languages

Tracked for Impact Factor

CITESCORE 0.9

Article

The Role Classifiers Play in Selecting the Referent of a Word

Weiyi Ma, Peng Zhou and Roberta Michnick Golinkoff

Special Issue
Current Research on Chinese Morphology

Edited by
Prof. Dr. Jidong Chen and Prof. Dr. Hongyin Tao

MDPI

https://doi.org/10.3390/languages8010084

languages

MDPI

Article

The Role Classifiers Play in Selecting the Referent of a Word

Weiyi Ma [1,*,†], Peng Zhou [2,†] and Roberta Michnick Golinkoff [3]

[1] School of Human Environmental Sciences, University of Arkansas, Fayetteville, AR 72701, USA
[2] Department of Foreign Languages and Literatures, Tsinghua University, Beijing 100084, China
[3] School of Education, Departments of Psychological and Brain Sciences and Linguistics and Cognitive Science, University of Delaware, Newark, DE 19716, USA
* Correspondence: weiyima@uark.edu
† These authors contributed equally to this work.

Abstract: An important cue to the meaning of a new noun is its accompanying classifier. For example, in English, X in "a sheet of X" should refer to a broad, flat object. A classifier is required in Chinese to quantify nouns. Using children's overt responses in an object/picture selection task, past research found reliable semantic knowledge of classifiers in Mandarin-reared children at around age three. However, it is unclear how children's semantic knowledge differs across different types of classifiers and how this difference develops with age. Here we use an arguably more sensitive measure of children's language knowledge (the intermodal preferential-looking paradigm) to examine Mandarin-reared three-, four-, and five-year-olds' semantic knowledge of four types of classifiers indicating animacy (human vs. animal distinction), configuration (how objects are arrayed), object shape, and vehicle function. Multiple factors were matched across classifier types: the number of classifiers, perceived familiarity and perceived typicality of the target, and the visual similarity of the two images paired together. Children's performances differed across classifier types, as they were better with animacy classifiers than with configuration and vehicle function classifiers. Their comprehension was reliable for animacy, object shape, and vehicle function classifiers but not for configuration classifiers. Furthermore, we did not find conclusive evidence for an age-dependent improvement in the child's performance. The analysis, including the oldest (five-year-olds) and youngest (three-year-olds) children, revealed a marginally significant age effect.

Keywords: Mandarin Chinese; classifier; intermodal preferential looking

Citation: Ma, Weiyi, Peng Zhou, and Roberta Michnick Golinkoff. 2023. The Role Classifiers Play in Selecting the Referent of a Word. *Languages* 8: 84. https://doi.org/10.3390/languages8010084

Academic Editors: Jidong Chen and Hongyin Tao

Received: 30 March 2022
Revised: 27 February 2023
Accepted: 28 February 2023
Published: 14 March 2023

1. Introduction

Various tools are available in the speech input addressed to children that can help them disambiguate the referent of a new word. One useful, though not infallible, cue to the meaning of a new word is its form class. In English, a novel word may be a noun if it follows an article (a/an/the); it may be a verb if it bears a morphological inflection (e.g., -ing). In Mandarin, there are also morphosyntactic cues to the noun–verb distinction (e.g., Ma et al. 2019), and children use these cues to disambiguate the referent of a new word (Ma et al. 2020b; Zhou and Ma 2018). However, even if a child understands that a new word is a noun and likely the name of an object, the indeterminacy of reference of the noun remains, as there may be multiple objects in the environment. Furthermore, a noun can refer to an object as a whole unit (e.g., a rabbit), a part of the object (e.g., ears of the rabbit), or any combination of these parts (e.g., Hollich et al. 2007; Gleitman and Trueswell 2020; Nelson 1988; Srinivasan and Snedeker 2014). In Chinese, another important cue to the meaning of a noun is its accompanying classifier. Although classifiers are not generally a feature of English or other European languages, classifier-like constructions are also used in English. Thus, X in "a sheet of X" should refer to a broad, flat object, while Y in "a chunk of Y" probably refers to a thick, solid object. Classifiers are important for word acquisition in such languages as Mandarin Chinese (Chao 1968), where classifiers are obligatory when

the number of entities needs to be specified. This study examines Mandarin-reared three-, four-, and five-year-olds' knowledge of classifiers and the course of their development.

1.1. The Acquisition of Mandarin Classifiers

Chinese grammar dictates that when a noun is preceded by a numeral, a classifier should be inserted before the noun. Thus, a phrase equivalent to "three books" should be expressed as *sān běn shū* [three classifiers (CL) book][1]. A classifier can either be a mass classifier or a count classifier (e.g., Allan 1977; Cheng and Sybesma 1998, 1999; Tai 1994; Zhang 2007). Mass classifiers are open-class words that can be used productively with nouns (Li et al. 2008, 2010) and are comparable to words denoting measurement in English (e.g., a *bowl* of apples). Count classifiers (e.g., *běn*) are closed-class words that have no direct translation in English, thus making them the prototypical classifiers for adult Chinese speakers (Erbaugh 2006; Killingley 1983). This study focuses mostly on children's comprehension of count classifiers.

One approach to understanding Mandarin-reared children's knowledge of classifiers is to scrutinize their *production* of classifiers. Although Chinese-reared children start to produce classifier structures at ages two–three, the accuracy of their classifier use is as low as 17% even at age four (e.g., Erbaugh 1986; Fang 1985; Hu 1993; Ying et al. 1983). Children struggle to recognize the co-occurrence between classifiers and nouns even at age five (Fang 1985; Ying et al. 1983). In addition, they first associate specific classifiers with only prototypical exemplars labeled by nouns (Hu 1993); both over- and under-generalized use of specific classifiers occurs in production (Erbaugh 1986; Hu 1993; Loke 1991). These findings suggest that children's semantic understanding is crucial for their acquisition of classifiers in language production.

There are only several experimental studies on Mandarin-reared children's *comprehension* of classifiers (Chien et al. 2003; Fang 1985; Hu 1993; Hao 2019; Li et al. 2008, 2010). In these studies, Mandarin-reared children were administered a forced-choice object or picture selection task, where they were asked to select the object/picture that could be labeled by the classifier phrase they were offered. Fang (1985) found that children's semantic knowledge of classifiers was unreliable at age four. Chien et al. (2003) found that three-year-olds' performance was better than chance for eight of the 14-count classifiers and three of the four mass classifiers they tested. There is also evidence that knowledge of the semantic distinctions between count and mass classifiers was fragile even at ages four–six (Li et al. 2008). In addition, some studies revealed an age-dependent improvement in classifier knowledge (Chien et al. 2003; Hu 1993; Li et al. 2008, 2010), while other studies failed to observe such an improvement between three- and five-year-olds (Hu 1993—for the overall score across the four object shape classifiers tested), or between four- and five-year-olds (Hao 2019). These findings suggest a difficulty with classifier acquisition, making the acquisition of classifiers an extended process. Thus, an age-dependent improvement in classifier knowledge may be unobservable between ages three and five.

While these studies pave the way for the present investigation, four issues remain for our understanding of how children acquire their comprehension of classifiers. First, none of these studies systematically examined children's semantic knowledge of sub-types of classifiers, such as the classifiers indicating animacy, configuration, and object shape. For example, only object shape classifiers were tested by Fang (1985) and Hao (2019); sub-types of count classifiers were not analyzed by Chien et al. (2003). In addition, the number of classifiers was not matched across the classifier types (e.g., Hu 1993; Li et al. 2010). For example, Li et al. (2010) found that children first noticed that a count classifier could specify the property of shape, suggesting that object shape classifiers may be easier to acquire than other types of classifiers. However, Li et al. (2010—Experiment 1) used nine object shape classifiers but only one animacy classifier, possibly promoting a practice effect for the object shape classifiers. Furthermore, it is possible that the finding is specific to the one animacy classifier used. Thus, the generalizability of the finding that object shape classifiers are exceptionally easy to acquire needs further examination.

Second, the factors that affect the acquisition of classifier knowledge remain understudied. Perhaps, object shape classifiers are acquired early in life (Li et al. 2010) because they support the shape bias that children use in early word acquisition—the tendency for names to extend to same-shaped objects rather than other characteristics, such as color or texture (e.g., Diesendruck et al. 2003; Landau et al. 1988). However, the existent research did not systematically compare child comprehension of object shape classifiers versus that of the classifiers unrelated to the word-learning biases, thus leaving the prediction untested.

Third, these studies used children's responses in an object/picture selection task or their speech production—cognitively demanding tasks. Thus, it is unclear whether the performance observed in these studies arose from the difficulty with classifier acquisition or the high cognitive demand of the tasks used. Research on the acquisition of classifiers in young children requires a method, such as the Intermodal Preferential Looking Paradigm (IPLP—Golinkoff et al. 2013). In an IPLP experiment, children's language comprehension is measured by their differential visual fixation to one of two images presented side-by-side when only one matches an accompanying linguistic stimulus. The IPLP allows children to reveal their language knowledge before children can use it in production, thus making it arguably a more sensitive measure of children's language than other tasks.

