The Dissertation Committee for Ying Hao certifies that this is the approved version of the following dissertation:

How do Mandarin-Speaking Children Learn Shape Classifiers?

Committee:

Chang Liu, Supervisor

Lisa M. Bedore

Li Sheng

Elizabeth D. Peña

Peng Zhou

How do Mandarin-Speaking Children Learn Shape Classifiers?

by

Ying Hao

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2019

Dedication

In loving memory of my grandpa, who was excited about my PhD application but passed away in an accident two months before my admission to the University of Texas at Austin. Despite all the hardship in life, you have been such a warm, considerate, trustworthy and humble person. You are an inspiration for my entire life. I think of you fondly, and I miss you more than words can say……

Acknowledgements

I would love to thank my advisors Dr. Lisa Bedore and Dr. Li Sheng. During the past five years, you have taught me so much: research methods, research proposal, paper writing, grant writing, presentation skills, teaching skills, mentoring skills……. Thank you for being patient. Thank you for allowing me to make mistakes and grow. Thank you for having faith in me to be an independent researcher. You are great models that I will always look up to.

I would love to thank Dr. Chang Liu, who took over the supervision during the last year of my PhD training. You are very generous to give me great suggestions for dissertation. I know it has been a hectic year for you, and I appreciate your time and efforts. I would love to thank Dr. Elizabeth Peña. I still remember that you told me "it is all about being persistent" when our paper about narrative skills in Mandarin-English bilingual paper got rejected. Your words meant a lot to me. I would also love to thank my *Dashixiong*, Dr. Peng Zhou. Thank you for helping me recruit participants in Beijing. You are a great powerhouse and a strong researcher, and my initial motivation of doing research is you. Thank you for your invaluable advice in different phases of my life. Believe it or not, we have been knowing each other for ten years!

I appreciate many others who helped me finish the dissertation: teachers, parents and children in Yueer Education in Beijing, research assistants from Tsinghua University, my husband Ju Gao, people in the HABLA lab (Javier, Stephanie, Kathleen……), people in the LLB lab (Danyang, Huanhuan, Sharon……), Liesbeth, Jesse Yang, my friend Victoria Li, my cousin Fanli Ji……

Many thanks to my parents for being supportive for all my crazy decisions in life: studying linguistics, working in India, studying in the US at an age when all my friends were settling down, traveling in France alone…… Thank you for being generous, considerate and diligent yourself. I hope I am a good reflection of you.

Abstract

How do Mandarin-Speaking Children Learn Shape Classifiers?

by

Ying Hao, Ph.D.

The University of Texas at Austin, 2019

Supervisor: Chang Liu

Mandarin is a classifier language. A classifier is inserted between a number and a noun for the purpose of quantification (e.g., 一**条**绳子 one **tiáo** rope). Each classifier marks semantic characteristics of the noun with which it co-occurs (e.g., 条 tiáo is typically paired with long, narrow and flexible objects). The semantic system of classifiers is complex, and classifier production is a vulnerable area for Mandarin-speaking children (e.g., Hao et al., 2018). However, it is unclear what learning mechanisms drive the acquisition of classifiers in Mandarin-speaking children. In the present study, we explored potential predictors, namely classifier-based semantic categorization and input frequency of classifiers. In addition, we hypothesized that existing vocabulary knowledge would be related to classifier learning.

Sixty-four typically-developing monolingual Mandarin-speaking children between 4;1 (year;month) and 6;5 completed two background tasks and two experimental tasks. The background tasks consisted of an object categorization task to index semantic categorization strategy, and a picture selection task and a picture naming task to measure vocabulary knowledge. In experiment 1, we implemented a comprehension and a production task for six real classifiers that emphasize shape. In experiment 2, we administered a learning task for two novel classifiers which encoded different semantic properties (i.e., curly-haired vs. broken). Frequency of classifier input was manipulated using a between-subject design.

We analyzed contributions of classifier-based semantic categorization, input frequency of classifiers, and vocabulary knowledge to classifier comprehension and production. Even though children preferred to categorize objects by shared classifier than by other semantic links in the object categorization task, this preference did not significantly predict real classifier comprehension and production in experiment 1. At the same time, vocabulary knowledge was a significant predictor for both. Children may find that the semantic system of real Mandarin classifiers is opaque, and they rely more heavily on an item-by-item learning approach that is used in vocabulary learning (i.e., idiosyncratic learning of individual words). In addition, children showed varied accuracy on different classifier-noun pairs for the same classifier, providing more evidence for item-based learning. For novel classifier learning in experiment 2, classifier-based object categorization was a marginally significant predictor for comprehension. The higher frequency group did not outperform the lower frequency group, and vocabulary knowledge was a significant predictor for neither comprehension nor production. These findings suggest that children mainly took a rule-based approach to learn novel classifiers with transparent semantic categorization. Overall, results from the two experiments showed that the learning approach children primarily use to learn classifiers depend on the transparency of the classifier system.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

STUDY MOTIVATION

Classifiers are a common phenomenon in a number of languages in Asia, Africa, Oceania and America (Allan, 1977). Mandarin is one of the most widely spoken classifier languages (Chien, Lust, & Chiang, 2003). The semantic system of classifiers is quite complex and difficult for children learning a classifier language (e.g., Ying, Chen, Song, Shao, & Guo, 1983; Salehuddin & Winskel, 2009).

Classifiers are a potential clinical marker to identify Mandarin-speaking children with developmental language disorder (DLD) (Cheung, 2009; Hao et al., 2018). DLD is a significant disorder in speaking and understanding language that cannot be attributed to hearing loss, low nonverbal intelligence, or neurological disorders (Leonard, 2014). It negatively influences children's literacy, mathematical thinking, and social interactions (Justice, Bowles, Pence Turnbull, & Skibbe 2009). Given a 7% prevalence rate of DLD in children (Tomblin et al., 1997) and the belief that DLD is equally prevalent across languages and cultures (Armon-Lotem, de Jong & Meir, 2015), it is estimated that in China there are about 5 million Mandarin-speaking children with DLD between ages four and nine (National Bureau of Statistics of China, 2010). Research also showed that classifiers are a vulnerable area for Mandarin-English bilingual children in the US (Hao, Bedore, Sheng, & Peña, 2018). In the US, Chinese is the next most frequently spoken minority language behind Spanish, with about three million speakers (U.S. Census Bureau, 2011). In both countries, in order to better assess and treat Mandarin-speaking children with DLD, there are pressing clinical needs to understand children's learning of this linguistic feature.

To the best of our knowledge, there are few studies about classifier learning (e.g., Chien et al., 2003; Li Barner, & Huang, 2008). Of these, even fewer have conducted in-depth investigations about how children learn classifiers (e.g., Uchida & Imai, 1999). It is still unclear what underlying mechanisms drive the learning and what influential factors are relevant. A study in typically-developing (TD) monolingual Mandarin-speaking children prepares future investigations in monolingual and bilingual Mandarin-speaking children with DLD. Moreover, the investigation of Mandarin classifiers provides new insights into the usage-based model of language acquisition by extending the focus from syntactic structures to semantic categorization. Classifiers are an interface feature, and the semantic aspect has been found to be more challenging than the syntactic constructions. The theoretical contribution of this study will be expanded later in the dissertation.

In the current study, we intended to conduct an in-depth investigation on how TD monolingual Mandarin-speaking children learn classifiers by exploring potential predictors. The organization of Chapter 1 is as follows. First, we introduced Mandarin as a classifier language. Then, we summarized typical stages of classifier development and findings on classifier comprehension and production in TD monolingual Mandarin-speaking children. Next, extant literature on potential predictors of classifier learning was reviewed. Lastly, based on the literature review, we specified our research questions and predictions.

MANDARIN AS A CLASSIFIER LANGUAGE

Mandarin is a classifier language. A classifier is placed between a number and a noun for the purpose of quantification. See (1). Another usage is to specify an object, in which a classifier is inserted between a demonstrative and a noun. See (2). In addition to quantify objects, some Mandarin classifiers are used to quantify actions. See (3). Nevertheless, the majority of Mandarin classifiers are noun classifiers (Erbaugh, 2006). A classifier is obligatory in the classifier construction, and all the illustrated usages are ungrammatical if the classifiers are omitted[1].

(1) 三 本 书

sān běn shū

three classifier-ben book

three books

(2) 这 本 书

zhè běn shū

this classifier-ben book

this book

(3) 去了一次北京

qù le yí cì Běijīng

go perfective-aspect one classifier-ci Beijing

went to Beijing once

[1] Based on personal observation, classifiers may be omitted in very colloquial usage in Beijing dialect when the number is equal or less than three. When a classifier is omitted, the phonological forms of the number are changed accordingly.

Classifiers are a syntactic-semantic interface phenomenon. In addition to being a syntactic feature, classifiers indicate semantic properties of an entity (e.g., animacy, shape, function and size) (Allan, 1977; Li & Thompson, 1981). For example, "条 tiáo" is a shape classifier that is typically paired with long, narrow and flexible objects. In example (4), (5) and (6), "条 tiáo" is used with *snake*, *rope* and *line* which share these semantic properties. The pairing of a classifier and an object is not a strict one-to-one correspondence. Another classifier, "根 gēn", can be paired with some objects that are paired with "条 tiáo". "根 gēn" is a semantically similar classifier, which is typically used with long, narrow but rigid objects (e.g., *stick*, *chopstick*, *pillar, cucumber, eggplant*). However, "根 gēn" can be paired with *rope* in (5) and *line* in (6) but not *snake* in (4). These exceptions are accepted as conventions by native speakers of Mandarin. In addition to specific classifiers, there is a general classifier "个 gè" which is more frequently used than specific classifiers (Erbaugh, 1986). See (7).

(4) 两 条 蛇

liǎng tiáo shé

two classifier-tiao snake

two snakes

(5) 三 条 绳子

sān tiáo shéngzi

three classifier-tiao rope

three ropes

(6) 四 条 线

sì tiáo xiàn

four classifier-tiao line

four lines

(7) 一 个 人

yí gè rén

one classifier-ge person

one person

Erbaugh's (2006) classification scheme best captured the full range of classifier meanings. She identified five types of classifiers[2]: measure classifiers (一**公里**路 one **kilometer** of road), collective classifiers (一**扎**啤酒 one **bundle** of beer), kind classifier (一**种**水果 one **type** of fruit), event classifiers (一**场**电影 one **performance** of movie) and sortal classifiers (一**条**蛇 one **tiáo** snake). She pointed out that perhaps every language has vocabulary to express concepts of measure, collection, kind, event but not sortal. For example, English measure words are similar to Mandarin measure classifiers in that both relate to length, weight, or volume (e.g., 两**磅**苹果 two **pounds** of apples, 三**升**水 three **liters** of water).

However, the sortal classifier, which mainly focuses on shape and size, is a strikingly distinctive feature in Mandarin that is not commonly present in non-classifier languages. Sortal classifiers in Mandarin provide unique insights into semantic organization of nouns by Mandarin speakers. Also, sortal classifiers appear to be more prototypical than other types of classifiers for native speakers of Mandarin. When adult Mandarin speakers were asked to list as many classifiers as they could, they listed mostly sortal classifiers (Chien et al., 2003). In the present study, we focused on sortal classifiers that emphasize semantic features of shape.

