to give decent verbal responses in the explanation task. The first reason, which aligns with the high selection randomness discussed above, would also provide further explanation for the U-shape detected in previous studies.

The U-shaped development in previous studies consists of three developmental periods: a good performance of the youngest group, a declined performance of the second youngest group, and a steady improvement of the oldest group. The present study suggests that the high selection randomness of the three-year-olds can, at least, partly explain why a U-shape could possibly occur in previous studies, especially why the youngest group could perform surprisingly well in the selection tasks.

However, as no U-shape is found in most of the measures in the present study, there is a need to explain the discrepancy in findings between the present study and previous studies. From the view of the authors, the discrepancy can be explained, from a methodological perspective, by the unique task design of picture selection experiment. The picture selection experiment in the present study consists of two sections (selection task and explanation task). The two tasks are not done totally separately but in an interweaving and interactive manner – participants, in each trial, needs to give a verbal explanation immediately after they make a choice. In this way, the participants will possibly be more cautious with their choice and their selection accuracy might incline towards the explanation score. Indeed, compared with the U-shape in previous studies, the trajectory found in selection accuracy in the present study is more similar to the explanation score developmental trajectory. In other words, although the present study finds and argues that three-year-olds can be highly random in choosing pictures, the participants of the youngest group in our study, influenced by the "tight" task design, are not so "random" as they could possibly be.

Also, as the present study takes different types of metonymy into consideration, instead of merely including novel metonymies as done in previous studies, it is possible that children's comprehension of conventional metonymy and of novel metonymy follows different trajectories, and that the criss-cross of the two trajectories can conceal the U-shape in the general finding. For example, the three-year-olds show an almost chance-level selection performance in conventional metonymy trials (refer to Table 6) but they are almost unable to give explanations for their choices in these trials, which, to some extent, pulls down their overall performance in the picture selection and explanation tasks.

4.3. Performance of age 4–5 weakened in selection task

Comparing the findings between eye tracking experiment and picture selection experiment, and also between the selection task and the explanation task, the present study argues four-to-five-year-old children's performance in the selection task can be underestimated and thus their real metonymy comprehension ability can be masked due to methodological reasons. Also, this argument can provide further explanations for the U-shaped curve (the decline of the four-to-five-year-olds) detected in previous studies,

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

although we do not find a clear drop (only reaction time data show a slight drop by mean) of performance among them.

Compared with the findings from eye tracking experiment where the age 4–5 group shows a significant improvement from the youngest group in both types of metonymy, especially in novel metonymy, the findings from the selection task suggest that the performance of the age 4–5 group is weakened as there is no significant difference between the two younger groups and the age 4–5 group even react more slowly than the age-3 group in choosing pictures. Why would there be such discrepancy in the age 4–5 group's performance between the two experiments?

The present study would argue that the difference of nature between online and offline tasks can cause the age 4–5 group to perform differently. In online tasks, children are not required to give any verbal response, and they are also under no influence from the examiner as they complete the task alone in the lab. Also, online eye tracking measures can collect data of children's live eye movements which reflects their ongoing processing of the information on the screen.

However, offline tasks, where children are required to choose the correct picture, can be more pragmatically demanding and complex for children at this age and thus can cause a delay in showing their improved performance. Indeed, children at 4 and 5 are experiencing a sharp increase in vocabulary and thus in conventional lexical meanings (Falkum, 2022). In other words, the age 4–5 children's growing sensitivity to conventions may not lead them to use and understand the words in a flexible and figurative manner, as figurative uses involve large departures from conventions (Falkum, 2022); and children at this age may also, being cautious, incline to the literal meanings when the context is uncertain, which was described as a literal preference in Köder and Falkum's study (2020). Such delay in comprehension ability development and late literal preference also has their empirical equivalence in studies on other figurative devices. In tasks where children are required to verbally paraphrase the metaphorical meaning or tell a story containing a metaphor (Pearson, 1990), and tasks where children are forced to choose the correct metaphorical meaning from options (Di Paola et al., 2020; Waggoner & Palermo, 1989), participants tend to show a lack of comprehension ability until 4 to 5 years old, which may stem from the complexity and high linguistic demand of the tasks used (Köder & Falkum, 2020).