Fourth, none of these studies controlled the major psycholinguistic attributes that can affect language processing performance, such as the typicality of the target image/object, the perceived familiarity of the target image/object, and the visual similarity between the two images/objects paired together. For example, young children tend to map words first to prototypical exemplars and later to less typical exemplars (e.g., Meints et al. 1999). In addition, high picture familiarity can facilitate children's performance in picture-naming tasks (Cycowicz et al. 1997). A comparison of the child's performances across multiple types of classifiers requires a control of the visual similarity between the two images/objects paired together across classifier types since it is easier to find the target between two visually dissimilar images/objects than between two visually similar images/objects.

Thus, a systematic comparison of child knowledge across sub-types of classifiers requires (a) a within-subject design where participants are tested on their knowledge of multiple sub-types of classifiers; (b) the control of major variables that could affect a child's performance; and (c) a method, such as the IPLP.

1.2. The Current Study

Here we asked whether Mandarin-reared three-, four-, and five-year-olds could use their semantic knowledge of classifiers to determine the referents of classifier phrases—"one CL *shénme*" (meaning "something/somebody") and whether children's performance differed across classifier types and improved with age. This is the first IPLP study that examined children's comprehension of multiple types of classifiers (animacy, object shape, configuration, and vehicle function classifiers) with several major variables that could affect a child's performance controlled: the number of classifiers, the typicality of the target image, the perceived familiarity of the target image, and the visual similarity between the two images paired together.

Four types of classifiers (i.e., animacy, object shape, configuration, and vehicle function classifiers) were used (Table 1). There were four classifiers within each classifier type. The animacy classifiers indicated the human versus animal distinction. The object shape classifiers indicated the shape of an object. The classifiers indicating the distinction among land-, water-, and air-based vehicles were referred to as vehicle function classifiers. Following Li et al. (2010), the classifiers indicating the arrangement of multiple objects (e.g., a queue of, a flock of) were referred to as configuration classifiers.

Table 1. The classifiers tested in this study.

Animacy	pair 1 pair 2	*wèi*: polite classifier for people *míng*: classifier for people	*zhī*: animals *tóu*: domesticated animals
Configuration	pair 1 pair 2	*qún:* group, herd *shuāng:* pair	*pái*: a straight line, queue *zhī*: a single one
Vehicle function	pair 1 pair 2	*liàng*: wheeled land vehicles (e.g., car) *jià*: winged flying vehicles (e.g., plane)	*sōu*: water vehicles (e.g., ship) *liè*: long, arrayed vehicles (e.g., train)
Object shape	pair 1 pair 2	*gēn*: thin, slender, pole, stick objects *lì*: small, grain-like objects	*zhāng*: flat objects *tiáo*: long, narrow objects

The four types of classifiers were used for several reasons. First, research suggested that object shape classifiers were easier to acquire than animacy classifiers (Li et al. 2010). The inclusion of the object shape and animacy classifiers allowed us to test the replicability of this finding when different object shape and animacy classifiers were used and when the number of classifiers was matched across classifier types. Second, the inclusion of the four types of classifiers allowed us to examine the factors that can affect the acquisition of classifiers. For example, the current design enabled us to determine (a) whether the classifiers that are related to early word learning biases (object shape classifiers) tend to be acquired earlier than the classifiers that are unrelated to these biases (animacy, configuration, vehicle function classifiers); (b) whether the classifiers that indicate a concept depicted by one single object (animacy, object shape, vehicle function classifiers) tend to be acquired earlier than the classifiers that indicate a concept depicted by multiple objects (configuration classifiers); and (c) whether the classifiers that are crucial for children's survival (animacy classifiers) tend to be acquired earlier than the classifiers that are less crucial for their survival (vehicle function classifiers).

This study examined three questions. First, could children—especially the youngest age group (age three)—comprehend some of the classifiers in an IPLP setting? Based on the past finding that three-year-olds' performance was better than chance for eight of the 14 count classifiers and three of the four mass classifiers (Chien et al. 2003), we predicted that the three-year-olds should be able to comprehend some of the classifiers in an IPLP setting. Second, did children's performance differ across classifier types? Based on Li et al.'s finding (2010), we predicted that classifier types should be differentially difficult, with some classifiers (e.g., object shape classifiers) being exceptionally easy to comprehend. Thus, children's performance should differ across classifier types. In addition, when children's performance was compared against chance, their comprehension should be significantly above chance with some classifier types but only emerging and fragile with other classifier types. Third, did the child's performance improve across age groups? Note that past research is divided on the observability of an age-dependent improvement in classifier comprehension. While some studies observed an age-dependent improvement between ages three and seven (Chien et al. 2003; Li et al. 2008, 2010), other studies did not find such an improvement between the ages three and five (Hu 1993) or between the ages four and five (Hao 2019). This study examined whether an age-dependent improvement was observable when children were tested in an IPLP setting.

2. Method

2.1. Child Participants

Ages for participation were based on prior research. The participants were 24 3-year-olds (M = 3.38; range = 3.02–3.65; female = 12), 24 4-year-olds (M = 4.29; range = 4.08–4.54; female = 12), and 24 5-year-olds (M = 5.33; range = 5.04–5.58; female = 12) recruited at the Hubei University of Technology Preschool in China. The 3-year-olds were recruited because research showed that 3-year-olds' comprehension of classifiers was above chance but fragile (Li et al. 2010). Thus, age 3 is an ideal age group to examine factors that can affect child comprehension of classifiers. The 4- and 5-year-olds were recruited to

explore the development of classifier knowledge. More importantly, this design allowed us to determine whether an age-dependent improvement in children's performance was observable between ages 3–5 *in an IPLP setting*. All children were from monolingual Mandarin-speaking households and had no history of auditory or visual impairments. The minimum sample size (n = 60) was established by conducting a power analysis using G*Power, based on an effect size of f = 0.25, α error probability of 0.05, power (1-β error probability) of 0.99, use of an ANOVA analysis: repeated measures, within–between interactions, containing three groups (three age groups) and four measures (four types of classifiers; Faul et al. 2007). This sample size is also consistent with previous IPLP research on children's word recognition (e.g., Ma et al. 2011, 2017, 2019; Mani and Plunkett 2007; Singh et al. 2014).

2.2. Visual and Auditory Stimuli

Table 1 shows the four categories of classifiers tested; the 16 classifiers were chosen because they appear with high frequency in Chinese adult texts (Da 2004) and, therefore, likely have a high frequency in child-directed speech as well. In addition, eight of the 16 classifiers (animacy: *zhī*, *tóu*, *wèi*; configuration: *pái*; vehicle function: *liàng*; shape: *zhāng*, *lì*, *tiáo*) also have the age of acquisition data based on parental reports on the MacArthur CDI (Tardif et al. 2008). Those CDI data showed that the eight classifiers were acquired by Mandarin-reared children by age 3—the youngest age range of the participants recruited in this study, suggesting that the participants in this study were likely to be familiar with the classifiers tested.

On each trial, children were shown two images side-by-side in the IPLP while the accompanying language prompted children to look at one of them. Two classifiers of the same type were paired together. The *wèi–zhī* and *míng–tóu* pairs were used for animacy classifiers; the *pái-qún* and *shuāng–zhī* pairs were used for configuration classifiers; the *jià–liè* and *liàng–sōu* pairs were used for vehicle function classifiers; the *gēn–zhāng* and *lì–tiáo* pairs were used for object shape classifiers. This design required children to decide between two classifiers of the same type, making this study a stringent test of children's classifier knowledge. These classifier pairs were chosen because the two images associated with the two classifiers were visually distinct, and the classifier pairs have different vowels and consonants, ensuring auditory discriminability. To minimize the influence of children's prior exposure to the combination of classifiers and visual stimuli, every effort was made to select visual stimuli for which children may not have ready names. In other words, the objects shown were either real objects seen relatively infrequently by children of these ages or imaginary objects. The visual stimuli were selected from the Novel Object and Unusual Name (NOUN) database (Horst 2009), online image databases for imaginary vehicles and animals, and images of Caucasian men—presumably seen relatively infrequently by Mandarin-reared children at this age. Sixteen pairs of images were used (two for each classifier pair), each testing children's knowledge of one classifier. Each image pair was shown only once per participant, ensuring that children could not use mutual exclusivity to identify the target across trials.