TYPICAL STAGES OF CLASSIFIER DEVELOPMENT

As classifiers are an interface phenomenon of syntax and semantics, our review of typical development trajectory is in both domains. The syntactic structure of classifiers is simple. Classifier constructions strictly follow the word order of "number classifier noun", as shown in the previous examples. Young Mandarin-speaking children, as early as age four, consistently produce this structure (Fang, 1985; Ying et al., 1983; Uchida & Imai, 1999). Tse, Li, and Leung (2007) analyzed language samples in free play[3], and found that none of the 492 Cantonese-speaking children omitted classifiers when classifier constructions were used. Children did, however, use inappropriate classifiers (mostly the general classifier "个 gè") to fill the classifier position.

In contrast, the semantic aspect of classifiers is challenging for young children. In Mandarin, each classifier encodes different semantic features. Mandarin-speaking children probably need to be exposed to multiple exemplars of a classifier before they can summarize

[2] There are other ways to divide Mandarin classifiers. See Chien et al. (2003) and Li et al. (2008).

[3] The play happened in a classroom, which was furnished with a set of toys, including cooking materials, food and fruit, furniture and electrical appliances, hospital materials and vehicles. The authors believed that these activities could represent different settings in children's daily life.

the corresponding semantic features for different classifiers. Given the complex classifier semantic system to be learned, it has been proposed that knowledge of classifiers may be related to skills of semantic categorization and synthetization (Uchida & Imai, 1999)[4].

Focusing on semantic organization, Uchida and Imai (1999) proposed three stages to describe typical development of classifiers. In the first stage, children under age three are unaware of the semantic features of classifiers, as nouns (not classifiers) carry more critical information in communication. Children learn classifier and noun pairs as unanalyzed lexical units. In the second stage, children are able to summarize and categorize some semantic features corresponding to several frequent classifiers. Yet, they are far away from fully mastering the semantic system of classifiers. At this stage, children may substitute a target classifier using an early-acquired classifier that is semantically similar. Alternatively, the general classifier "个 gè" is widely overgeneralized as a place holder. In the final stage, accompanied by increased ability to synthesize fragmented semantic knowledge into coherent rules, children at age six are able to distinguish a variety of different specific classifiers. Their classifier usage becomes more diverse and includes fewer errors. Overall, this proposal emphasizes the role of semantic categorization in classifier development.

CLASSIFIER COMPREHENSION AND PRODUCTION

Three previous studies documented the comprehension of classifiers in monolingual Mandarin-speaking children who were presumed to be typically developing (Chien et al., 2003; Li, Huang, & Hsiao, 2010; Sera, Johnson, & Kuo, 2013). They conducted cross-sectional studies using a forced-choice object/picture selection task. Children were presented with three objects/pictures and were asked to select one. Among the three objects/pictures (e.g., *wallet, snake, schoolbag*), only one (i.e., *snake*) could be paired with the targeted classifier (e.g., 条 tiáo) in a carrier phrase (e.g., 米老鼠说他想要一条 something. *Mickey Mouse says that he wants one tiáo something.*).

Results from the three studies converged in showing that the comprehension of classifiers increased with age. Accuracy in 2-year-old Mandarin-speaking children was below chance on most specific classifiers examined, suggesting that children lacked the understanding of specific classifiers at this age. From three to five years of age, children demonstrated steady

[4] In section 5, we will review other proposals of classifier learning. Here, we focus on Uchida and Imai (1999)'s proposal.

increase from chance-level to around 80% accuracy. Children older than six demonstrated ceiling performance, indicating the mastery of specific classifiers in comprehension.

The three studies did not manipulate the three selections of objects in each test trial. Some objects were perceptually-similar to target objects (e.g., pencil and rope), which should be paired with semantically-similar classifiers to the target classifiers. Other selections of objects were perceptually-distinct from target objects (e.g., mirror and rope), and they should be paired with semantically-distant classifiers to the target classifiers. This caused varied difficulty level across different trials. Because of this, the dominant type of errors in comprehension were unknown, but such information could provide insights into children's classifier development. This is discussed in more details later, as it pertains to "semantic relatedness".

To examine classifier production, Ying et al. (1983) and Fang (1985) asked Mandarin-speaking children to complete a counting task. Children were presented with pictures depicting sets of objects in different quantities. As classifiers are obligatory in classifier constructions, children had to include classifiers. Similar to comprehension, production accuracy steadily increased as children grew older. In Ying et al. (1983), four-year-old children achieved 17% accuracy and only a few specific classifiers were appropriately used. These young children widely overgeneralized the general classifier "个 gè" to occasions when specific classifiers should be used. Between the ages of five and six, accuracy increased to around 60%, but children still produced some inappropriate pairings of classifiers and nouns. For example, some children used verbs that were typically paired with the nouns to replace the target specific classifiers (e.g., "一**开**汽车" *one **drive** car* for "一**辆**汽车" one **liàng** car). For another example, the classifier "辆 liàng" should be paired with cars but not planes and ships. Some children used it for all vehicles. Such errors indicated that children were exploring accurate semantic properties of classifiers, but perhaps more exposure was needed to extract the exact semantic representations. It was not until age seven that children's accuracy of classifier usage reached 82%, indicating improved but not perfect mastery. Even though Ying et al. (1983) and Fang (1985) provided some examples of children's errors, these studies did not provide a systematic analysis of errors that could inform the learning process.

It is important to note that the production of classifiers emerged later than comprehension. While six-year-olds were almost at ceiling on classifier comprehension (Chien et al., 2003; Li et al., 2010; Sera et al., 2013), children at the same age were around 60% accurate in production (Ying et al., 1983; Fang, 1985). Extended development of classifier production has been found in children speaking other classifier languages. Salehuddin and Winskel (2009) found that 9-year-old Malay-speaking children only correctly used classifiers half of the time in a counting task. These children tended to substitute later-acquired classifiers with earlier-acquired classifiers that were more frequent in their input.

POTENTIAL PREDICTORS OF CLASSIFIER LEARNING

The studies above highlight the developmental sequence of classifiers and illustrate that classifiers are acquired over an extended period of time. Uchida and Imai (1999) posit that classifier-based semantic categorization is an important mechanism in acquisition. However, there are other potential predictors of classifier learning. In this section, we focused on four potential predictors of classifier learning.

Classifier-based semantic categorization

Humans organize words that share similar semantic representations (Markman & Hutchinson, 1984). The words *hands* and *gloves* go together, as they are thematically related. The words *cat*, *dog* and *rabbit* can be grouped together, as they are all animals. A unique kind of word organization that may be formed by native Mandarin speakers is based on shared classifiers. Core semantic properties of a classifier could be extracted by increased exposure to the classifier and its corresponding nouns. For example, Mandarin-speaking children are exposed to the classifier-noun combination of "条 tiáo" and *snake*. When they receive increased exposure, they find that other nouns are paired with "条 tiáo", such as *fish*, *tie*, *trousers* and *rope*. Then, children may summarize and synthesize the shared semantic properties among these objects (i.e., long, narrow and flexible), and encode them in the classifier "条 tiáo". When children need to quantify objects that are long, narrow and flexible, "条 tiáo" is likely to be activated.

Some researchers believe that semantic categorization is a mechanism that supports the learning of classifiers. Uchida and Imai (1999) considered that the development of classifier-based semantic categorization explained the development of classifiers, although no evidence was provided to support their proposal. Erbaugh (2006) regarded "classifiers as a system of noun categorization" (p3), which could be analogous to German or Spanish gender morphemes.

Sera, Johnson, and Kuo (2013)'s tested Mandarin-speaking children and English-speaking children on the categorization of objects to understand the kinds of information children used to classify objects. A picture was firstly given to children (e.g., *snake*). Then the experimenter presented two pictures (e.g., *rope* and *rabbit*) and asked children to select the picture that was more similar to the first picture. *Rope* and *snake* were related as they shared the same classifier "条 tiáo". *Snake* and *rabbit* were related as they were both animals. The hypothesis was that if classifier learning and classifier-based semantic categorization were related, Mandarin-speaking children should be more likely to match *rope* with *snake* than English-speaking children. Results showed that Mandarin-speaking children chose *rope* in more than 70% of opportunities whereas English-speaking children chose *rope* and *rabbit* equally often. The patterns were the same in adult Mandarin speakers and English speakers. In addition, the researchers measured classifier comprehension among these Mandarin-speaking children. Probability of categorizing objects by classifiers was significantly higher in children who demonstrated better classifier comprehension. This was the only study providing some evidence about the correlation between classifier learning and classifier-based semantic categorization. The findings need to be replicated and validated.

Semantic relatedness

The second potential predictor we considered was degree of semantic relatedness. Some classifiers are similar with subtle distinguishing semantic properties. For example, "片 piàn" and "面 miàn" are semantically-similar classifiers that are paired with flat objects. A fine-grained difference between them is that "片 piàn" is related to flat thin objects (e.g., *bread*, *leaf*) and "面 miàn" emphasizes the flat surface (e.g., *wall*, *drum, mirror*). Other classifiers are more distant semantically. For example, "片 piàn"/ "面 miàn" are semantically distinct from "条 tiáo"/"支 zhī", as "条 tiáo" and "支 zhī" are paired with long objects.

Semantic distance could affect Mandarin-speaking children's classifier learning. In an error detection task, Mandarin-speaking children were asked to determine whether a puppet's usage of a classifier was correct (Uchida & Imai, 1999). Three nouns were paired with the same classifier (e.g., 条 tiáo): a target object (e.g., *worm*), a semantically-similar but non-target object (e.g., *mouse*), and a semantically-distant and non-target object (e.g., *car*). The results showed that children were more likely to identify the pairs of classifiers and the semantically-distant objects as wrong usages. They showed lower performance on the pairing of classifier and semantically-similar objects.

However, in Uchida and Imai (1999), the semantically-related objects were not well defined. For example, *mouse* was used as a semantically-similar but non-target object when "条 tiáo" (long, narrow and flexible) was tested. It may not be convincing that *mouse* is long and narrow. In the current study, we defined semantically-related objects as that they should be paired with a semantically-similar classifier. Additionally, we obtained adult agreement on classifier-noun combinations in our comprehension test (specified in the method section).

In production, Salehuddin and Winskel (2009) found that Malay-speaking children commonly substituted a target classifier using a classifier with shared but only one distinguishing semantic feature. For example, a predominant substitution for "keping" (2D, +rigid) was "helai" (2D, -rigid). In order to successfully learn classifiers, children need to make fine distinctions among semantically-similar classifiers. Since Malay does not have a general classifier to be used as a default, Malay-speaking children used specific classifiers in substitutions. Mandarin has the general classifier "个 gè", so in production Mandarin-speaking children should use it more predominantly than specific classifiers, as many previous studies have shown (e.g., Tse et al., 2007).

Frequency of classifier input

Mandarin classifiers have many exceptions to the general "rules" of semantic categorization. Native Mandarin speakers use "条 tiáo" to quantify *news* and *clues*, which are apparently not long, narrow and flexible in the physical sense. These exceptions do not align with the typical semantic properties of classifiers, which cast doubt on the proposal of semantic categorization. Maybe children do not take a rule-based approach to categorize semantic features but rely on rote memory to learn each individual classifier. Thus, some researchers have proposed that semantic categorization alone was insufficient to support classifier learning (Li et al., 2010). Children need to have adequate exposure or receive explicit instructions for classifiers from their language learning environment (Ying et al., 1983). Frequency of classifiers in children's input should be more strongly related to their success of classifier learning (Salehuddin & Winskel, 2009). Alternatively, classifier-based semantic categorization may interact with frequency of classifier input. When the input frequency of a classifier was held constant across different children, variations of classifier-based semantic categorization could relate to children's different performance on the classifier.