In the offline task findings of the present study, there is evidence that would suggest age 4–5 group's increased linguistic ability and high cautiousness in selection. Age 4–5 group are found to be surprisingly slower in reaction than the youngest group, which would suggest that the age 4–5 children are more cautious and may experience more competition in the decision-making process. Being cautious and hesitating, the age 4–5 children are more likely to think too much of the choice itself and misinterpret the intention of the examiners, and thus to stick to literal choices which they think safer. Among the explanation responses, more than one child from the age 4–5 group, after choosing the literal picture in metonymy trials, gave similar responses to the following utterance "(I choose this picture) because you said it's the big ear (itself), then I pick the big ear"; also, some of them added "(but) the boy there is also big ear", which means that they may actually understand the expression but they, taking into account of their interpretation of the examiner's intention and the task's aim, refrain their real thoughts and do not choose the correct picture.

Furthermore, not only the comparison between the online and offline results, but also the comparison within the offline tasks between the selection and explanation score results can provide explanations for the age 4–5 children's weakened performance in the

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

selection task. The ratio of explanation score to selection score (shown in Figure 6) of some of the age 4–5 children is greater than 2, which means that there must be cases where the age 4–5 children choose the wrong picture in the selection task, but can still give correct explanations to the meaning of the target metonymy expression, which is in line with the point that age 4–5 children's performance can be underestimated in the selection task.

Combining the two offline tasks and one online task, the present study would argue that, compared with the age-3 group, the age 4–5 children, in the selection task, show a higher vocabulary and language production ability, adopt a more cautious strategy in the decision-making process, and thus are more likely to stick to literal options in metonymy trials. However, online tasks suggest that the above-mentioned literal preference found in age 4–5 children is only a seeming inclination, which means that their declined performance in the selection task does not equal to a declined metonymy comprehension ability; instead, the apparent preference for literal meanings can be indicative of their growing ability of attending to conventions, an ability with great functional importance in children's language and social learning (Falkum, 2022; Kalish & Sabbagh, 2007), as age 4–5 children's comprehension ability in both types of metonymy improved, compared with their age-3 counterparts in the results from the eye tracking task.

4.4. Socio-cultural experience as a source of difficulty until age 6

The hypothesised 3-type difficulty categorisation of novel metonymy has been found not to be supported as children's performance in both selection and explanation tasks is not found to decline as the hypothesised level of difficulty increases. As the 3-type categorisation is based on children's animacy preference (Piaget, 1964) and the metonymy expression's degree of departure from the prototypical metonymy (Peirsman & Geeraerts, 2006), the present study would argue that the difficulty of metonymy comprehension for children does not stem from the inanimacy of content or the distance of internal-domain mapping.[3]

The present study, as one of the first to take into account different types of metonymy, did not stop at the state where the 3-type difficulty for children has been proven wrong. We merged the three novel metonymy sub-types into one 'novel metonymy type' and compared it with the conventional metonymy type. The results show significant declines from novel to conventional metonymy and significant delays in learning of conventional metonymy. As is attested in both eye tracking and explanation tasks, children's comprehension development in novel metonymy and conventional metonymy follows different trajectories: children's comprehension development of novel metonymy shows a chronological trend, with steady improvements from age 3 to age 7; while in terms of conventional metonymy, significant improvements appears only after the age of 4–5 and it is only the age 6–7 group that start to show a good command of understanding and interpreting conventional metonymies. Considering this, the present study would argue that sociocultural conventions and experience can be the source of difficulty for children's metonymy comprehension development.

As is discussed in the literature review, the difficulty of a figurative expression can be estimated from the perspective of its familiarity and semantic transparency (Cacciari &