A female native speaker of Beijing Mandarin produced auditory stimuli in a sound-attenuated recording chamber (Table 2). Speech stimuli were produced in a child-directed manner (Cooper and Aslin 1990; Fernald 1985; Ma et al. 2011, 2020a, 2022a; Werker et al. 1994). To maintain the children's attention, slightly different carrier sentences were used across classifiers. Each carrier sentence that the speaker produced contained a classifier structure (e.g., *kàn! zhè shì yí* CL [Look! This is a CL]) and one indefinite pronoun—*shénme* (meaning "something/somebody"). Then, using Audacity 2.0.3, the existential indefinite, *shénme,* was synthesized (using Audacity 2.0.3) into each of the carrier phrases at the sentence-final position (e.g., *kàn! zhè shì yí CL shénme* [Look! This is a CL something/somebody) (Table 2). The use of *shénme* required children to rely on their semantic knowledge of the classifier to find the target.

Table 2. Carrier phrases used in one of the conditions.

Trial	Carrier Phrase	Character by Character English Translation
1	Kuài Kàn! Zhè-er yǒu yì pái	Quickly look! There is one classifier . . .
2	Kàn! Nà shì yì gēn . . .	Look! That is one classifier . . .
3	Kàn kàn! Nǎ gè shì yí lì . . .	Look, look! Which one is one classifier . . .
4	Qiáo! Zhè-lǐ yǒu yì sōu . . .	Look! Here exists one classifier . . .
5	Nǐ qiáo! Zhè shì yì zhī . . .	You look! Here is one classifier . . .
6	Kuài qiáo! Nà-er shì yì shuāng . . .	Quickly look! There is one classifier . . .
7	Kuài qiáo! Nà-er shì yì míng . . .	Quickly look! There is one classifier . . .
8	Kuài Kàn! Zhè-er yǒu yì liàng . . .	Quickly look! Here exists one classifier . . .
9	Kuài qiáo! Nà-er shì yì tiáo . . .	Quickly look! There is one classifier . . .
10	Kàn kàn! Nǎ gè shì yì zhī . . .	Look, look! Which one is one classifier . . .
11	Qiáo! Zhè-lǐ yǒu yì qún . . .	Look! Here exists one classifier . . .
12	Nǐ qiáo! Zhè shì yì zhāng . . .	You look! This is one classifier . . .
13	Nǐ qiáo! Zhè shì yí liè . . .	You look! This is one classifier . . .
14	Kàn kàn! Nǎ gè shì yì tóu . . .	Look, look! Which one is one classifier . . .
15	Kàn! Nà shì yí wèi . . .	Look! That is one classifier . . .
16	Kàn! Nà shì yí jià . . .	Look! That is one classifier . . .

2.3. Apparatus and Procedure

The experimental procedure was almost identical to that of Singh et al. (2014, 2015) and Ma et al. (2017—Experiment 2). Participants were tested in a quiet testing booth at their school. Participants sat on a blindfolded female research assistant's lap, facing a 39-inch LED TV monitor 1 m from the center of the screen. Visual stimuli were displayed to the left and right of the screen at eye level. Auditory stimuli were presented through internal speakers of the TV monitor. A hidden camera recorded children's visual fixation on the display. Video recordings were then coded offline.

Before each trial, children saw an attention-getter (e.g., a giggling boy) in the center of the screen. An experiment consisted of a task familiarization phase and a test phase. In the task familiarization phase (2 trials), children were presented with images of a chicken and a car side-by-side and were directed to look at the chicken in one trial and the car in the other. In the test phase (16 trials), on each trial, children were presented with two images side-by-side, accompanied by a pre-recorded carrier sentence containing a classifier phrase that can be used to quantify one of the two images. Classifiers of the same type were not presented on more than two consecutive trials. Four stimulus orders were created. The left/right position of target images was counterbalanced across subjects.

Children saw two images side-by-side for 6 s on each trial, while the onset of the vocalized classifier began 2633 ms into a trial (Table 3). Each trial was segmented into two 3-s phases: a pre-classifier phase and a post-classifier phase. Based on analysis standards set by prior research (e.g., Gonzalez-Gomez et al. 2013; Ma et al. 2017; Ma and Zhou 2019; Mani and Plunkett 2007; Singh et al. 2014; Swingley and Aslin 2000, 2002; White and Aslin 2011), visual fixation on the post-classifier phase was calculated from 367 ms after the onset of the classifier to remove the time taken to launch an eye movement in response to auditory input (the results reported here remained the same when other minima, e.g., 200 and 400 ms, were used). The two-phase design is a typical IPLP procedure used to investigate young children's sensitivity to familiar words (e.g., Ma et al. 2017; Mani and Plunkett 2007; Singh et al. 2015). Since the vocalized classifier began 2633 ms into a trial, children would have looked randomly or roughly equally at the two images in the pre-classifier phase because there was no match to be found before the onset of the classifier, and because children did not yet have sufficient time (367 ms) to process the classifier after its onset in the pre-classifier phase. In the post-classifier phase, if children comprehend the meaning of the classifier, they should look at the target more than the distractor. Thus, an increase in looking time to the target image across phases indicates that the participant mapped the verbal label onto the visual target. This design controlled for any potential preference for one of the images on trial because even if children preferred the distractor or the target, they

should have still looked more at the target in the post-classifier phase than the pre-classifier phase if children understood the meaning of the classifier.

Table 3. An example of trials for the visual and speech stimuli.

	Left Side	Right Side	Speech Stimuli
Animacy			Kàn [look]! Nà [that] shì [is] yí [one] wèi [classifier] shénme [something]
Configuration			Kuài [quickly] Kàn [Look]! Zhè er [here] yǒu [exist] yì [one] pái [classifier] shénme [something]
Vehicle function			Kàn [Look]! Nà [that] shì yí [one] jià [classifier] shénme [something]
Object shape			Nǐ [you] qiáo [look]! Zhè [this] shì [is] yì [one] zhāng [classifier] shénme [something]

Note: The test phase consisted of 16 trials, each testing the child's knowledge of one classifier. Classifiers of the same type were not presented on more than two consecutive trials. Four stimulus orders were created. The left/right position of target images was counterbalanced across conditions.

2.4. Coding and Data Analysis

Using SuperCoder (Hollich 2005), participants' eye movements were coded frame-by-frame to 1/30 of a second with the audio turned off so that the coder was blind to the condition. Coding of 20% of the subjects by another coder yielded an inter-coder agreement of 98%. Based on the established procedure, data analysis only included the trials in which children had an attention span of more than 20% in both the pre-classifier and post-classifier phases (Ma et al. 2017; Quam and Swingley 2010; Singh et al. 2014) and during which the children fixated on both the target and the distractor in the pre-classifier phase (Ma et al. 2017; Mani and Plunkett 2007). Based on these criteria, we excluded 79 trials across all participants. Thus, the final dataset contained 1073 trials (16 trials × 72 participants).

For each participant, on each trial, the proportion of time spent fixating the target image (target fixation [TF]) was calculated within each phase (i.e., the pre- and post-classifier phases). In each phase, TF was calculated by dividing the length of looking time to the target by the total length of looking time to the target and non-target (Ma et al. 2017). Then, the cross-phase TF increase (post-classifier TF minus pre-classifier TF) was calculated for each trial. In the post-classifier phase, if children comprehend the meaning of the classifier, the TF should be longer in the post-classifier phase than in the pre-classifier phase, leading to a cross-phase TF increase that was greater than 0. By contrast, if children did not comprehend the meaning of the classifier in the post-classifier phase, their visual fixation should not have significantly differed between phases, leading to a cross-phase TF increase that did not significantly exceed 0.

2.5. Stimulus Verification, Typicality Rating, Familiarity, and Visual Similarity Rating in Adults

Adult participants were recruited to verify the target assignment used in the current task. To further validate the comparison of the child's performance across the four types of classifiers, adult participants were also asked to rate the typicality of the target image, its perceived familiarity, and the perceptual similarity of the two images paired together.

Thirty adult native Mandarin speakers (M = 19.10 years, range = 17–22; 17 females) completed these tasks sequentially. First, they were asked to select the image that could

be labeled by the classifier; if they deemed both images could be labeled by the classifier phrase, they were asked to so indicate and select the image that could best be labeled by that classifier. This task aimed to confirm the target assignment used in the current task. Each task consisted of 16 trials presented in random order on a 13-inch computer screen. Adults were presented with an image pair and a pre-recorded classifier phrase as used in the child study. Second, adults were asked to rate the typicality of the target image for that classifier on a 7-point Likert scale (1 = a poor example; 7 = a great example). Finally, adults rated the perceived familiarity of the target image on a 7-point Likert scale (1 = not familiar; 7 = highly familiar). The typicality and familiarity rating tasks aimed to determine whether the typicality and perceived familiarity of the target images differed across classifier types.