The usage-based account of language acquisition emphasizes item-based learning of linguistic forms and the key role of input frequency in idiosyncratic learning (Bybee, 1995, 2007; Tomasello, 2003; Ambridge & Lieven, 2011)[5]. Language acquisition starts from the learning of item-based linguistic constructions, such as concrete words, phrases and utterances. Bannard, Lieven and Tomasello (2009) used computational modeling to evaluate explanatory power of idiosyncratic words and phrases in the language production of two English-speaking children at two years old. The results showed that the lexically specific model provided good explanation, indicating that a child's knowledge of language starts from concrete speech.

As more linguistic materials are accumulated in exposure, language acquisition progresses to abstract constructions as a result of more mature cognitive abilities of pattern finding and analogy (Ibbotson, 2013). These abstract constructions range from word combinations (e.g., more X) to complex syntactic structures (e.g., X VERBed Y the Z.). However, it is important to note that concrete and idiosyncratic expressions are never completely abandoned by mature language learners, and these specific linguistic forms and the influence of frequency remain crucial to language usage (e.g., Arnon & Snider, 2010). The learning process thus could be a complex interplay of rule abstraction and item-by-item memorization.

Since used-based accounts focus on item-based learning in the process of language acquisition, input frequency is regarded to play a significant role (Ibbotson, 2013; Tomasello, 2003). Constructions that are consistently and frequently used are more likely to become automatized, yielding more successful learning. As a result, the type and amount of language input has been a particular focus by usage-based linguists. For example, Cameron-Faulkner, Lieven, and Tomasello, (2003) analyzed language production by 12 English-speaking mothers and their 2-year-old children from dyadic mother-child interactions. Mothers' input was examined in an item-based manner (e.g., frequency of specific copula constructions like

---

[5] Also see Yang (2016) for a mathematical equation to predict the dominant usage of rules or memorization in child language acquisition. To make language learning more efficiently, a threshold for irregular forms (e.g., held for hold) is calculated which determines whether learning is rule-based or item-based. Productive rules are applied if number of exceptions and irregularities is below a critical threshold. Otherwise, children tend to learn idiosyncratic constructions by memorization.

"There's NP", "That's NP", "It's NP"). The results showed that children's frequency of early usage of these frames was significantly correlated with input frequency by their mothers.

Applying the usage-based model to classifier learning, children may start from learning specific constructions with classifiers that are more frequent in language input and later progress to more abstract semantic categorization for different classifiers. Along the entire process, frequency of input remains significant in the success of classifier learning. Repeated exposure and use strengthen mental representations and increase the accessibility of classifiers. Higher-frequency classifiers are more automatic and easier to retrieve, and lower-frequency classifiers are less automatic and harder to access. As the learning process may involve both the item-based and rule-based learning strategies, input frequency may interact with semantic categorization by shared classifiers. This application in Mandarin classifier is an extension of the usage-based model, as the model is predominantly used to explain the acquisition of syntactic patterns instead of semantic patterns (e.g., Tomasello, 2003).

Salehuddin and Winskel (2009) measured production accuracy for eight classifiers in Malay-speaking children. The researchers estimated input frequency of these classifiers using a corpus of 150,000 words that were collected in a range of situations (e.g., children's television programs, storybooks). The results showed a significant positive relationship (r_s=.69, p<0.05) between production accuracy and frequency of classifiers in the corpus. Nevertheless, the authors claimed that quantifying input frequency for real classifiers is difficult. It is very likely that the corpus being measured is not fully representative of children's actual linguistic environment. In addition, it is hard to determine which settings are most beneficial for children's classifier learning.

A learning task controlling for frequency of input for invented classifiers can help address the issue of frequency, because experimenters can directly manipulate the number of demonstrations in children's learning process. To the best of our knowledge, there are two classifier learning studies (i.e., Uchida & Imai, 1999; Culbertson & Wilson, 2013). Although none of them manipulated frequency of classifier, we reviewed them to guide the design of a classifier learning task in the present study.

Uchida and Imai (1999) conducted a classifier learning task in Japanese-speaking four-year-old and five-year-old children. They taught two real Japanese classifiers[6]. The children

[6] Although they did not test invented classifiers, "a pretest was conducted on a pool of children to select those who had not yet learned the usage of *hiki* and *tou*" (p60). The two Japanese real classifiers thus were considered to be unknown to these

then were assigned to one of three learning conditions: rule-explicitly-given (i.e., rules of the two classifiers were explicitly taught), exemplars-only (i.e., only exemplars of the two classifiers were displayed) and control (i.e., no training was provided). Immediately after the training, a post-test was administered using a forced-choice selection task. Results[7] showed that the two age groups were comparable on the control and the rule-explicitly-given conditions: performance on the control condition was at floor, and performance on the rule-explicitly-given condition was at ceiling. The most interesting findings were from the exemplar-only condition: children at age four only achieved below chance accuracy, whereas accuracy in children at age five was over 70%. The-exemplar-only condition is highly similar to how classifier learning happens in naturalistic input, in which caregivers are unlikely to explicitly teach semantic rules for classifiers but specific exemplars. Uchida and Imai (1999) purported that the difference in accuracy between 4- and 5-year-old children was related to the improved ability to categorize classifier-based semantic properties. Additionally, we think that the 5-year-old children may have gained more exposure and more receptive knowledge of classifiers than the 4-year-old children. Therefore, even though the cognitive maturation account is likely to be true, the input frequency account cannot be ruled out.

The other learning study was conducted in English-speaking adults by Culbertson and Wilson (2013). The researchers invented two classifiers which were modeled on two Cantonese classifiers. One classifier was used with objects that were rigid, narrow and long (e.g., pen), and the other was paired with objects that were flexible, broad and flat (e.g., towel). The learning phase included 48 trials (24 for each classifier). In each learning trial, participants listened to an auditory stimulus and selected one written form from four while looking at an image demonstrating varied quantities of an object. Ten participants were presented with classifier-noun pairings that followed the above-mentioned semantic rules (i.e., semantic condition), and 10 participants were presented with random pairing that should not lead to rule learning (i.e., random condition). After a brief break, participants continued to a test phase by attending the same task. In addition to familiar objects that had been demonstrated in the

children. A detailed description about this pretest was not specified in the paper. It is likely that these children had not started using the two real classifiers in production, but they should have some receptive knowledge of the classifiers.

[7] As the paper is a review of an earlier study written in Japanese, many details were not disclosed. The researchers only reported descriptive data.

learning phase, they were tested on novel objects to test generalization of classifier knowledge[8]. The results showed that as expected, accuracy was significantly higher in the semantic condition than the random condition (86% vs. 45%). Participants showed comparable accuracy for familiar and novel objects, indicating that they were able to generalize the semantic rules to new contexts.

The two learning studies (Uchida & Imai, 1999; Culbertson & Wilson, 2013) guided our design of a learning task in Mandarin-speaking children. First, to avoid ceiling effect and to more closely mimic real-life classifier learning, we decided to present exemplars of invented classifiers without explicit teaching. Second, we determined that the frequency of input for invented classifiers should exceed the amount provided to adults. By consulting previous learning studies of novel English noun morphemes in preschool children, picture stimuli were used as exemplars to illustrate the invented morpheme, and frequency of input ranged from 20 to 40 (e.g., Kohnert & Danahy, 2007; Kaushanskaya, Gross, Sheena, & Roman, 2017). As a more complex semantic system, classifiers may be more challenging than English noun morphemes. We decided to use more than 40 times exposures in the current learning task.

Third, we included a novel object condition to examine children's generalization of classifier knowledge. For real classifiers, a novel object condition also helps control children's varied exposure with the range of nouns that can be paired with the same classifier (Li et al., 2010). For instance, one child hears "条 tiáo" and pairs it with *snake*, because his parents have told stories about snakes and used "条 tiáo" to quantify snakes. Another child does not have such exposure and cannot pair 条 tiáo" with *snake*. Both Li et al. (2010) and Fang (1985) included a novel object condition in addition to a familiar object condition, and consistent accuracy was found in comprehension (Li et al., 2010) and production (Fang, 1985). However, we noticed that in Li et al. (2010) adult agreement on three classifiers in the novel object condition did not reach 90%. In Fang (1985), there was no agreement obtained from adults. In this study, we intended to obtain adult agreement and set the minimum level at 90%.

Vocabulary

Vocabulary knowledge may predict classifier development. On one hand, many classifiers serve as nouns in Mandarin noun-noun compounds. For example, "面 miàn" is a

---

[8] No novel objects were presented to participants in the random condition, as there were no correct choices for novel trials under this condition.

word in multiple noun-noun compounds, and it means “the surface of something” (e.g., 湖面 lake surface; 桌面 table surface). The meaning of the noun “面 miàn” and the classifier “面 miàn” share the same semantic representations. Therefore, the expansion of vocabulary could benefit the learning of classifiers.

On the other hand, children need to know different semantic properties of words before they can summarize and form a category for them. In order to store and access words more efficiently, children with larger vocabulary may have stronger needs to categorize words. It is thus more likely for them to form semantic categories for different classifiers than children with smaller vocabulary. Individual variations of vocabulary knowledge (especially the breadth) thus need to be taken into consideration when measuring the contribution of semantic categorization.

Vocabulary knowledge may serve to be a reliable indicator of classifier development. Extant classifier studies exclusively used age to index classifier development, but there is an increasing trend to use other indices (e.g., mean length of utterances) to more reliably predict child language development (e.g., Baron, Bedore, Peña, Lovgren-Uribe, López, & Villagran, 2018). Here, we intend to measure vocabulary knowledge in order to more accurately predict children’s performance on classifiers.

Summary of potential predictors

These findings suggest that factors beyond age potentially predict children’s classifier knowledge. These predictors emphasize different approaches of language learning. Semantic categorization on the basis of classifiers and semantic relatedness highlight rule-based categorization and synthetization of semantic features. Frequency of input and vocabulary knowledge draw our attention to item-by-item learning of each individual classifier. Because none of these factors have been studied together, it is important to consider the relative contribution of each of these.

RESEARCH GOALS AND PREDICTIONS

In the present study, our goal was to conduct an in-depth investigation to understand how TD monolingual Mandarin-speaking children learn classifiers. We wanted to systematically explore potential predictors of children’s learning of classifiers. Two experiments were implemented. In experiment 1, we measured children’s comprehension and production of real Mandarin classifiers. Vocabulary knowledge and semantic categorization strategies were measured to examine their correlations with classifier comprehension and production. Semantic relatedness was manipulated in the comprehension task. Experiment 2 was a learning task, in which children were trained on two invented classifiers. Frequency of classifier input was directly manipulated using a between-subject design.

Based on previous literature, we predicted that in addition to age, classifier-based semantic categorization, semantic relatedness, frequency of classifier input and vocabulary knowledge would relate to classifier comprehension and production in Mandarin-speaking children. Here are specific predictions:

1) Classifier-based semantic categorization: We predicted that children with a stronger preference for noun categorization by shared classifiers would demonstrate better performance on classifiers (Kuo & Sera, 2009; Sera et al., 2013). There would be an interaction between categorization and age. Uchida and Imai (1999) found that children at age five showed increased classifier learning than children at age four, which was regarded to relate to increased semantic categorization pertaining to classifiers.