[3]The longer distance of mapping, according to Peirsman and Geeraerts' (2006) model, means larger departure from the prototypical spatial part-whole metonymy.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Padovani, 2012). For one thing, the familiarity of a figurative expression reflects the degree of exposure of a speaker to it (Cacciari & Padovani, 2012). Highly socio-culturally related metonymy expressions can be more unfamiliar to children because such metonymy expressions are used on limited occasions and can not be comprehended until a child has experienced such occasions (e.g., the metonymy "Chairman Mao – bank notes/money" can only be understood by interlocutors who have communicated about bank notes in shopping-related contexts). For another, the semantic transparency of metonymy, as a measure of the relation between literal (default) and figurative meaning, could explain the difficulty of metonymy from conceptual aspects (e.g., strength of contiguity and distance of mapping) but would not capture the variability in usage aspects (e.g., socio-cultural experience). In other words, one metonymy expression, with fixed type of contiguity in formal analysis, can be of different difficulty for people with different social experience, as in the metonymy of "drink white", which could provide mental access to "drink white wine" in a restaurant context only for people, unlikely as children, who have such conversational experience (e.g., being exposed to wine-related conversations at parties or formal dinners).

In what ways can children older than six be more socio-culturally experienced and thus perform far better than their younger counterparts in understanding conventional metonymies? From our perspective, children's socio-cultural experience in figurative language acquisition can be influenced by family and education in terms of richness and diversity of language input.

For one thing, in mainland China, children at the age of six enter primary school and begin to receive standard education, which, from the very beginning, includes the subject of "yuwen" (literally translated as language and literature). From "yuwen" classes, primary school children receive focused input of literary reading materials which contain rich figurative uses of language and socio-cultural knowledge than daily colloquial language does, which familiarises primary school children with non-literal language uses.

Besides providing standard language education, primary school is also a place where children begin to build diverse interpersonal relationships (e.g., with schoolmates and teachers), which renders a more diverse language input and more chances for error corrections, especially teacher's correction (Lasagabaster & Sierra, 2005), compared to kindergarten period when children's language input almost comes from their family members, namely parents. As is discussed in §4.3, children at the age of four to five experience a sharp increase in vocabulary and an increasing sensitivity to conventions but still make mistakes in metonymy picture selection tasks, especially in conventional metonymy; it is just possible that age 6–7 children benefit from the correction process, especially peer correction and teacher correction, during which the learned vocabulary becomes more flexible in use.

Indeed, metonymy in real use covers a wider variety than is categorised in Peirsman and Geeraerts' (2006) model; the novelty and conventionality of metonymy are notions constructed by adults with richer language and social experiences, and thus would have opposing effects in predicting children's metonymy comprehension (conventional being harder than novel). From the view of the present study, the comprehension process of metonymy not only relies on schematic ability of category levels rendering internal domain mapping but also largely depends on socio-cultural experiences, which could be the source of the children's delayed acquisition of conventional metonymies.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

5. Conclusion

This study examined the development of metonymy comprehension among Mandarin-speaking children aged 3–7, using both online (eye-tracking) and offline (picture selection and explanation) tasks. The results provide several insights into how children's comprehension of metonymy evolves with age and across different types of metonymy.

In the offline experiment, children's performance in the explanation task improved significantly with age, while no significant differences were found between adjacent age groups in the selection task. In the eye tracking experiment, children's fixation proportion of target areas after target expression increases with age, particularly between the two younger groups.

The analysis combining selection score data and explanation score data, suggests that three-year-old children's performance can be highly random, potentially overestimating their true comprehension in selection tasks. In contrast, four-to-five-year-olds' selection performance can be underestimated, due to their cautious approach as a result of their sharp increase in linguistic capacity and sensitivity to conventions, reflected in longer reaction times. Furthermore, their decision-making process could be influenced by their perception of the examiners' intention. This discrepancy may explain why the U-shaped curve of development was found in previous studies (Falkum et al., 2017; Jiang, 2019; Köder & Falkum, 2020), although the present study does not find such U-shaped pattern in any clear form.

The present study also argues that the absence of a clear U-shaped curve in most measures (except for reaction time) can be attributed to the unique design of our study. For one thing, the interweaving of explanation task immediately after the selection task in each trial would make the children more careful in the selection process and would reduce the selection randomness of the age-3 group in our study. For another, the possible differing developmental trajectories for novel and conventional metonymy might obscure a U-shaped pattern when combined. As is to be noted, the argument of age-3 children's possible high randomness in selection tasks and the argument of their reduced randomness in the present selection task do not contradict with each other – we find evidence suggesting age-3 children's possible high randomness in other studies which only employ picture selection tasks, while the selection randomness of the youngest group in our study is indeed high, but not that high, due to methodological reasons.