Another group of 30 adult native Mandarin speakers (M = 20.07 years, range = 17–23; 18 females)—who did not participate in the above tasks—rated the visual similarity of the two images paired together. On each trial, the participants were shown one of the 16 image pairs as used in the child study and rated how visually similar the two images were on a 7-point Likert scale (1 = not similar; 7 = highly similar).

3. Results

3.1. Adults' Data: Experimental Checks

The image name match task. On 468 trials (97.5% of all trials) among the 480 trials (16 trials × 30 participants), adults selected the assigned target image as the only image that could be labeled by the classifier in the two-option forced-choice task, thus verifying the assignment of classifiers to the targets.

The typicality rating task. A direct-entry logistic regression analysis was performed with 480 ratings (16 trials × 30 participants), where the typicality rating served as the dependent variable and classifier type (animacy, configuration, vehicle function, object shape) served as predictors. Results showed that classifier type did not predict the rating (p = 0.74), suggesting that typicality ratings did not differ across classifier types. Then, within each adult, we calculated an average typicality rating for each classifier type: object shape (M = 4.52, SD = 0.96), configuration (M = 4.48, SD = 0.93), animacy (M = 4.41, SD = 0.89), and vehicle function (M = 4.31, SD = 1.00) classifiers. Separate paired sample t-tests—comparing the average typicality ratings between classifier types—revealed no significant results (p's > 0.42).

The familiarity rating task. A direct-entry logistic regression analysis was performed with 480 ratings, where the familiarity rating served as the dependent variable, and the classifier type served as the predictor. Results showed that classifier type did not predict the rating (p = 0.83), suggesting that familiarity ratings did not differ across classifier types. Then, within each adult participant, we calculated an average familiarity rating for each classifier type: vehicle function (M = 2.94, SD = 0.65), configuration (M = 2.91, SD = 0.80), animacy (M = 2.85, SD = 0.61), and object shape (M = 2.75, SD = 0.68) classifiers. Separate paired sample t-tests—comparing the average familiarity ratings between classifier types—revealed no significant results (p's > 0.28).

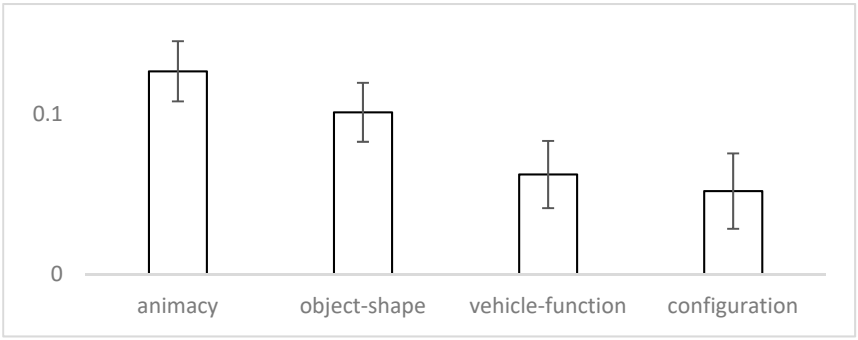
The visual similarity rating task. A direct-entry logistic regression analysis was performed with 480 ratings, where the similarity rating served as the dependent variable, and the classifier type served as the predictors. Results showed that classifier type did not predict the rating (p = 0.41), suggesting that perceptual similarity ratings did not differ across classifier types. Then, within each adult participant, we calculated an average rating for each classifier type: vehicle function (M = 3.88, SD = 1.19), configuration (M = 3.66, SD = 0.97), object shape (M = 3.66, SD = 0.83), and animacy (M = 3.51, SD = 0.78) classifiers. Separate paired sample t-tests—comparing the average perceptual similarity ratings between classifier types—revealed no significant results (p's > 0.22).

Thus, adult participants' data confirmed the target assignment in the two-choice selection task. In addition, their data verified that the typicality and perceived familiarity of the target images did not differ across classifier types; nor did the virtual similarity of two images paired together on a trial, thus validating the comparative analyses of children's data across classifier types.

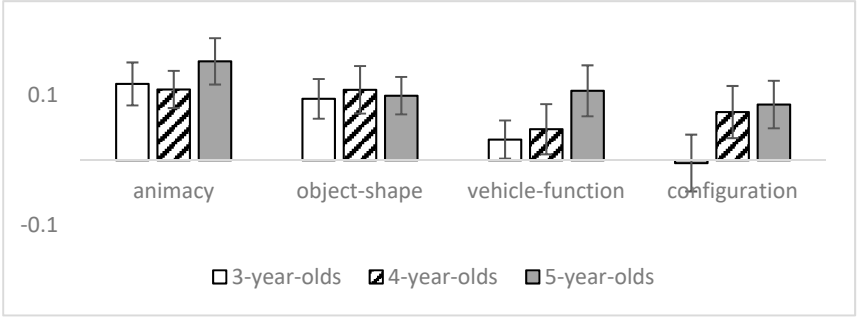
3.2. Children's Data

For each child, on each trial, we first calculated the target fixation (TF) within each phase (post-classifier TF, pre-classifier TF). Then, we calculated the TF increase on each trial (post-classifier TF minus pre-classifier TF) and the average TF increase across the four trials for each classifier type.

Did the average TF increase differ across classifier types and age groups? A 4 × 3 mixed model ANOVA with the within-subject factor of classifier type (animacy, configuration, object shape, vehicle function) and the between-subject factor of age groups (three-, four-, and five-year-olds) analyzed the average TF increase. A significant main effect of classifier type emerged ($F(3,207) = 3.00$, $p = 0.03$, $\eta_p^2 = 0.04$), but neither the main effect of age group ($F(2,69) = 1.82$, $p = 0.17$) nor the classifier type × age group interaction ($F(6,207) = 0.54$, $p = 0.78$) was significant, suggesting that children's performance differed across classifier types for all age groups. Since children's performance did not differ by age group, post-hoc analyses analyzed the three age groups' combined data. Descriptive analyses showed that animacy classifiers had the highest average TF increase ($M = 0.13$, $SD = 0.16$), followed by object shape ($M = 0.10$, $SD = 0.16$), vehicle function ($M = 0.06$, $SD = 0.18$), and configuration ($M = 0.05$, $SD = 0.20$) classifiers (Figure 1). Post-hoc analyses first compared the configuration and vehicle function classifiers (the two types of classifiers with a lower average TF increase) against animacy classifiers. Two paired sample *t*-tests showed that the average TF increase was higher with animacy classifiers than with configuration ($t(71) = 2.60$, $p = 0.01$) and vehicle function classifiers ($t(71) = 2.30$, $p = 0.02$). A significance cutoff level of 0.025 [0.05/2] was used in two-comparison *t*-tests. Then, two paired sample *t*-tests compared the configuration and the vehicle function classifiers against the object shape classifiers. Results showed that the average TF increase did not differ between the object shape and configuration classifiers ($t(71) = 1.72$, $p = 0.09$) or between the object shape and vehicle function classifiers ($t(71) = 1.48$, $p = 0.14$).



(**a**)



(**b**)

Figure 1. (**a**) The TF increase for each type of classifiers across all participants. (**b**) The TF increase for each type of classifiers within each age group. Error bars reflect SEM.

To further explore the age group difference, A 4 × 2 mixed model ANOVA with the within-subject factor of classifier type and the between-subject factor of age group (three- and

five-year-olds) analyzed the average TF increase data. Results revealed a marginally significant main effect of age group ($F(1,46) = 3.52$, $p = 0.067$, $\eta_p^2 = 0.07$), a significant main effect of classifier type ($F(3,138) = 2.84$, $p = 0.04$, $\eta_p^2 = 0.06$), and an insignificant classifier type × age group interaction ($F(3,138) = 0.66$, $p = 0.58$). Thus, children's performance only marginally improved with age, even when the oldest and youngest age groups were analyzed together.

Did the average TF increase for each classifier type significantly exceed 0? If children understood the classifier, the post-classifier TF should be greater than the pre-classifier TF, thus leading to a cross-phase TF increase that was greater than 0. Since children's performance did not significantly differ across age groups, the three age groups' combined data were analyzed. Four one-sample *t*-tests examined whether the average TF increased significantly differed from 0. An adjusted significance cutoff level of 0.012 [0.05/4] was used in four-comparison analyses throughout this study. Results showed that the average TF increase significantly exceeded 0 for the animacy ($t(71) = 6.75$, $p < 0.001$), object shape ($t(71) = 5.49$, $p < 0.001$), and vehicle function classifiers ($t(71) = 2.98$, $p = 0.004$), but only marginally exceeded 0 for the configuration classifiers ($t(71) = 2.21$, $p = 0.03$) (Figure 1). Thus, children had reliable knowledge of animacy, object shape, and vehicle function classifiers, but their knowledge of configuration classifier was still emerging.