2) Semantic relatedness: It has been found that semantic relatedness affected classifier comprehension and production (Uchida & Imai, 1999; Salehuddin & Winskel, 2009). In comprehension, if children make errors, they would be more likely to select an object that is paired with a semantically-similar classifier than a semantically-distant classifier to the target classifier. In production, children would be more likely to substitute a target classifier using a semantically-similar classifier than a semantically-distant classifier. Alternatively, since Mandarin has a general classifier, a non-semantic strategy may be used by children. The general classifier "个 gè" may be predominantly used as a placeholder to replace specific classifiers, as the classifier construction requires a classifier.

3) Input frequency: Based on the usage-based approach (Bybee, 1995; 2007; Tomasello, 2003; Ibbotson, 2013), input frequency of individual classifiers would predict the success of the learning of the classifiers. We aimed to manipulate frequency for invented classifiers in experiment 2. We predicted that the more exposure a child had with a classifier, the higher accuracy he or she would exhibit on this classifier. Frequency may interact with classifier-based object categorization strategy. For example, when two children received the same amount of classifier input, the child who prefers to categorize objects by classifiers would outperform the child who does not.

4) Vocabulary: We were the first to include vocabulary as a predictor in classifier learning. We considered that many classifiers serve as nouns in noun-noun compounds, so the learning of these nouns would benefit classifier learning. On the other hand, vocabulary size would shape children's needs to semantic categorization. Also, vocabulary knowledge would serve as a more reliable index of classifier development than age.

5) Age: As previous studies consistently showed age-related changes in Mandarin-speaking children's classifier comprehension and production (e.g., Chien et al., 2003; Ying et al., 1983), we predicted the same in the current study. We included Mandarin-speaking children between four and six years of age. As age was not a major focus of the current study, the age range was intended not to be broad. Based on the literature, this age range could allow us to capture relatively mature comprehension and progressive production of classifiers.

6) Comprehension and production: Consistent with previous studies (e.g., Fang, 1985), We predicted that classifier comprehension would precede production.

Chapter 2: Method

PARTICIPATNTS

We recruited 73 Mandarin-speaking children from two preschools in Beijing, China. Thirty-five children were from one preschool, and 38 children were from the other preschool. The two preschools were identified in collaboration with the Child Cognition Lab by Dr. Peng Zhou at Tsinghua University. Ethical approval for this project was granted by Tsinghua University (IRB protocol number: 20170018). Since the experimenter is the first author from the University of Texas at Austin, she obtained ethical approval from IRB offices at University of Texas at Austin (IRB protocol number: 2018050131). The preschools informed parents and teachers about the project, and children between the ages of four and six years were encouraged to participate. If parents were interested, they volunteered to participate. Parents signed consent forms and completed a questionnaire. Teachers also signed a consent form and completed their questionnaire.

The 73 children participated in both experiment 1 and 2 (detailed information below). Their parents completed a parent questionnaire, in which demographic information (e.g., birthdate, gender, maternal education) and medical information relating to speech and language difficulties (i.e., hearing loss, autism, cerebral palsy, Down syndrome, Williams syndrome, other neurological or genetic disorders) were requested. None of the 73 children was reported to have a neurological or genetic disorder.

As we only intended to include TD children, parents and teachers evaluated children's oral language performance to inform us concerns of language impairment. To guide parents'/teachers' decisions of concern, we used the Inventory To Assess Language Knowledge (ITALK) (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018). The ITALK is a parent/teacher questionnaire including five questions in five different areas of language, including vocabulary, speech, sentence length, grammar and comprehension. The five questions were translated into Mandarin. To explain the questions, we replaced the English examples with Mandarin examples. For instance, for the question about grammar, we provided examples of Mandarin grammatical features (e.g., aspect markers, "bei" construction) and asked parents to evaluate the frequency of accurate usage of these features. These examples we selected were vulnerable areas, based on the literature of child language development in Mandarin-speaking children (e.g., Hao et al., 2018; Li & To, 2017). For each question, there was a 0-5 rating scale (0 represents the lowest performance and 5 represents the highest). The experimenter explained the five questions to parents/teachers and addressed any questions they had. After completing the ITALK, we asked parents and teachers if they had any concerns regarding the child's language development and to explain the nature of their concerns.

We planned to include monolingual children. To know about children's language experience, we used the Bilingual Input Output Survey (BIOS; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2018). In this survey, parents reported children's hour-by-hour language input and output on a typical week day and weekend day. The timetable was later calculated into percentages of input and output in Mandarin and any other languages (e.g., Cantonese, English). The survey helped exclude children who had more than 20% input in other languages than Mandarin.

On the basis of the above questionnaires in parents and teachers, nine children were removed for any of the following reasons: 1) There were concerns from either parents or teachers about children's oral language, as guided by the ITALK. Five children were removed, and the concerns included "having difficulties with language comprehension", "taking a long pause to organize a sentence before uttering the sentence", "having difficulties with long sentence organization", "repeating a sentence multiple times but could not express meaning clearly", "demonstrating very low vocabulary compared to peers". Two parents reported that their children were very shy in front of strangers. Since this was not a concern about language but personality, we included the two children and regarded them as TD. 2) Based on parents' report on the BOIS, two children were not regarded as monolingual Mandarin-speaking. One child's Cantonese input was more than 20%. The other child had recently spent more than three months in the US. 3) Two children could not complete all the tasks but were able to participate in the categorization task, the first task that only required children to do picture selection. Even though the experimenter made efforts to elicit productions, they did not respond in the second task, the classifier production task. Both of them stopped at the second task and did not proceed to the following tasks.

The final sample thus had 64 children (35 boys). The mean age was 5;0 (year;month) (SD: 8.6 months), and the age range was 4;1 to 6;5. Maternal education was ranked: 1 indicates middle school or below, 2 indicates high school, 3 indicates associate degree, 4 indicates bachelor's degree, and 5 indicates master's degree or above. The average rank of maternal education was 4.0 (SD: 0.6), and the range was 3-5.

EXPERIMENT 1

Materials

In experiment 1, we tested children's knowledge of six real classifiers using categorization, production and comprehension tasks. These real classifiers were selected to cover semantic features of different shapes. See Table 1. Five of the six shape classifiers were selected from Li et al. (2010), which were reported to be highly frequent. We added another classifier (i.e., 粒 lì) to create a third pair to constitute three pairs of semantically-related classifiers, as one of our aims was to study the influence of semantic relatedness. For each classifier, we draw two pictures depicting two corresponding real-life objects that are typically paired with the classifier.

Pairs	Classifiers	Semantic features	Picture Stimuli
Long and narrow	条 (tiáo)	long, narrow, flexible	necklace, scarf
	支 (zhī)	long, narrow, rigid	candle, pencil
Flat	面 (miàn)	flat, smooth surface	mirror, wall
	片 (piàn)	flat, thin	leaf, bread
Cube-like	块 (kuài)	lump, cube	soap, cake
	粒 (lì)	tiny round particles	rice, pill

Table 1. Real classifiers, corresponding semantic features, and typical objects that can be paired with the real classifiers in experiment 1.

We included a novel object condition to control for children's varied exposure to the range of nouns that could be paired with the same real classifier. The novel object condition also served to test children's generalization of the knowledge of real classifiers in novel contexts. To verify test materials, we conducted a pre-test in 17 native Mandarin-speaking adults living in Beijing. They participated in a picture-selection task which included 24 trials. Twelve trials included real-life objects (See Figure 4) and 12 trials included novel objects that were created based on the typical semantic features of the six real classifiers (See Figure 1). In each trial, adults were asked to select one picture from three that corresponded to the experimenter's instructions (e.g., *Micky Mouse says he wants one zhī something*). See the comprehension task below for more detailed descriptions of the procedure.

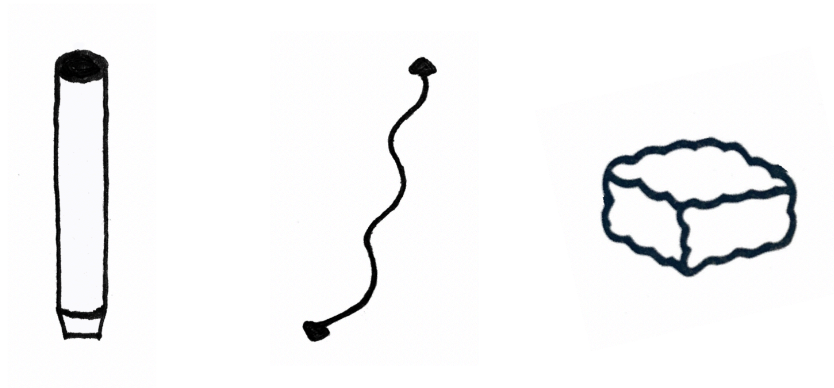


Figure 1. A sample test plate of the novel object condition. The novel objects were created on the basis of typical semantic properties of real classifiers. The left novel object was designed to illustrate the classifier "支 zhī", the middle novel object was designed to illustrate "条 tiáo", and the right novel object was created to illustrate "块 kuài". The target classifier was "支 zhī" and the target selection was the left novel object.

Based on adults' selections, we calculated adult agreement on each classifier (Table 2). Agreement was more than 90% for all classifiers in the familiar object condition, attesting to the familiarity of the stimuli. However, in the novel object condition, agreement on "片 piàn" was unacceptably low. As adult agreement for all novel objects was not reached, we excluded the novel condition from the following tasks in experiment 1, including the categorization task, the classifier production task and the classifier comprehension task.

Classifiers	Adult Agreement	
	Familiar Object Condition	Novel Object Condition
条 (tiáo)	100%	97%
支 (zhī)	100%	97%
面 (miàn)	97%	97%
片 (piàn)	97%	**53%**
块 (kuài)	97%	97%
粒 (lì)	100%	100%

Table 2. Agreement on familiar and novel object conditions by adults.

Categorization task

Following Kuo and Sera (2009) and Sera et al. (2013), we asked children to identify objects by similarity, in order to understand if children were more likely to judge objects as similar when they were expected to be paired with the same classifier. They were presented with a picture on top and then asked to select a picture from two below (Figure 2). The instruction was "下面哪一张和上边这张更像? *Which picture is more similar to the top picture?*". One choice matched the top picture by representing objects sharing the same classifier (e.g., pencil and bamboo). The other choice depicted an object that was thematically related (e.g., pencil and eraser). Here, we intended to present a contrast of semantic categorization by classifiers, which primarily encodes the perceptual features of objects, and by other types of non-perceptual semantic features.

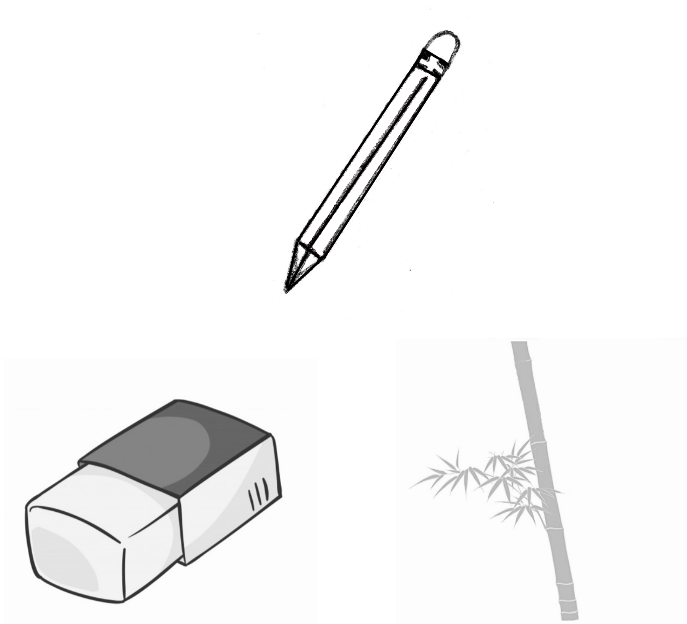


Figure 2. A sample trial of the categorization task.