In terms of the influence from metonymy type, both the offline and online tasks found a delayed acquisition of conventional metonymy compared with novel metonymy. Eye tracking data showed that the youngest group displayed an insensitivity to stimuli (reflected in the change of fixation proportion before and after target expressions) in one novel metonymy trial and in all conventional trials. This insensitivity decreases with age, and it is the oldest group (age 6–7 group) that show a clear understanding of both novel and conventional metonymy stimuli.

Therefore, the present study argues that, on the one hand, inanimacy of content and distance of mapping may not be the source of difficulty for children's metonymy acquisition; on the other hand, the difficulty, for children under six, may stem from children's lack of socio-cultural experience and conventional knowledge, which can be enriched in the form of more diversified social engagements and language input.

The study further argues that online designs are not inherently superior to offline ones because both collect data concerning different aspects of linguistic and cognitive competence. Therefore, researchers with different aims choose different methods which yield different findings. Taking the present study as an example, the explanation task results

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

align with the online results in many aspects (e.g., chronological development, influence from metonymy type), which indicates that offline tasks, if properly designed, although not measuring moment-by-moment processing, can also reflect participants' relatively pure comprehension processes, and thus offer valuable data that complements online methods.

Given limited research on the development of figurative language comprehension, particularly metonymy, as compared to metaphor and idiom (Cacciari & Padovani, 2012), the present study adds balance and diversity to this field. Specifically, by providing further explanations to children's metonymy comprehension developmental trajectory, the present study can also give novel input into the discussion of children's early onset and literal stage in figurative language.

Furthermore, the present study, as one of the pioneers to add different types of metonymy into the consideration, suggests that future studies should continue to explore the diversity within certain figurative device and consider both cognitive development and socio-cultural factors that would influence the acquisition process Our findings underline the need for a broader investigation into how children understand and use different types of figurative language, integrating various methodological approaches for a more comprehensive view.

Admittedly, the present study also comes with potential limitations, with one being the variation in sample sizes across different age groups, as the data were collected during the pandemic period. Although the overall sample sizes were sufficient for analysis, smaller subgroups, particularly in the younger age groups, may have influenced the ability to detect differences between groups. Future studies could benefit from recruiting larger and more evenly distributed samples across all age ranges to enhance the robustness of comparisons.

Acknowledgements. This research was supported by Planning Project Grant from Shanghai Planning Office of Philosophy and Social Science (project no. 2022ZYY001) and Open Project Grant from Shanghai Key Laboratory of Brain-Machine Intelligence or Information Behaviour (project no. 2021KFKT009) awarded to the second author. We would like to thank all the children who participated in our study, as well as the parents and staff at the kindergartens for their collaboration: this research would not have been possible without them.

Competing interest. The authors declare none.

Disclosure of use of AI tools. The authors declare no use of AI tools during the preparation of the manuscript.

References

Barcelona, A. (Ed.). (2003). *Metaphor and metonymy at the crossroads: A cognitive perspective*. Berlin and New York: Mouton de Gruyter.

Billow, R. M. (1981). Observing spontaneous metaphor in children. *Journal of Experimental Child Psychology*, **31**, 430–445. doi: 10.1016/0022-0965(81)90028-X

Brdar-Szabó, R., & **Brdar, M.** (2003). The MANNER FOR ACTIVITY metonymy across domains and languages. *Jezikoslovlje*, **4**, 43–69.

Cacciari, C., & **Padovani, R.** (2012). The development of figurative language. In M. J. Spivey, K. McRae, & M. F. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 505–522). Cambridge: Cambridge University Press.

Caillies, S., & **Le Sourn-Bissaoui, S.** (2008). Children's understanding of idioms and theory of mind development. *Developmental Science*, **11**, 703–711. doi: 10.1111/j.1467-7687.2008.00720.x

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Clark, E. V. (2019). Perspective-taking and pretend-play: Precursors to figurative language use in young children. *Journal of Pragmatics*, **156**, 100–109. doi: 10.1016/j.pragma.2018.12.012

Colston, H. L. (2007). What figurative language development reveals about the mind. *Mental States: Language and Cognitive Structure*, **2**, 191.