Did the TF increase for each classifier significantly exceed 0? To examine the children's comprehension of each classifier, the TF increase for each classifier was analyzed. Within each classifier type, four one-sample *t*-tests examined whether the TF increase significantly differed from 0 for each classifier. A significance cutoff level of 0.012 [0.05/4] was used. Among the animacy classifiers, the TF increase significantly exceeded 0 with three classifiers—*tóu* ($t(68) = 5.03$, $p < 0.001$), *zhī* ($t(68) = 4.93$, $p < 0.001$), and *wèi* ($t(68) = 3.52$, $p < 0.001$), but not with *míng* ($p = 0.27$) (Figure 2). Among the object shape classifiers, the TF increase significantly exceeded 0 with two classifiers—*tiáo* ($t(66) = 3.49$, $p < 0.001$) and *zhāng* ($t(67) = 3.25$, $p = 0.002$), but not with *lì* ($p = 0.09$) or *gēn* ($p = 0.20$). Among the vehicle function classifiers, the TF increase significantly exceeded 0 with *liàng* ($t(65) = 4.06$, $p < 0.001$) but not with *sōu* ($p = 0.12$), *jià* ($p = 0.58$), or *liè* ($p = 0.65$). Among the configuration classifiers, the TF increase significantly exceeded 0 with *qún* ($t(66) = 2.69$, $p = 0.009$) but not with *pái* ($p = 0.11$), *zhī* ($p = 0.83$), or *shuāng* ($p = 0.90$).

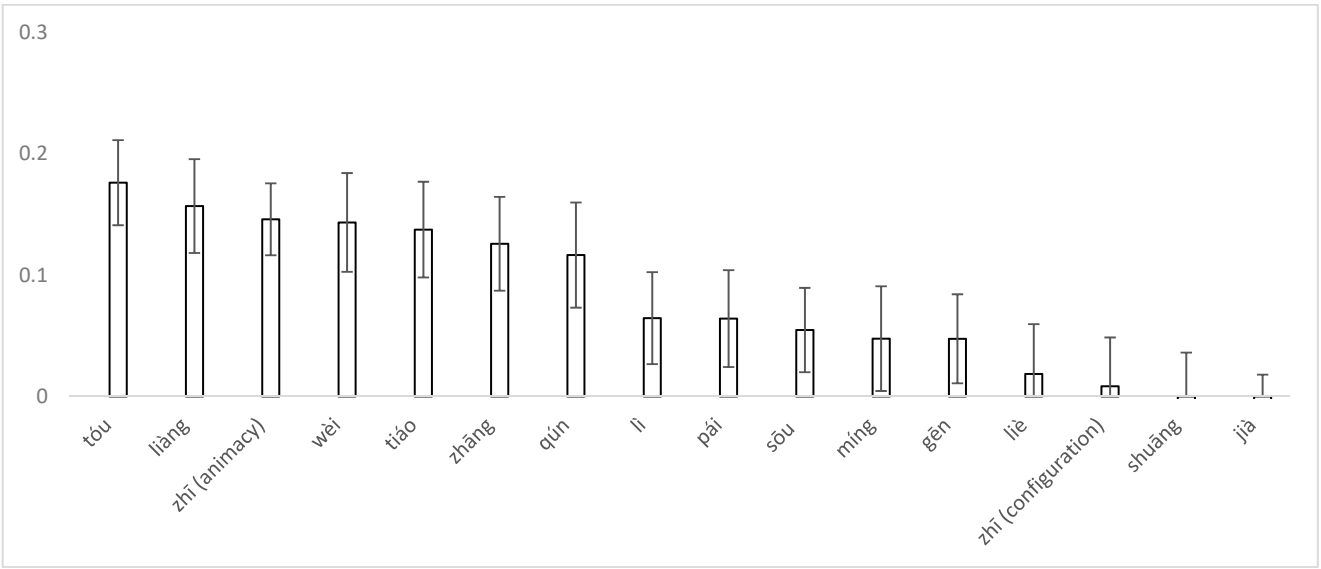


Figure 2. The TF increase for each classifier across all participants. The TF increase significantly exceeded 0 for *tóu*, *liàng*, *zhī* (animacy), *wèi*, *tiáo*, *zhāng*, *qún*, but not for *lì*, *pái*, *sōu*, *míng*, *gēn*, *liè*, *zhī* (configuration), *shuāng*, and *jià*. Error bars reflect SEM. The means (standard deviations) are provided here: *tóu*: 0.18 (0.29); *liàng*: 0.16 (0.31); *zhī* (animacy): 0.15 (0.25); *wèi*: 0.14 (0.34); *tiáo*: 0.14 (0.32); *zhāng*: 0.13 (0.32); *qún*: 0.12 (0.36); *lì*: 0.06 (0.31); *pái*: 0.06 (0.33); *sōu*: 0.05 (0.29); *míng*: 0.05 (0.36); *gēn*: 0.05 (0.28); *liè*: 0.02 (0.33); *zhī* (configuration): 0.008 (0.33); *shuāng*: −0.005 (0.35); *jià*: −0.02 (0.31).

4. Discussion

This study examined Mandarin-reared three-, four-, and five-year-olds' comprehension of four types of classifiers: animacy, configuration, vehicle function, and object shape classifiers. A within-subject design was used so that each child was tested on their knowledge of four types of classifiers. Unlike prior studies, this study was more controlled. We conducted three separate tasks with adults to be certain that any results could not be ascribed to unanticipated differences between the stimuli. First, the adults' familiarity ratings indicated that perceived familiarity with the target images did not differ across classifier types. Second, typicality ratings of the target images did not differ across classifier types, and finally, the visual similarity rating of the two images paired together did not differ across classifier types. These experimental controls afford us the opportunity to consider our findings relatively free from stimulus artifacts.

Question 1: Could children comprehend classifiers at ages three–five?

The one-sample *t*-tests showed that the average cross-phase TF increase significantly exceeded 0 for the animacy, object shape, and vehicle functional classifiers. Thus, children's overall performance was reliable with animacy, object shape, and vehicle function classifiers at the ages tested. This finding is also supported by the prior work revealing reliable comprehension of classifiers at ages three (Chien et al. 2003; Li et al. 2010) and four (Li et al. 2008) and Mandarin-speaking parental report of their children's vocabulary knowledge (Tardif et al. 2008). Notably, Fang (1985) did not find reliable knowledge of four object shape classifiers at age four, but Fang required that children's responses be correct in all three trials for a single classifier. In the current study and other research (Chien et al. 2003; Li et al. 2010), reliable knowledge of classifiers was defined as the accuracy rate that was significantly above chance in picture or object-choice tasks. In an IPLP study, since the visual stimuli within a pair are designed to be equated for attractiveness, children are not expected to exclusively look at the target even if they understand the meaning of the target words.

Question 2: Did the child's performance differ based on classifier type?

First, the ANOVA analysis on the average TF increase showed a significant main effect of classifier type and an insignificant age × classifier type interaction, suggesting that the child's performance differed by classifier types for all age groups. Second, paired sample *t*-tests on the average TF increase showed that the child's performance was better with animacy classifiers than with configuration and vehicle function classifiers. Third, one-sample *t*-tests on the average TF increase found that child comprehension was reliable for animacy, object shape, and vehicle function classifiers but not for configuration classifiers. Fourth, one-sample *t*-tests on the TF increase on each trial revealed reliable comprehension for 3/4 of the animacy classifiers (*tóu*, *zhī*, *wèi*) and 2/4 of the object shape classifiers (*tiáo*, *zhāng*), but only for 1/4 of the vehicle function classifiers (*liàng*) and 1/4 of the configuration classifier (*qún*), suggesting that vehicle function and configuration classifiers may be exceptionally hard to acquire, and that children did not respond in the same way to all four classifiers within a type.

Here, we propose three explanations for the learnability of classifiers. First, the learnability of classifiers may be related to children's early sensitivity to the semantic concepts encoded by the classifiers. Supporting this explanation is the current finding that children's overall performance was reliable for object shape and animacy classifiers. An important mechanism children use in word acquisition is *shape bias*, which is evident in typically developing children as early as 18 months of age (Landau et al. 1988; Graham and Poulin-Dubois 1999; Samuelson and Smith 2000; Perry and Samuelson 2011). Past research has observed the shape bias in children's learning of nouns (Smith 2000) and verbs (Golinkoff et al. 1996), suggesting that this bias may apply to the acquisition of words across form classes. This explanation is also supported by children's superior performance with animacy classifiers than with vehicle function and configuration classifiers and that

children's comprehension was reliable with $\frac{3}{4}$ of the animacy classifiers. The ability to distinguish between humans and animals occurs early in the first year of life. Even 3.5- and 6-month-olds prefer to attend to human beings than non-human primates (e.g., a gorilla or monkey—Heron-Delaney et al. 2011), and 3-month-old infants have categories of humans and non-human animals (e.g., cats and horses) (Quinn and Eimas 1998).