All pictures depicted real objects. The top pictures were the same pictures for the six real classifiers as in the comprehension and production tests below. Each classifier was tested twice for a total of 12 trials. There were three practice trials to familiarize children with the task. The practice objects could not be paired with any of the targeted classifiers. To ensure that children's choices in this task were not primed by the experimenter's usage of classifiers, the administration of the categorization task preceded the production and comprehension tasks. We created two forms, in which the trials were randomly sequenced in two different orders.

Children's selections were recorded using a response sheet (Appendix 1). Later, the selections were coded on the basis of whether their responses could be linked by the use of the same classifiers. For example, if a child selected an object that shared the same classifier with the top object, the response was coded 1. If the child chose an object that shared other semantic relations with the top object, the response was coded 0.

Production task

The next task we implemented was the classifier production task. To elicit classifier production, children were asked to count the number of objects depicted in a picture. Because of the lack of agreement by adults as described before, the novel condition was deleted. The number of objects ranged from 1 to 4 in each trial. See Figure 3 for an example. We created two forms with different sequences of the test trials in two random orders.

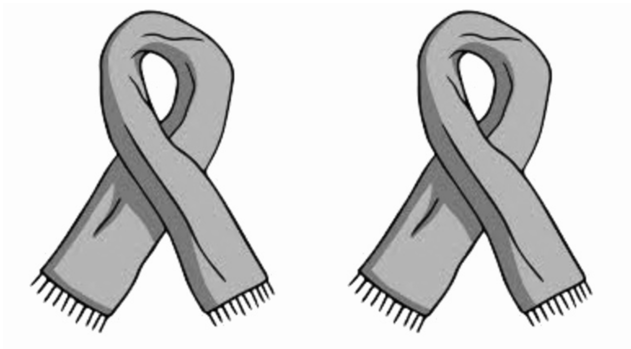


Figure 3. A sample trial of the production task in experiment 1.

Our production task included two practice trials and 12 trials with familiar object (two trials for each classifier). The experimenter provided three types of prompts to elicit responses in each trial: 1) Initially “有多少？ *How many?*”; 2) If the child only produced a number (e.g., 二 *two*), the experimenter prompted the child by asking “两什么？ *two what?*”; 3) If the child only produced the noun (e.g., 围巾 *scarf*), the experimenter prompted the child by asking “多少围巾？ *How many scarfs?*”. Note that all prompts were designed so that no classifiers would be modeled. The use of classifiers was optional in all modeled prompts. We did not provide further prompts if the child omitted classifiers once they had identified the target object (e.g., 两围巾 *two scarfs*), as classifier omission is potentially a type of error (Stokes & So, 1997).

Each child’s responses were recorded using an open response sheet (See Appendix 2). The experimenter wrote down the exact responses children provided. All productions were audio recorded for reliability. A second researcher listened to 20% of the recording and transcribed children’s responses. An agreement of 100% was reached between the two researchers.

We scored children’s responses using a binary scoring system. Target classifiers were coded 1. If a child used a non-target specific classifier but it is an acceptable alternative in Mandarin, the response was coded 1. These acceptable alternatives of specific classifiers included “三**根**蜡烛 three gēn candle” for “三支蜡烛 three zhī candle” (3 occurrences), “四**根**铅笔 four gēn pencil” for “四支铅笔 four zhī pencil” (5 occurrences), “一**颗**药 one kē pill” for “一粒药 one lì pill” (2 occurrences) and “三**块**镜子 three kuài mirror” for “三面镜子 three miàn mirror" (2 occurrences). Altogether, there were 12 occurrences of acceptable alternative classifiers. Non-target classifiers were coded 0, including the general classifier "个 gè", classifier omission and specific classifiers that were not targeted and not acceptable alternatives.

Comprehension task

A forced-choice picture selection task was administered afterwards to test children's comprehension of the six real classifiers. Semantic relatedness was manipulated in this task. The six classifiers included three pairs of semantically-similar classifiers (Table 1). For example, in the pair of "条 tiáo" and "支 zhī", "条 tiáo" is typically used with long, narrow and flexible objects, but "支 zhī" is typically used with long, narrow but rigid objects. In the picture selection task, an object that is used with "条 tiáo" could be a semantically-similar but non-target distractor when "支 zhī" was tested.

On each test plate, three possible selections were provided on the basis of the semantic-relatedness: 1) an object that should be paired with the target classifier (e.g., *tiáo-scarf*: long, narrow and flexible objects); 2) an object that should be paired with a semantically-similar classifier (e.g., *candle-zhī*: long, narrow and rigid objects); 3) a random object that should be paired with a classifier in the other two pairs of classifiers with relatively distinct semantic features (e.g., *bread-piàn*: flat slices). Figure 4 represents a sample test plate for "条 tiáo". The left (*scarf*) is the target. The middle (*bread*) is the distractor that should be paired with the semantically-distant classifier "片 piàn". The right (*candle*) should be paired with "支 zhī", which is a semantically-similar classifier to "条 tiáo". *Scarf* and *candle* are both long and narrow, but *scarf* is flexible, and *candle* is rigid. The position of the target object was controlled: 1/3 of the target objects appeared on the left, 1/3 appeared in the middle, and 1/3 appeared on the right. The other two objects were randomly positioned.

The procedure of the test was similar to Chien et al. (2003) and Li et al. (2010). Children were told that Mickey Mouse, a hand puppet from the US, wanted to play a guessing game with them. However, Mickey Mouse could not speak much Mandarin but could speak good English. If Mickey Mouse did not know the name of an object, he would say he wanted "something". In each trial, the child was asked to select the object that corresponded to the target classifier in a carrier phrase "米老鼠说他想要一 CLASSIFIER something" (*Mickey Mouse says that he wants one CLASSIFIER something*)". There were two forms, in which the test trials were randomly sequenced into two different orders.

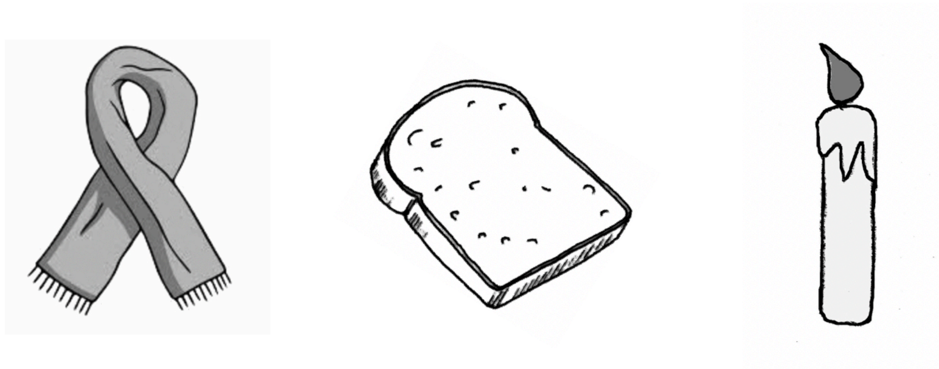


Figure 4. A sample test plate of the comprehension task in experiment 1. The test sentence was "米老鼠说他想要一条something" (*Mickey Mouse says that he wants one tiáo something*). The target object was positioned on the left (*scarf*).

There were 12 familiar object trials, and each real classifier was tested twice. Two practice trials were included in the beginning to familiarize children with the task, in which no target classifiers were included. All children selected the correct pictures in the two practice trials. A novel object condition was not included because of the lack of adult agreement. Each child's selections were recorded using a response sheet (See Appendix 3). Later, correct selections were coded 1, and incorrect selections were coded 0. Incorrect selections were recorded using the sheet and analyzed.

Vocabulary test

The final task we administered in experiment 1 was a vocabulary test. To understand how vocabulary knowledge was related to classifier knowledge, we tested children's vocabulary using the Receptive and Expressive Vocabulary Screener of Mandarin (RAEVSOM) (Sheng, Lam, Wang, Chow, & Zheng, in progress). The test included 16 trials of picture selection for receptive vocabulary and 16 trials of picture naming for expressive vocabulary. All words included in the two tests were nouns. The receptive vocabulary test has been administered to 580 Mandarin-speaking children, and the expressive vocabulary test has been given to 569 Mandarin-speaking children. Internal consistency was satisfactory for both the receptive vocabulary test (Cronbach's alpha=0.789) and the expressive vocabulary test (Cronbach's alpha=0.785). These values exceeded generally accepted threshold for good reliability (Henson, 2001), indicating that the trials on these tasks were consistently assessing the same skills. Among these children, 131 completed a sentence repetition task, and the correlation (i.e., external validity) with sentence repetition was significant ($r=0.492$, $p<0.01$).

Each receptive vocabulary trial had 4 choices, and only one choice corresponded to the target word the experimenter verbalized. The experimenter used a response sheet to record children's selections. Correct selections were coded 1, and incorrect selections were coded 0. The total score was the averaged accuracy of the 16 trials.

For expressive vocabulary, children were required to name objects depicted in pictures. Each trial had a list of acceptable answers, which were based on responses from 10 native Mandarin-speaking adults. The experimenter wrote down the exact responses using an open response sheet. The production was audio recorded, and a second researcher listened and transcribed 20% of children's responses. An agreement of 99% was reached between the two researchers. One response was unintelligible, and the two researchers had different interpretations. The two researchers agreed that the response was not a target response listed in the acceptable answers. Acceptable responses were coded 1, and unacceptable responses were coded 0. The total score was the averaged accuracy of the 16 trials.

EXPERIMENT 2

In experiment 2, we implemented a learning task, in which we manipulated input frequency of two invented classifiers. To measure influences of semantic categorization strategies and vocabulary knowledge, we included the data from the categorization task and the vocabulary task in experiment 1 as potential predictors of novel classifier learning. This could inform us about how children's current knowledge of real classifiers relates to their learning of invented classifiers.

Materials

We invented two classifiers. One classifier displayed fluffy and curly-haired imagined objects. The other classifier was paired with broken imagined objects. These semantic features were not encoded in any Mandarin classifiers and morphemes. We created a total of 24 novel objects, and each classifier was illustrated by 12 novel objects. To control for animacy, half of these novel objects for each classifier were animate, and the other half were inanimate.

We tried to minimize the processing load for children with regard to the phonological forms of the two novel classifiers. The criteria were: 1) the phonological form of each classifier consisted of one syllable with a CV (consonant vowel) structure. The CV structure is a very simple structure that is widely used in Mandarin and is common in many classifiers; 2) the selected consonants, vowels and tones have been found to be acquired by 3-year-old Mandarin-speaking children across different studies (see Li and To (2017) for a review); 3) phonological forms of the two invented classifiers did not correspond to any Chinese characters (Modern Chinese Dictionary, 2013). See Table 3 for semantic features, phonological forms and sample pictures of the two classifiers.

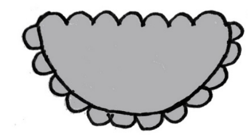
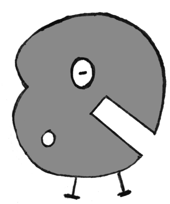
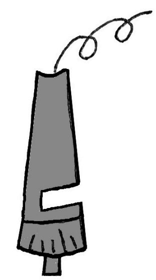
Classifiers	Semantic features	Sample pictures	
		Animate	Inanimate
dě	Fluffy and curly-haired		
tá	Broken		

Table 3. Phonological forms, semantic features and sample pictures of the two invented classifiers.