Colston, H. L. (2020). Figurative language development/acquisition research: Status and ways forward. *Journal of Pragmatics*, **156**, 176–190. doi: 10.1016/j.pragma.2019.07.002

Di Paola, S., **Domaneschi, F.**, & **Pouscoulous, N.** (2020). Metaphorical developing minds: The role of multiple factors in the development of metaphor comprehension. *Journal of Pragmatics*, **156**, 235–251. doi: 10.1016/j.pragma.2019.08.008

Falkum, I. L. (2022). The development of non-literal uses of language: Sense conventions and pragmatic competence. *Journal of Pragmatics*, **188**, 97–107. doi: 10.1016/j.pragma.2021.12.002

Falkum, I. L., **Recasens, M.**, & **Clark, E. V.** (2017). "*The moustache* sits down first": On the acquisition of metonymy. *Journal of Child Language*, **44**, 87–119. doi: 10.1017/S0305000915000720

Frisson, S., & **Pickering, M. J.** (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, **22**, 595–613. doi: 10.1080/01690960601017013

Gardner, H., **Kircher, M.**, **Winner, E.**, & **Perkins, D.** (1975). Children's metaphoric productions and preferences. *Journal of Child Language*, **2**, 125–141. doi: 10.1017/S0305000900000921

Jiang, X. H. (2019). An empirical study of preschoolers' metonymic ability development. *Modern Foreign Languages*, **4**, 487–500.

Kalish, C. W., & **Sabbagh, M. A.** (2007). Conventionality and cognitive development: Learning to think the right way. *New Directions for Child and Adolescent Development*, **115**, 1–9.

Katsos, N. (2021). The implicature and perspective-taking task: A novel way of investigating the relation between pragmatics and mind-reading. *Cambridge Occasional Papers in Linguistics*, **13**, 7.

Kecskes, I., & **Zhang, F.** (2009). Activating, seeking, and creating common ground: A socio-cognitive approach. *Pragmatics & Cognition*, **17**, 331–355. doi: 10.1075/pc.17.2.06kec

Köder, F., & **Falkum, I. L.** (2020). Children's metonymy comprehension: Evidence from eye-tracking and picture selection. *Journal of Pragmatics*, **156**, 191–205. doi: 10.1016/j.pragma.2019.07.007

Köder, F., & **Falkum, I. L.** (2021). Irony and perspective-taking in children: The roles of norm violations and tone of voice. *Frontiers in Psychology*, **12**, 624604. doi: 10.3389/fpsyg.2021.624604

Kövecses, Z. (2002). *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Z., & **Radden, G.** (1998). Metonymy: Developing a cognitive linguistic view. *Cognitive Linguistics*, **9**, 37–78. doi: 10.1515/cogl.1998.9.1.37

Lakoff, G., & **Johnson, M.** (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

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

Lasagabaster, D., & **Sierra, J. M.** (2005). Error correction: Students' versus teachers' perceptions. *Language Awareness*, **14**, 112–127. doi: 10.1080/09658410508668828

Levorato, M. C., & **Cacciari, C.** (1992). Children's comprehension and production of idioms: The role of context and familiarity. *Journal of Child Language*, **19**, 415–433. doi: 10.1017/S0305000900011478

Littlemore, J. (2015). *Metonymy: Hidden shortcuts in language, thought and communication*. Cambridge: Cambridge University Press.

Martín-González, I. R. Ronderos, C., **Castroviejo, E.**, **Schroeder, K.**, **Lossius-Falkum, I.**, & **Vicente, A.** (2024). That kid is a grasshopper! Metaphor development from 3 to 9 years of age. *Journal of Child Language*, **51**, 1–26. doi: 10.1017/S0305000924000187

Nerlich, B. (1999). "Mummy, I like being a sandwich": Metonymy in language acquisition. In K.-U. Panther & G. Radden (Eds.), *Human cognitive processing* (Vol. **4**, p. 361–383). Amsterdam and Philadelphia: John Benjamins.

Nunberg, G. (1995). Transfers of meaning. *Journal of Semantics*, **12**, 109–132.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, **3**, 143–184.