This study found that animacy classifiers had the highest average TF increase, followed by object shape classifiers. Note that the current finding is inconsistent with the past finding of Li et al. (2010). This cross-study difference should be related to the fact that Li et al. (2010) used nine object shape classifiers but only one animacy classifier, which might have enhanced the child's performance with object shape classifiers due to a practice effect. It is also possible that the finding of Li et al. (2010) is specific to that particular animacy classifier used. Nevertheless, the current finding suggests that the classifiers that are related to early word learning biases (i.e., object shape classifiers) may not be acquired earlier than the classifiers that are unrelated to these biases. The acquisition of classifiers may be affected by multiple factors beyond their relationship to early word-learning mechanisms. In addition, this study found that the child's performance was better with animacy classifiers than with vehicle function classifiers, thus supporting the possibility that the classifiers that are crucial for children's survival (i.e., animacy classifiers) tend to be acquired earlier than the classifiers that are less crucial for their survival (i.e., vehicle function classifiers). The high learnability of animacy classifiers may also arise from the fact that animacy classifiers encode a concept that infants become sensitive to early in life, and that is important for the survival of children. These two explanations are not mutually exclusive since children may become sensitive to a concept early in life because that concept is important for their survival.

Second, the learnability of classifiers may be related to the abstractness of the semantic concept encoded by the classifiers. To learn words in any language, children must first attend to and isolate a referent in their environment for the word, then abstract the commonalities shared by the instances labeled by that word, and then make word-referent mappings and extend the label to new, within-category exemplars (e.g., Golinkoff et al. 2002). This study found that configuration classifiers are exceptionally difficult to acquire. While an object shape classifier indicates the shape of *one* object, a configuration classifier indicates the spatial arrangement of *multiple* objects. The current finding supports the possibility that the classifiers that indicate a concept depicted by multiple objects (i.e., configuration classifiers) tend to be acquired later than the classifiers that indicate a concept depicted by a single object. Compared with an object shape classifier, children may have more difficulty searching for what perceptual feature the items in an arrangement have in common. Children may require repeated exposure to move their attention to the commonality independent of the specific objects (e.g., Maguire et al. 2008). Thus, abstracting a common feature across a smaller, visually similar set of exemplars (e.g., an object shape classifier) may be easier than that across a larger, visually variable set of exemplars (e.g., a configuration classifier)—a trend that may apply in word acquisition across form classes (e.g., Ma et al. 2009, 2022b). This explanation is also supported by the finding that Mandarin-reared children first associate specific classifiers with only prototypical exemplars labeled by nouns (Hu 1993)—a typicality effect that was also observed in Mandarin-reared children's meaning construal of familiar verbs (Ma et al. 2021). Arguably, it is easier to find commonalities from a set of prototypical exemplars labeled by a word than from a set of atypical exemplars—a semantic development trend observed in the acquisition of nouns, propositions, and verbs (Meints et al. 1999, 2002, 2008).

Third, the learnability of classifiers may be related to their input frequency in child-directed speech. The early-acquired classifiers tend to have high input frequency in child-directed speech. Supporting this explanation is the past finding that Chinese-reared children produce the generic classifier, *gè*, before specific classifiers (e.g., Hu 1993; Tse et al. 2007). A recent corpus study confirmed that *gè* is the most frequently used classifier in Mandarin child-directed speech (Ma et al. 2019) because *gè* is a generic classifier that can be

used productively with nouns. In addition, Mandarin-speaking parents also overuse *gè* in child-directed speech even when it is an inappropriate word choice[2], perhaps because parents tend to simplify the morphosyntactic structure of child-directed speech to facilitate children's language processing, thus further increasing the input frequency of *gè* in child-directed speech.

Children did not respond in the same way to all four classifiers within a type. For example, despite children's superior overall performance with animacy classifiers, children's comprehension of *míng* (indicating humans) was not reliable. According to an online corpus of modern Chinese text (Da 2004), *míng* has a lower input frequency than the other three animacy classifiers. Although the input frequency is calculated based on written text, it may reflect the input frequency of *míng* in child-directed speech. Children's comprehension of *lì* (indicating small, grain-like objects) was also not reliable, and it, too, had a lower input frequency than other object shape classifiers according to the corpus of modern Chinese text. However, since none of the existing Mandarin child-directed speech corpora offer input frequency data for all the 16 classifiers tested here, the effect of input frequency in a child-directed speech on classifier acquisition still requires further research. Furthermore, given that *lì* refers to a concept to which children are sensitive early on (i.e., shape) and indicates a narrow set of exemplars (i.e., a specific shape), the late age of acquisition of *lì* may be due to its low input frequency in child-directed speech. Perhaps, the acquisition of a classifier requires a minimum input frequency in child-directed speech—a threshold that the input frequency of *lì* does not meet.

Furthermore, the input frequency of classifiers is inevitably driven by that of the nouns they quantify. The current study observed reliable comprehension of the vehicle function classifier, *liàng*, which is used to quantify words like *chē* [automobile] and *qì-chē* [cars]—nouns that have high input frequency in Mandarin child-directed speech and are acquired early in life (Ma et al. 2019). Notably, the words (*huǒ-chē* [train]; *fēi-jī* [airplane]; *chuán* [ship]), which are quantified by the other three vehicle function classifiers (*liè* [quantifying trains], *sōu* [quantifying trains ships], *jià* [quantifying airplanes]), have lower input frequency than *chē* in child-directed speech (Ma et al. 2019). Thus, the high learnability of *liàng* may be a by-product of the learnability of the nouns that they quantify.

The three explanations may not be mutually exclusive. Perhaps, the classifiers that encode semantic concepts to which children have an early sensitivity, tend to indicate a narrow set of exemplars and be used more often by parents in child-directed speech. Perhaps, parents label these concepts to accommodate young children's limited cognitive abilities. It is also possible that objects quantified by these classifiers are readily available in infants' environment (e.g., balls, cars, humans, animals). The current design does not allow an estimate of the causal effect of these factors on classifier acquisition.

In this study, two classifiers of the same type were paired together on each trial. This design required children to decide between two classifiers of the same type—a stringent test of children's classifier knowledge. Better child performance may be observable if two classifiers of different types are paired together. For example, children may be able to find the target of *jià* (indicating an air-based vehicle) if an airplane is paired with an animal. However, this design can only reveal children's partial knowledge of a classifier (i.e., *jià* is used to quantify vehicle) rather than more complete knowledge of the classifier. Notably, the adults' data showed that on 97.5% of the 480 trials (16 trials × 30 participants), adults selected the assigned target image as the only image that could be labeled by the classifier in the two-option forced-choice task, thus verifying the assignment of classifiers to the targets. This finding also demonstrated that the acquisition of adult-like classifier knowledge is a lengthy process.

Question 3: Did the child's performance improve with age?

Neither the main effect of the age group nor the classifier x age group interaction was significant. Thus, this study revealed no conclusive evidence for an age-dependent improvement in the child's performance. However, some past studies showed a significant

age-dependent improvement in classifier knowledge between ages three–five (Chien et al. 2003), ages four–six (Li et al. 2008), and ages two–six (Li et al. 2010). There are two explanations for the cross-study differences. First, the cross-study differences might have arisen given that the selection of classifiers differed across studies. Some classifiers used in this study (e.g., the configuration and vehicle function classifiers) may be acquired after age five. Thus, a significant age-dependent improvement should not be evident between ages three and five. Some of the classifiers used here were not tested before, leaving the cross-study comparison hardly evaluable. Second, an age-dependent improvement in classifier knowledge may be more likely evident when a wider age range is included. This explanation is supported by the current finding that the analysis of the oldest (five-year-olds) and youngest (three-year-olds) age ranges revealed a marginally significant effect of age, thus revealing suggestive evidence for an age-dependent improvement. Note that some past studies did not reveal an age-dependent improvement in children's classifier comprehension either. For example, an age-dependent improvement was not observed between three- and five-year-olds in their comprehension of four object shape classifiers (Hu 1993), or between four- and five-year-olds in their comprehension of six object shape classifiers (Hao 2019). These findings demonstrated that the acquisition of classifiers is an extended process that continues beyond age five (Li et al. 2008, 2010). Nevertheless, even six-year-olds have not mastered the distinction between count and mass classifiers (Li et al. 2008). Note that this study did not use modern eye-tracking devices. Future research should further examine the acquisition of classifiers using modern eye-tracking devices.