Procedure

There were two phases in this learning task: a teaching phase and a testing phase. In the teaching phase, the experimenter engaged in interactive play with the child to count Mickey Mouse's collections of novel objects with novel features. The two invented classifiers were embedded in the counting. To make it as a counting activity, the number of demonstrations for each picture stimulus varied (ranging from 6 to 8).

We manipulated frequency of input using a between-subject design. The lower frequency group included 30 children, and they received 42 demonstrations for each classifier. The remaining 34 children received 84 demonstrations, and the frequency of demonstrations in this higher frequency group doubled that in the lower frequency group. The children were randomly assigned into the lower and higher frequency groups, and at the time of assignment, the experimenter was blind to children's demographic information.

The teaching combined modeling and imitation. Previous studies showed that children learned new morphemes more effectively by making contrast (e.g., Swisher & Snow, 1994). Therefore, in each teaching trial, we followed a sequence of repeated modeling, imitation and contrastive modeling. See Figure 5 and the script below it for how we demonstrated the two invented classifiers.

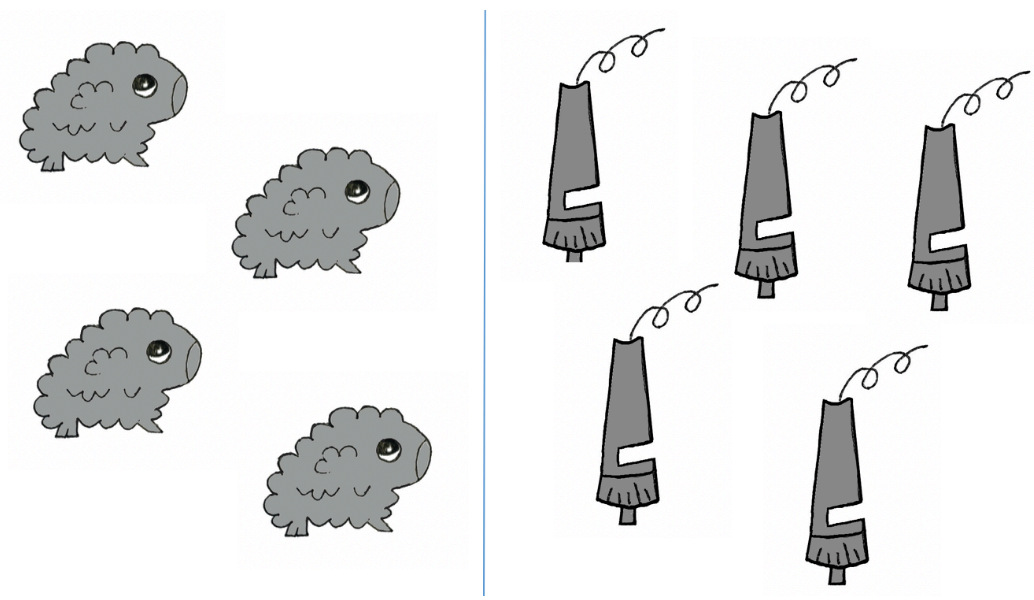


Figure 5. A sample teaching trial in experiment 2. The frequency of demonstration for the left invented classifier dě was 6, including 4-time repeated modeling, 1-time imitation, and 1-time contrastive modeling. The frequency of exposure for the right invented classifier tá was 7, including 5-time repeated modeling, 1-time imitation, and 1-time contrastive modeling.

Script: "看左边，这里有一 dě something，两 dě something， 三 dě something，四 dě something。米老鼠一共有___________（四 dě something）。 看右边，这里有一 tá something，两 tá something， 三 tá something，四 tá something，五 tá something 。米老鼠一共有___________（五 tá something）。所以，左边一共有四 dě something，右边一共有五 tá something。 "

Look at the left, here is one dě something, two dě something, three dě something, four dě something. Mickey Mouse altogether has___________(four dě something). Look at the right, this is one tá something, two tá something, three tá something, four tá something, five tá something. Mickey Mouse altogether has___________(five tá something). So, on the left, there are four dě something, and on the right, there are five tá something.

Immediately after the teaching phase, children continued to a testing phase. We included a production test and a comprehension test. Similar to experiment 1, the production test preceded the comprehension test to avoid providing additional models.

Familiarity was manipulated to measure children's generalization of classifier knowledge. There were familiar and novel conditions. We referred to the novel objects that children were exposed to in the teaching phase as familiar objects, as they saw these objects in the learning trials. There were 12 trials including familiar objects. Twelve trials included novel objects that had not previously been demonstrated in the teaching phase. Altogether, there were 24 trials for the examination of comprehension and production respectively.

Production task

Production was tested using a counting task, which followed the exact procedure of the production task in experiment 1. Similar to experiment 1, the production test preceded the comprehension test. The number of objects ranged from 1 to 4 (Figure 6). The experimenter guided children to name all objects by saying "something". Some children preferred to give a name to the novel objects, and we did not force them to use "something". All prompts were the same as in experiment 1. The experimenter documented production responses using an open response sheet (Appendix 4). We audio recorded children's responses, and a second researcher listened to 20% of the responses and transcribed these responses. Agreement of 100% was met between the two researchers.

If the responses were the targeted invented classifiers, we scored them as 1. If the responses were not the targeted invented classifiers, we scored them as 0. As the target classifiers were invented classifiers, no acceptable alternatives of non-targeted real classifiers were permitted. This is different from what we did in experiment 1, in which alternative non-target real classifiers were accepted (e.g., "三根蜡烛 three gēn candles" for "三支蜡烛 three zhī candles").

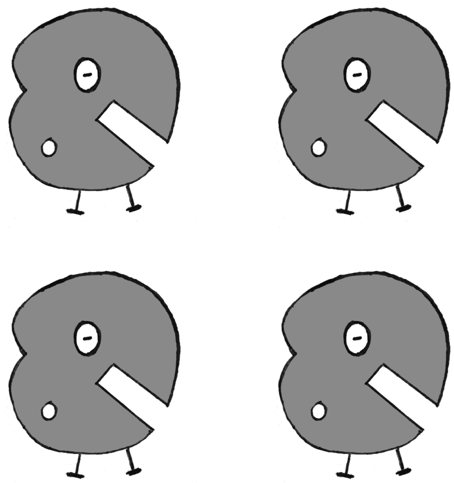


Figure 6. A sample trial to examine the production of invented classifiers in experiment 2.

Comprehension task

Comprehension was examined using a forced-choice picture selection task. Two selections were provided. One selection was the target, and the other was a random object associated with the other invented classifier. We controlled for animacy. The two selections were either both animate or inanimate. In addition, the two selections were either both familiar (novel objects that had been demonstrated in the teaching phase) or novel (novel objects that had not been demonstrated in the teaching phase). See Figure 7 for an example. The position of the target objects was controlled: half of the target pictures were on the left, and the other half target pictures were on the right. Children's selections were recorded using a comprehension response sheet (See Appendix 5). Correct responses were coded as 1, and incorrect responses were coded as 0.

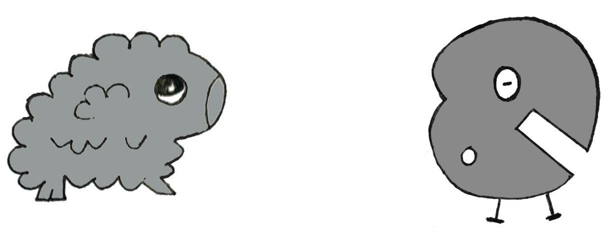


Figure 7. A sample trial to examine the comprehension of invented classifiers in experiment 2.

Chapter 3: Results

EXPERIMENT 1

As previous studies showed age-related changes for real classifiers (e.g., Chien et al., 2003; Li et al., 2010), here we display the demographic and descriptive data by age group. Children under age five were categorized into the younger group, and children above five were in the older group. This cutoff was selected to match previous literature by Uchida and Imai (1999), in which the researchers compared classifier learning in Japanese-speaking children at 4- and 5-year-old. See Table 4 for more detailed demographic information about the two age groups. There was no difference in maternal education between these two groups of children (t(62)=.754, *p*=.454).

Measures		Age Groups	
		Younger Group	Older Group
n		34	30
Gender (M:F)		17:17	18:12
Age (year;month)	Mean (SD)	4;4 (2.8)	5;8 (5.2)
	Range	4;1-4;11	5;1-6;5
Maternal Education	Mean (SD)	3.9 (0.7)	4.0 (0.6)
	Range	3-5	3-5
ITALK Rating	Mean (SD)	4.5 (0.29)	4.7 (0.21)
	Range	3.7-5	4.1-4.9

Table 4. Demographic information of participants by age. The younger group included children under five years old and the older group included children above age five. SD of age is displayed in month. Maternal education was ranked: 1=middle school or below; 2=high school; 3=associate degree; 4=Bachelor's degree; 5= Master's degree or above. The ITALK score is the average of parent's and teacher's rating.

We calculated averaged accuracy, standard deviations, and ranges of the measures in the two age groups, including categorization preference to objects that shared the same classifiers, production, comprehension, receptive vocabulary and expressive vocabulary. See Table 5. Preference to object categorization by shared classifiers was about 70% in both age groups. Accuracy of classifier production was extremely low in the younger group but higher in the older group. In the classifier comprehension task, both groups achieved over 70% accuracy, which was much higher than production accuracy. For receptive vocabulary and expressive vocabulary, younger children achieved above 60% accuracy, and accuracy of older children increased to more than 75%.

Measures		Age Group	
		Younger Group	Older Group
Categorization	Preference (SD)	0.70 (0.15)	0.74 (0.16)
	Range	0.33-0.92	0.25-1
Production	Accuracy (SD)	0.05 (0.07)	0.21 (0.20)
	Range	0-0.33	0-0.58
Comprehension	Accuracy (SD)	0.75 (0.16)	0.88 (0.10)
	Range	0.33-1	0.67-1
Receptive Vocabulary	Accuracy (SD)	0.64 (0.14)	0.76 (0.16)
	Range	0.25-0.94	0.38-0.94
Expressive Vocabulary	Accuracy (SD)	0.63 (0.16)	0.77 (0.12)
	Range	0.31-0.94	0.5-0.94

Table 5. Averaged accuracy, standard deviations, and ranges of dependent variables in the younger and older age group. For categorization, we calculated percentages of object choice that shared same classifiers with the target objects.

Comprehension

We employed a generalized linear mixed model for binary data to analyze results for comprehension. Children's accuracy on the forced-choice selection task was the dependent variable. It was entered as binary data: 0 represents incorrect choice and 1 represents correct choice. We then entered the categorization preference that corresponded to each comprehension trial (0 indicates a choice that did not share a same classifier with the target object and 1 indicates a choice that shared the same classifier with the target object), classifier (from 1 to 6 - six real classifiers), age group (1 indicates the younger group and 2 indicates the older group), overall vocabulary accuracy (averaged accuracy on receptive and expressive vocabulary tests), maternal education (from 1 to 5 - middle school or below to Masters' or above) and form (1 indicates form A and 2 indicates form B) into the model[9].