Panther, K.-U., & **Thornburg, L. L.** (Eds.). (2003). *Metonymy and pragmatic inferencing* (Vol. **113**). Amsterdam and Philadelphia: John Benjamins.

Pearson, B. (1990). The comprehension of metaphor by preschool children. *Journal of Child Language*, **17**, 185–203.

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press

Peirsman, Y., & **Geeraerts, D.** (2006). Metonymy as a prototypical category. *Cognitive Linguistics*, **17**, 269–316. doi: 10.1515/COG.2006.007

Piaget, J. (1964). Part I: Cognitive development in children – Piaget development and learning. *Journal of Research in Science Teaching*, **40**.

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics*, **25**, 51–79. doi: 10.1075/bjl.25.04pou

Pouscoulous, N., & **Tomasello, M.** (2020). Early birds: Metaphor understanding in 3-year-olds. *Journal of Pragmatics*, **156**, 160–167. doi: 10.1016/j.pragma.2019.05.021

Radden, G., & **Kövecses, Z.** (1999). Towards a theory of metonymy. In K.-U. Panther & G. Radden (Eds.), *Metonymy in language and thought* (pp. 17–59). Amsterdam and Philadelphia: John Benjamins.

Rundblad, G., & **Annaz, D.** (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, **28**, 547–563. doi: 10.1348/026151009X454373

Schumacher, P. B., **Weiland-Breckle, H.**, **Reul, G.**, & **Brilmayer, I.** (2023). Tracking meaning evolution in the brain: Processing consequences of conventionalization. *Cognition*, **240**, 105598. doi: 10.1016/j.cognition.2023.105598

Slabakova, R., **Cabrelli Amaro, J.**, & **Kang, S. K.** (2013). Regular and novel metonymy in native Korean, Spanish, and English: Experimental evidence for various acceptability. *Metaphor and Symbol*, **28**, 275–293. doi: 10.1080/10926488.2013.826556

Slabakova, R., **Cabrelli Amaro, J.**, & **Kyun Kang, S.** (2016). Regular and novel metonymy: Can you curl up with a good Agatha Christie in your second language? *Applied Linguistics*, **37**, 175–197. doi: 10.1093/applin/amu003

Tantucci, V., & **Wang, A.** (2020). From co-actions to intersubjectivity throughout Chinese ontogeny: A usage-based analysis of knowledge ascription and expected agreement. *Journal of Pragmatics*, **167**, 98–115. doi: 10.1016/j.pragma.2020.05.011

Van Herwegen, J., **Dimitriou, D.**, & **Rundblad, G.** (2013). Development of novel metaphor and metonymy comprehension in typically developing children and Williams syndrome. *Research in Developmental Disabilities*, **34**, 1300–1311. doi: 10.1016/j.ridd.2013.01.017

Vicente, A., & **Falkum, I. L.** (2023). Accounting for the preference for literal meanings in autism spectrum conditions. *Mind & Language*, **38**, 119–140. doi: 10.1111/mila.12371

Waggoner, J. E., & **Palermo, D. S.** (1989). Betty is a bouncing bubble: Children's comprehension of emotion-descriptive metaphors. *Developmental Psychology*, **25**, 152–163. doi: 10.1037/0012-1649.25.1.152

Winner, E., **Rosenstiel, A. K.**, & **Gardner, H.** (1976). The development of metaphoric understanding. *Developmental Psychology*, **12**, 289–297. doi: 10.1037/0012-1649.12.4.289

Zhang, H., & **Lu, W. Z.** (2010). *Cognitive metonymy*. Shanghai, China: Shanghai Foreign Language Education Press.

Zheng, Q., **Jia, Z.**, & **Liang, D.** (2015). Metaphor and metonymy comprehension in Chinese-speaking children with high-functioning autism. *Research in Autism Spectrum Disorders*, **10**, 51–58. doi: 10.1016/j.rasd.2014.11.007

Cite this article: Xie, S., & He, C. (2026). An empirical study on native Mandarin-speaking children's metonymy comprehension development. *Journal of Child Language* **53,** 80–107, https://doi.org/10.1017/S0305000924000539

https://doi.org/10.1017/S0305000924000539 Published online by Cambridge University Press