This study found that the child's performance differed across classifiers. However, the meaning of Mandarin classifiers can be complex. For example, the object shape classifier, *tiáo* can quantify not only long and slender objects but also animals (e.g., *yì tiáo gǒu* [a dog]; *yì tiáo yú* [a fish]), humans in idiomatic usage (e.g., *yì tiáo hǎo-hàn* [a brave man]), and even abstract concepts (e.g., *yì tiáo xiāoxī* [a piece of news]). This study tested children's comprehension of classifiers in a visual fixation task using two images. Children's apparent use of a classifier in this study by no means indicates a full mastery of its meaning. Classifiers are difficult to acquire.

5. Conclusions

This study was the first to examine three-, four-, and five-year-old Mandarin-reared children's comprehension of four types of classifiers when various factors were matched across classifier types (i.e., the number of classifiers, the perceived familiarity and typicality of the target images, and the visual similarity of the two images paired together on a trial). Reliable classifier knowledge was observed at ages three–five. Performance differed across classifier types. However, this study did not reveal conclusive evidence for age-dependent improvements in a child's performance.

Author Contributions: Conceptualization, W.M. and P.Z.; project administration, W.M. and P.Z.; data analysis: W.M., P.Z. and R.M.G.; writing, W.M. and R.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: P.Z. is supported by a National Natural Science Grant (U20B2062) by National Natural Science Foundation of China.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of the University of Arkansas (protocol code 1710079596, date of approval: 11/22/2017).

Informed Consent Statement: Parental informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available upon request to the first author.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] The strike above a vowel denotes the lexical tone of the Mandarin Chinese character. There are four basic tones in Mandarin. Tones are distinctive features in Mandarin. For example, *mā* with a high, level tone means mother; *má* with a rising tone means numb; *mǎ* with a dipping tone means horse; and *mà* with a falling tone means to curse.

[2] We also examined a subset of the early-acquired nouns that were analyzed by Ma et al. (2019). Only words that occurred with classifiers at least once in the CHILDES Beijing Corpus were analyzed. The final sample contained 29 nouns, 28 of which occurred with the generic classifier (*gè*), and one occurred only with specific classifiers. We first asked adult participants to list the most appropriate classifiers that could modify these nouns and then to judge the appropriateness of the use of *gè* in child-directed speech. To rule out possible influences between tasks, two separate groups of Mandarin-speaking undergraduate students (majors in disciplines other than linguistics) in China participated. For the fill-in-the-blank task, 24 participants were asked to provide the *most* appropriate classifier for each of the 29 nouns, based on the following instruction written in Chinese: "Please write down the most appropriate classifier to fill in the blank. For example, in *yí* [one] ___ *rén* [people], you may write down *gè* in the blank if you believe that it is the most appropriate classifier to fill in the blank based on your knowledge of Chinese." For each noun, we calculated the rate of *gè* use by dividing the token frequency of *gè* over the total token frequency of classifiers (combining the generic and specific classifiers). The mean rate of *gè* use was 0.76 (SD = 0.33) across the 29 nouns. A one-sample t-test comparing the mean rate of *gè* against chance (0.5; generic vs. specific) found that it was significantly above chance ($t(28) = 4.16$, $p < 0.001$), suggesting that caregivers were more likely to use *gè* than specific classifiers. Is *gè* the most appropriate classifier to modify these nouns? The result showed that for only 7 (24% of 29) nouns, *gè* was listed (among specific classifiers) as the most appropriate classifier. A sign test showed that across the 29 nouns, *gè* was less likely to be the most appropriate classifier than specific classifiers ($p < 0.01$). For each noun, we divided the responses of *gè* being listed as the most appropriate classifier over the total number of responses (n = 24). A one-sample t-test comparing the mean rate of *gè* being listed as the most appropriate classifier against chance (0.5; generic vs. specific) showed that it was significantly below chance ($t(28) = 5.08$, $p < 0.001$, Cohen's d = 2.07). Thus, *gè* is less likely to be the most appropriate classifier than specific classifiers to modify the 29 nouns selected. Then, the use of *gè* was further analyzed. Another 24 undergraduate students rated the appropriateness of the phrases containing *gè* and each of the 28 nouns that appeared with *gè* at least once in CHILDES Beijing Corpus. The task consisted of 28 items, each presenting a phrase in the form of "*yí* [one] *gè* [classifier] noun." The instruction was presented in Chinese, "Please judge whether the following phrases are good examples of Chinese based on a scale from 1 to 9, where 1 means very bad, 3 means bad, 5 means not bad or not good, 7 means good, 9 means very good, and the numbers in between mean moderately good or bad. Please focus on the use of classifiers during the task." This analysis focused on the 28 nouns that appeared with *gè* at least once in the child-directed speech corpus. The use of *gè* received a mean appropriateness rating of 3.98 (SD = 2.59). Since the fill-in-the-blank task showed that 7 of the nouns could most appropriately occur with *gè*, and 21 of them could only occur with specific classifiers, the two types of words were analyzed separately. We found that the 7 nouns received a rating of 7.98 (SD = 0.51), verifying the results of the fill-in-the-blank task. The 21 nouns received a rating of 2.65 (SD = 1.22), suggesting that *gè* was used inappropriately with these nouns based on Mandarin grammar. Taken together, these findings show that Mandarin-speaking caregivers tend to overuse the generic classifier—*gè*, even when it is inappropriate.

References

Allan, Keith. 1977. Classifiers. *Language* 53: 285–311. [CrossRef]

Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Cheng, Lisa L. S., and Rint Sybesma. 1998. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies* 28: 385–412.

Cheng, Lisa L. S., and Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30: 509–42. [CrossRef]

Chien, Yu-Chin, Barbara Lust, and Chi-Pang Chiang. 2003. Chinese children's comprehension of count-classifiers and mass-classifiers. *Journal of East Asian Linguistics* 12: 91–120. [CrossRef]

Cooper, Robin P., and Richard N. Aslin. 1990. Preference for infant-directed speech in the first month after birth. *Child Development* 61: 1584–95. [CrossRef]

Cycowicz, Yael M., David Friedman, Mairay Rothstein, and Joan Gay Snodgrass. 1997. Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 65: 171–237. [CrossRef]

Da, Jun. 2004. A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In *Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese*. Edited by Pu Zhang, Tianwei Xie and Juan Xu. Beijing: Tsinghua University Press, pp. 501–11.

Diesendruck, Gil, Lori Markson, and Paul Bloom. 2003. Children's reliance on creator's intent in extending names for artifacts. *Psychological Science* 14: 164–68. [CrossRef]

Erbaugh, Mary S. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. In *Noun Classes and Categorization*. Edited by Colette G. Craig. Amsterdam and Philadelphia: John Benjamins, pp. 399–436.

Erbaugh, Mary S. 2006. Chinese classifiers: Their use and acquisition. In *Handbook of East Asian Psycholinguistics: Chinese*. Edited by Ping Li, Lihai Tan, Elizabeth Bates and Ovid J. L. Tzeng. Cambridge: Cambridge University Press, pp. 39–51.

Fang, Fuxi. 1985. An experiment on the use of classifiers by 4- to 6-year-olds. *Acta Psychologica Sinica* 17: 384–92.