We investigated main effects of all independent variables. As previous studies suggested that categorization may interact with age (Uchida & Imai, 1999), we included the two-way interaction of categorization*age. We hypothesized on the potential interaction between categorization and frequency of input. However, as frequency was not manipulated in the examination of real classifiers, the interaction was not included in the model.

The results showed two significant main effects. Vocabulary significantly predicted children's classifier comprehension, $F(1, 755)=6.45$, $p=.01$, odds ratio=0.075. Children with larger vocabulary tended to score higher in the classifier comprehension task than children with smaller vocabulary. Per 1% increase in vocabulary accuracy, there was 7.5% increase in the accuracy of real classifier comprehension. The other significant predictor was classifier ($F(5, 755)=12.34$, $p<.001$). We then conducted multiple pairwise comparisons with LSD corrections. Accuracy on "条 tiáo" was lower than any other classifiers, including "支 zhī" ($t(755)=4.75$, $p<.001$), "面 miàn" ($t(755)=4.05$, $p<0.001$), "片 piàn" ($t(755)=5.30$, $p<.001$), "块 kuaì" ($t(755)=6.19$, $p<.001$) and "粒 lì" ($t(755)=6.53$, $p<.001$). Children's comprehension of "粒 lì" was better than "支 zhī" ($t(755)=2.88$, $p=.02$) and "面 miàn" ($t(755)=2.88$, $p=.002$). Accuracy on "块 kuaì" was higher than "面 miàn" ($t(755)=2.55$, $p=.009$). All the other comparisons did not reach significance. Odds ratios were reported in a pairwise manner. "块 kuaì" was compared with all the other classifiers. Children were 1.67 times more likely to be accurate in "块 kuaì" than "片 piàn", 0.78 times more likely to be accurate in "粒 lì" than "块 kuaì", 2.78 times more likely to be accurate in "块 kuaì" than "面 miàn", 9.07 times more likely to be accurate in "块 kuaì" than "条 tiáo", 2.22 times more likely to be accurate in "块 kuaì" than "支 zhī". See Figure 8.

[9] Among all independent variables, age group and overall vocabulary accuracy were correlated ($r=.485$, $p<.01$). Vocabulary knowledge was one of the major predictors. Even though age was not the main focus of the study, based on Uchida and Imai (1999), the interaction between age and semantic categorization strategy was our major interest. As required by the model, all the variables in an interaction should be included as independent variables. Therefore, both vocabulary and age were kept in the model. Note that the model estimates the association between a given independent variable and the outcome holding all other independent variables constant. While the main effect of vocabulary knowledge was estimated, age was controlled.

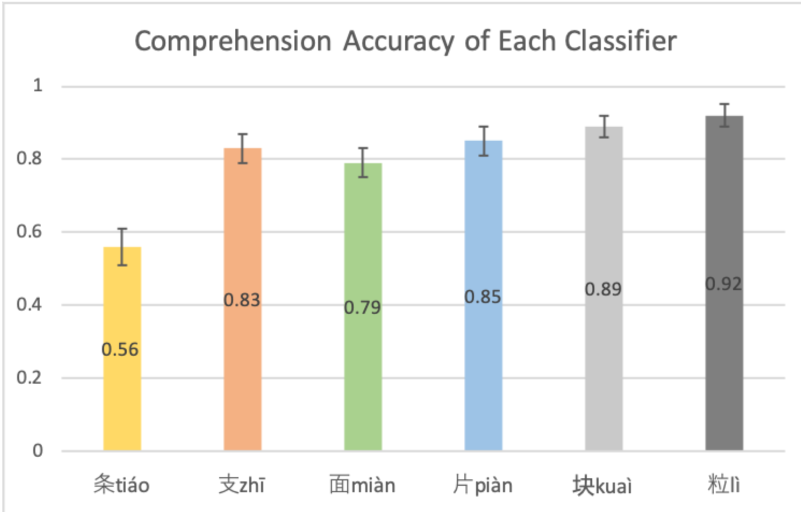


Figure 8. Comprehension accuracy of individual classifiers in experiment 1.

As each classifier was paired with two nouns, children may display different performance on the two trials of the same classifier. We thus examined accuracy of each individual trial (Figure 9). Children demonstrated an accuracy gap of 0.16 between the classifier-noun combinations of *tiáo_necklace* and *tiáo_scarf*. For other classifiers, there were relatively small differences between the two trials. Overall, comprehension accuracy was consistent across the two test trials for each classifier.

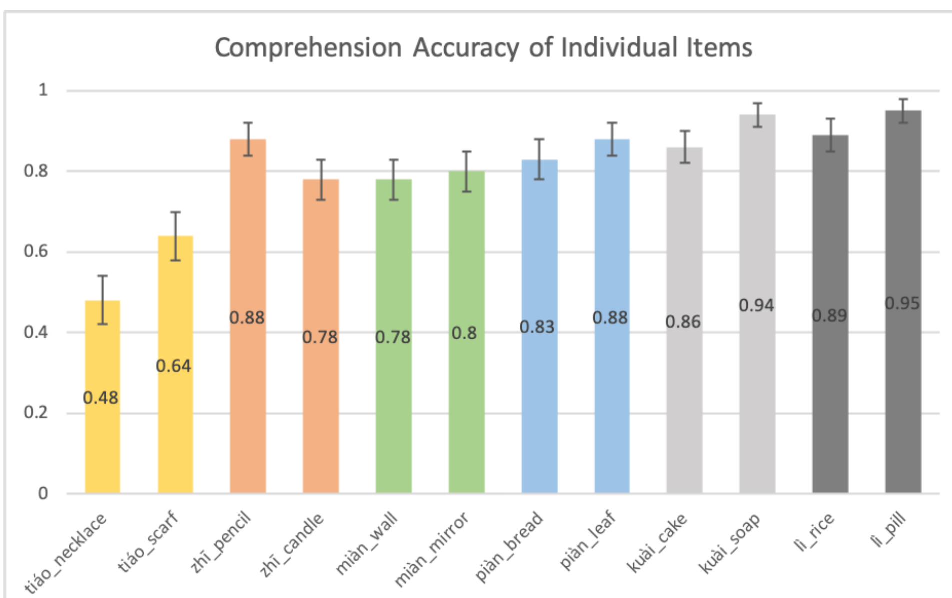


Figure 9. Comprehension accuracy of individual classifier-noun combinations in experiment 1.

There were no significant main effects of categorization (F(1, 755)=2.35, *p*=.13, odds ratio=1.93), age group (F(1, 755)=2.59, *p*=.11, odds ratio=2.14), maternal education (F(1, 755)=0.63, *p*=.53[10]) and form (F(1, 755)=0.44, *p*=.51, odds ratio=1.20). The interaction between age and categorization was not significant, (F(1, 755)=1.24, *p*=.27, odds ratio=0.59).

We examined children's errors in comprehension to see if semantically-related selections were the dominant errors compared to semantically-unrelated selections. On average, 81.2% of errors children made were semantically-related errors, and the rest 18.8% were semantically-unrelated errors. As we predicted, the dominant error type was semantically-related choices in the comprehension test.

Production

Results of the production test were analyzed using a generalized linear mixed model for binary data. The dependent variable was accuracy on the production test (0 indicates an incorrect response and 1 indicates a correct response). We entered the same independent variables as the previous mixed model, including categorization preference, classifier, age group, overall vocabulary accuracy, maternal education and form. Similarly, we focused on the main effects of all the independent variables and the two-way interaction between categorization and age.

Categorization was not a significant predictor for the accurate production of real classifiers, F(1, 755)=1.52, *p*=.22, odds ratio=1.36. The main effect of form was not significant, F(1, 755)=.80, *p*=.37, odds ratio=1.38. The interaction between age and categorization was also not significant (F(1, 755)=.16, *p*=.69, odds ratio=1.33).

We found three significant main effects and a marginally significant main effect. Age was a significant predictor for children's production accuracy, F(1, 755)=6.79, *p*=.009, odds ratio=2.88. The older group scored higher on the production task than the younger group, and the older children were 2.88 times more likely to produce accurate classifiers than the younger children. Similar to results in comprehension, there was a main effect of vocabulary (F(1, 755)=7.90, *p*=.005, odds ratio=.014). Children who scored higher in vocabulary achieved better performance on the classifier production task than children who scored lower in vocabulary. For every 1% increase in vocabulary accuracy, there was a 1.4% increase in classifier

[10] Odds ratios were reported in a pairwise manner, and the reference level of maternal education was 5. The odds ratio for 3 and 5 was 0.62, and the odds ratio for 4 and 5 was 0.87.

production accuracy. A marginally significant main effect of maternal education was found (F(2, 755)=2.90, *p*=.056). Marginally, children whose mothers earned a Master's degree or above achieved higher classifier production accuracy than children whose mothers earned a bachelor's degree (t(755)=1.89, *p*=.06). Odds ratios were reported in a pairwise manner. The reference level of maternal education was Master's degree and above. Children whose mothers received Master's degree and above were 2.35 more likely to produce accurate classifiers than children whose mothers received associate degree, and 2.62 more likely to produce accurate classifiers than children whose mothers received bachelor's degree. Note that all mothers in the current sample had at least a Bachelor's degree. All the other pairwise comparisons were not significant.

There was a main effect of classifier, F(5, 755)=6.58, *p*<.001. Multiple pairwise comparisons were conducted with LSD corrections. Children more accurately produced "片 piàn" than "条 tiáo" (t(755)=3.63, *p*<.001), "支 zhī" (t(755)=3.02, *p*=.003), "面 miàn" (t(755)=3.04, *p*=.002) and "粒 lì" (t(755)=3.87, *p*<.001). Accuracy of "块 kuài" were higher than "粒 lì" (t(755)=2.84, *p*=.005) and "条 tiáo" (t(755)=2.43, *p*=.02). All the other pairwise comparisons did not reach significance. The reference classifier to report odds ratios was "块 kuài". Children were 0.53 times more likely to accurately produce "片 piàn" than "块 kuài", 5.31 times more likely to accurately produce "块 kuài" than "粒 lì", 1.96 times more likely to accurately produce "块 kuài" than "面 miàn", 3.42 times more likely to accurately produce "块 kuài" than "条 tiáo", 1.91 times more likely to accurately produce "块 kuài" than "支 zhī". See Figure 10.

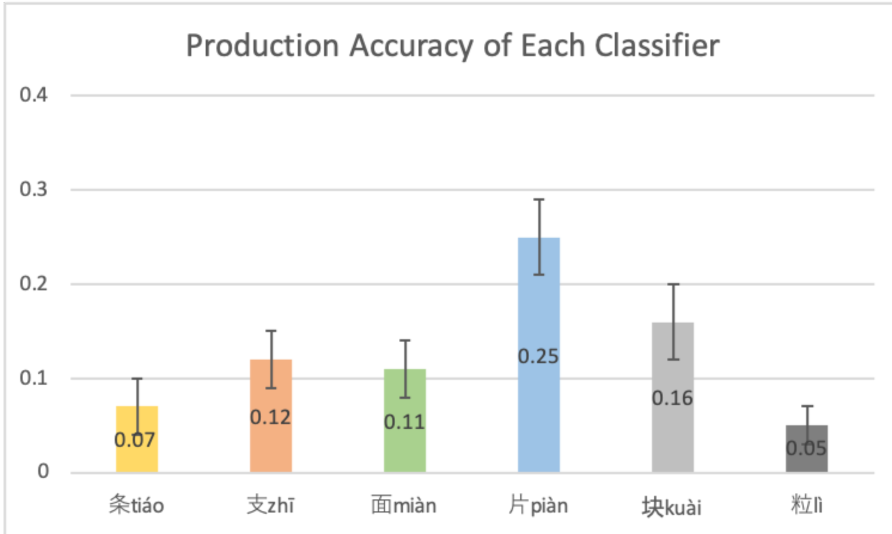


Figure 10. Production accuracy of individual classifiers in experiment 1. Children's averaged accuracy for individual classifiers was below 0.3.