Faul, Franz, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39: 175–91. [CrossRef]
Fernald, Anne. 1985. Four-month-olds prefer to listen to motherese. *Infant Behavior and Development* 8: 181–95. [CrossRef]
Gleitman, Lila R., and John C. Trueswell. 2020. Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science* 12: 22–47. [CrossRef] [PubMed]
Golinkoff, Roberta Michnick, He Len Chung, Kathy Hirsh-Pasek, Jing Liu, Bennett I. Bertenthal, Rebecca Brand, Mandy J. Maguire, and Elizabeth Hennon. 2002. Young children can extend motion verbs to point-light displays. *Developmental Psychology* 4: 604–15. [CrossRef] [PubMed]
Golinkoff, Roberta Michnick, Roberta Jacquet, Kathy Hirsh-Pasek, and Ratna Nandakumar. 1996. Lexical principles may underlie the learning of verbs. *Child Development* 67: 3101–19. [CrossRef] [PubMed]
Golinkoff, Roberta M., Weiyi Ma, Lulu Song, and Kathy Hirsh-Pasek. 2013. Twenty-five years using the intermodal preferential looking paradigm to study language acquisition: What have we learned? *Perspectives on Psychological Science* 8: 316–39. [CrossRef]
Gonzalez-Gomez, Nayeli, Silvana Poltrock, and Thierry Nazzi. 2013. A "bat" is easier to learn than a "tab": Effects of relative phonotactic frequency on infant word learning. *PLoS ONE* 8: e59601. [CrossRef]
Graham, Susan A., and Diane Poulin-Dubois. 1999. Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language* 26: 295–320. [CrossRef]
Hao, Ying. 2019. How Do Mandarin-Speaking Children Learn Shape Classifiers? Ph.D. dissertation, The University of Texas at Austin, Austin, TX, USA.
Heron-Delaney, Michelle, Sylvia Wirth, and Olivier Pascalis. 2011. Infants' knowledge of their own species. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366: 1753–63. [CrossRef]
Hollich, George. 2005. *Supercoder: A Program for Coding Preferential Looking*. Version 1.5, Computer Software. West Lafayette: Purdue University.
Hollich, George, Roberta M. Golinkoff, and Kathy Hirsh-Pasek. 2007. Young children associate novel words with complex objects rather than salient parts. *Developmental Psychology* 43: 1051–61. [CrossRef]
Horst, Jessica S. 2009. Novel Object and Unusual Name (NOUN) Database. Available online: http://www.sussex.ac.uk/wordlab/noun (accessed on 9 January 2017).
Hu, Qian. 1993. The Acquisition of Chinese Classifiers by Young Mandarin-Speaking Children. Ph.D. dissertation, Boston University, Boston, MA, USA.
Killingley, Siew-Yue. 1983. *Cantonese Classifiers: Syntax and Semantics*. Newcastle-upon-Tyne: Grevatt and Grevatt.
Landau, Barbara, Linda B. Smith, and Susan S. Jones. 1988. The importance of shape in early lexical learning. *Cognitive Development* 3: 299–321. [CrossRef]
Li, Peggy, Becky Huang, and Yaling Hsiao. 2010. Learning that classifiers count: Mandarin-speaking children's acquisition of sortal and mensural classifiers. *Journal of East Asian Linguistics* 19: 207–30. [CrossRef]
Li, Peggy, David Barner, and Becky H. Huang. 2008. Classifiers as count syntax: Individuation and measurement in the acquisition of Mandarin Chinese. *Language Learning and Development* 4: 249–90. [CrossRef] [PubMed]
Loke, K. K. 1991. A Semantic Analysis of Young Children's Use of Mandarin Shape Classifiers. In *Child Language Development in Singapore and Malaysia*. Edited by Anna Kwan-Terry. Singapore: Singapore University Press, pp. 98–116.
Ma, Weiyi, and Peng Zhou. 2019. Three-year-old tone language learners are tolerant of tone mispronunciations spoken with familiar and novel tones. *Cogent Psychology* 6: 1690816. [CrossRef]
Ma, Weiyi, Anna Fiveash, Elizabeth Margulis, Douglas Behrend, and Willian Forde Thompson. 2020a. Song and infant-directed speech facilitate word learning. *Quarterly Journal of Experimental Psychology* 73: 1036–54. [CrossRef] [PubMed]
Ma, Weiyi, Peng Zhou, and Roberta M. Golinkoff. 2020b. Young Mandarin learners use function words to distinguish between nouns and verbs. *Developmental Science* 23: e12927. [CrossRef] [PubMed]
Ma, Weiyi, Peng Zhou, and William Forde F Thompson. 2022a. Children's decoding of emotional prosody in four languages. *Emotion* 22: 198–212. [CrossRef] [PubMed]
Ma, Weiyi, Peng Zhou, Leher Singh, and Liqun Gao. 2017. Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition* 159: 139–55. [CrossRef] [PubMed]
Ma, Weiyi, Peng Zhou, Roberta M. Golinkoff, Joanne Lee, and Kathy Hirsh-Pasek. 2019. Syntactic cues to the noun and verb distinction in Mandarin child-directed speech. *First Language* 39: 433–61. [CrossRef]
Ma, Weiyi, Roberta M. Golinkoff, Derek Houston, and Kathy Hirsh-Pasek. 2011. Word learning in infant-and adult-directed speech. *Language Learning and Development* 7: 209–25. [CrossRef]
Ma, Weiyi, Roberta M. Golinkoff, Kathy Hirsh-Pasek, Colleen McDonough, and Twila Tardif. 2009. Imageability predicts the age of acquisition of verbs in Chinese. *Journal of Child Language* 36: 405–23. [CrossRef]
Ma, Weiyi, Roberta M. Golinkoff, Lulu Song, and Kathy Hirsh-Pasek. 2021. Using verb extension to gauge children's verb meaning construals: The case of Chinese. *Frontiers in Psychology* 11: 572198. [CrossRef]
Ma, Weiyi, Rufan Luo, Robert M. Golinkoff, and Kathy Hirsh-Pasek. 2022b. The influence of exemplar variability on young children's construal of verb meaning. *Language Learning and Development*. (*advance online publication*).
Maguire, Mandy J., Kathy Hirsh-Pasek, Roberta M. Golinkoff, and Amanda C. Brandone. 2008. Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science* 11: 628–34. [CrossRef]

Mani, Nivedita, and Kim Plunkett. 2007. Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language* 57: 252–72. [CrossRef]
Meints, Kerstin, Kim Plunkett, and Paul L. Harris. 1999. When does and ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology* 35: 1072–78. [CrossRef] [PubMed]
Meints, Kerstin, Kim Plunkett, and Paul L. Harris. 2008. Eating apples and houseplants: Typicality constraints on thematic roles in early verb learning. *Language and Cognitive Processes* 23: 434–63. [CrossRef]
Meints, Kerstin, Kim Plunkett, Paul L. Harris, and Debbie Dimmock. 2002. What is 'on' and 'under' for 15-, 18-and 24-month-olds? Typicality effects in early comprehension of spatial prepositions. *British Journal of Developmental Psychology* 20: 113–30. [CrossRef]
Nelson, Katherine. 1988. Constraints on word learning? *Cognitive Development* 3: 221–46. [CrossRef]
Perry, Lynn K., and Larissa K. Samuelson. 2011. The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology* 2: 345. [CrossRef]
Quam, Carolyn, and Dniel Swingley. 2010. Phonological knowledge guides two-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language* 62: 135–50. [CrossRef]
Quinn, Paul C., and Peter D. Eimas. 1998. Evidence for a global categorical representation of humans by young infants. *Journal of Experimental Child Psychology* 69: 151–74. [CrossRef]
Samuelson, Larissa K., and Linda B. Smith. 2000. Children's attention to rigid and deformable shape in naming and non-naming tasks. *Child Development* 71: 1555–70. [CrossRef]
Singh, Leher, Hwee Hwee Goh, and Thilanga D. Wewalaarachchi. 2015. Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition* 142: 1–11. [CrossRef] [PubMed]
Singh, Leher, Tam Jun Hui, Calista Chan, and Roberta M. Golinkoff. 2014. Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science* 17: 94–109. [CrossRef] [PubMed]
Smith, Linda B. 2000. Learning how to learn words: An associative crane. In *Breaking the Word Learning Barrier: What Does It Take?* Edited by Roberta M. Golinkoff and Kathy Hirsh-Pasek. New York: Oxford Press, pp. 51–80.
Srinivasan, Mahesh, and Jesse Snedeker. 2014. Polysemy and the taxonomic constraint: Children's representation of words that label multiple kinds. *Language Learning and Development* 10: 97–128. [CrossRef]
Swingley, Daniel, and Richard N. Aslin. 2000. Spoken word recognition and lexical representation in very young children. *Cognition* 76: 147–66. [CrossRef] [PubMed]
Swingley, Daniel, and Richard N. Aslin. 2002. Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science* 13: 480–84. [CrossRef]
Tai, J. H. 1994. Chinese classifier systems and human categorization. In *In Honor of William S.-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Nottingham: Pyramid Press, pp. 479–94.
Tardif, Twila, Paul Fletcher, Zhixiang Zhang, Weilan Liang, and Q. H. Zuo. 2008. *The Chinese Communicative Development Inventory (Putonghua and Cantonese Versions): Manual, Forms, and Norms*. Beijing: Peking University Medical Press.
Tse, Shek Kam, Hui Li, and Shing On Leung. 2007. The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language* 34: 495–517. [CrossRef]
Werker, Janet F., Judith E. Pegg, and Peter McLeod. 1994. A cross-language comparison of infant preference for infant-directed speech: English and Cantonese. *Infant Behavior and Development* 17: 321–31. [CrossRef]
White, Katherine S., and Richard N. Aslin. 2011. Adaptation to novel accents by toddlers. *Developmental Science* 14: 372–84. [CrossRef]
Ying, Houchang, Guopeng Chen, Zhengguo Song, Weiming Shao, and Ying Guo. 1983. 4–7 Sui Ertong Zhangwo Liangci De Tedian [Characteristics of 4-to-7-year-olds in Mastering Classifiers]. *Information on Psychological Sciences* 26: 24–32.
Zhang, Hong. 2007. Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics* 16: 43–59. [CrossRef]
Zhou, Peng, and Weiyi Ma. 2018. Children's use of morphological cues in real-time event representation. *Journal of Psycholinguistic Research* 47: 241–60. [CrossRef] [PubMed]