For the same reason, we explored production accuracy of individual test item. As can be seen from Figure 11, there was a big discrepancy (0.34) between the two test items for "片 piàn". Children were more likely to produce "片 piàn" when it was combined with *leaf* than *bread*. Regarding the classifier "支 zhī", children tended to pair it with *pencil* more often than *candle* (discrepancy - 0.14). Similarly, for "面 miàn", children were more likely to produce "面 miàn" when it was combined with *wall* (discrepancy - 0.12) than *mirror*. Differences were relatively small for the other three classifiers.

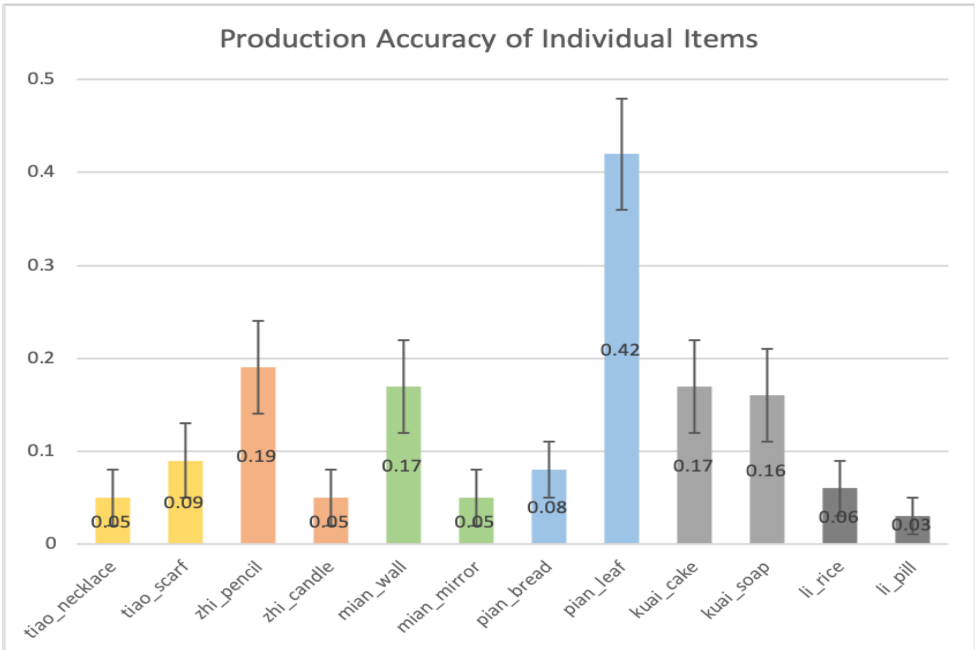


Figure 11. Production accuracy of individual classifier-noun combination in experiment 1. Children's averaged accuracy for individual items was below 0.5.

We analyzed the composition of children's responses for individual items in production as well. As shown in Figure 12, the general classifier "个 gè" was produced most often. Except for *pian_leaf*, the targeted specific classifiers were rarely used. Two researchers coded children's errors to determine if substitutions were semantically-related or semantically-unrelated. The agreement between the two researchers reached 100%. Non-targeted specific classifiers that were either semantically-unrelated or semantically-related were both rarely used.

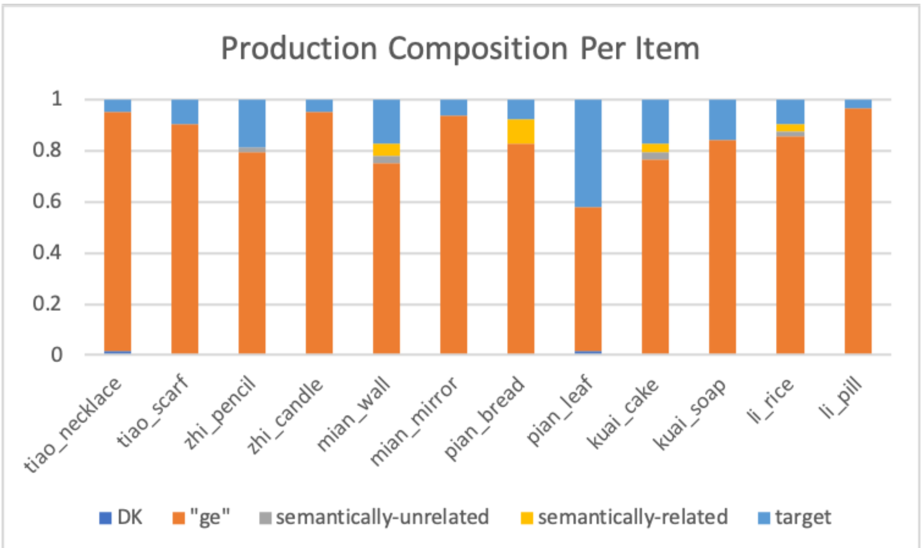


Figure 12. Distribution of production for each classifier-noun combination in experiment 1. "DK"="don't know".

We then further explored specific substitutions using semantically-related and semantically-unrelated classifiers (Table 6). Percentages of semantically-unrelated and semantically-related substitutions were both very low. We focused on semantically-related substitutions. "四块面包" (four kuài bread) had the highest frequency (7 times). In our stimuli, we demonstrated slices of bread, and the target classifier should be "片 piàn". These children may more regularly see cube-like bread than slices of bread in real life. "墙" (wall) was paired with "幅 fú" (for painting), "片 piàn" (for bread), "张 zhāng" (for paper), whereas the target classifier "面 miàn" emphasizes surfaces (for wall and mirror). All these inaccurate specific classifiers share similar semantic properties "flat and thin", but they represent fine-grained differences with the targeted specific classifier "面 miàn". Similarly, the other three semantically-related substitutions are associated with fine-grained distinctions of semantic features. "颗 kē" and "块 kuài" are related to larger cube-like or round objects, but rice should be paired with "粒 lì" that denotes tiny round objects. Between "颗 kē" and "块 kuài", "颗 kē" is typically round and "块 kuài" does not need to be round, and cake should be paired with "块 kuài" instead of "颗 kē".

Again, the majority of children's production was using the general classifier "个 gè" to replace the target classifier. Among the 64 children, 21 children used "个 gè" exclusively, and they did not use any specific classifiers. There were only two occurrences of classifier omission, and they were produced by the same child.

Type	Occurrence	Percentage	Errors
Omission	2	0.3%	一树叶 one leaf (片 piàn) -1 一项链 one necklace (条 tiáo) -1
"个 gè" substitution	647	84.2%	
Semantically-unrelated substitution	5	0.7%	四块铅笔 four kuài pencil (支 zhī) -1 两片米 two piàn rice (粒 lì) -1 四片蛋糕 four piàn cake (块 kuài) -1 两座墙 two zuò wall (面 miàn) -1 两部砖头 two bù brick (块 kuài) -1
Semantically-related substitution	14	1.8%	四块面包 four kuài bread (片 piàn) -7 两幅墙 two fú wall (面 miàn) -2 两片墙 two piàn wall (面 miàn) -1 两张墙 two zhāng wall (面 miàn) -1 两颗米粒 two kē rice (粒 lì) -1 两块米粒 two kuài rice (粒 lì) -1 四颗蛋糕 four kē cake (块 kuài) -1
Target or other acceptable classifier	100	13.0%	

Table 6. Occurrences and percentages of production types. For each specific error, the accurate and targeted classifier is displayed in parentheses. The number following each error indicates the frequency of occurrence.

Comprehension and production

We explored the correlation between comprehension accuracy and production accuracy for real classifiers in experiment 1. We intended to submit the data for Pearson's correlation tests, but the data did not meet the assumption of normality. We then ranked the data and submitted it for a Spearman's rho correlation test, which is non-parametric and does not assume normal distribution. There was a significant correlation between comprehension and production (r_s=.353, *p*=.004) in experiment 1.

Overall, production appears to be more challenging than comprehension. While the mean accuracy of comprehension was 82%, production accuracy was only 13%. To explore different comprehension-production patterns, we set cutoffs as medians of comprehension accuracy (i.e., 0.67) and production accuracy (i.e., 0.29) as shown on Figure 13. There were six children who fell on the cutoff of comprehension, and they were categorized as demonstrating good comprehension. There were three comprehension-production patterns: 1) high comprehension and relatively high production; 2) high comprehension and low production; 3) relatively low comprehension and low production. No children demonstrated low comprehension but high production.

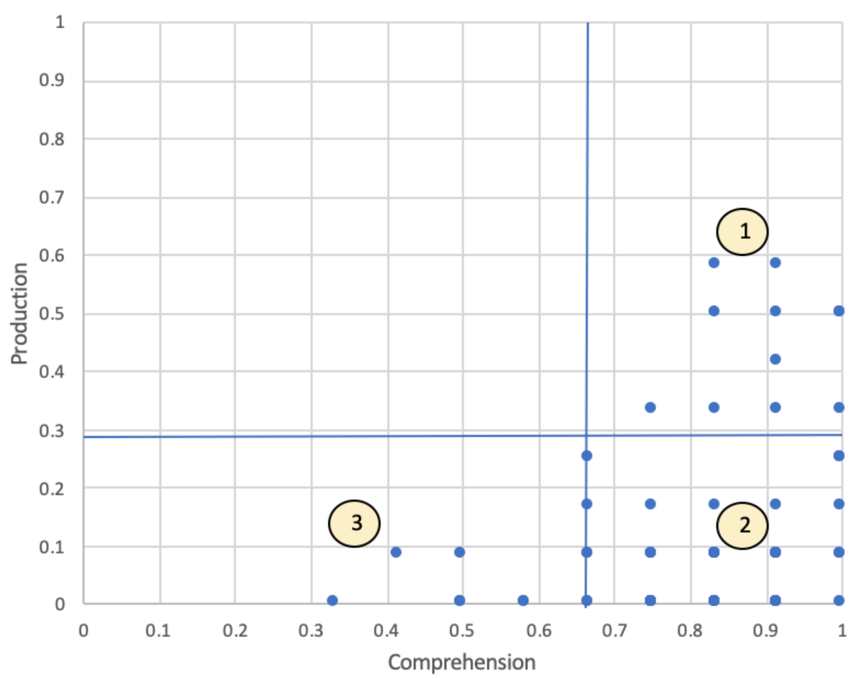


Figure 13. Correlation scatterplot between comprehension and production in experiment 1. The numbers indicate three comprehension-production patterns. Pattern 1=high comprehension and relatively high production; Pattern 2=high comprehension and low production; Pattern 3=relatively low comprehension and low production.

Table 7 shows descriptive data of the number of children, age, maternal education, vocabulary, categorization, ITALK score, comprehension accuracy and production accuracy for the three patterns. A majority of children was categorized in pattern 2, which demonstrated high comprehension and low production. Children in pattern 1 were older than those in pattern 2 and 3. Maternal education of pattern 1 was slightly higher than pattern 2 and 3. Regarding vocabulary, there was a decreasing trend from pattern 1 to 3. Finally, children in pattern 1 and 2 appeared to be more likely to choose objects by shared classifiers than children in pattern 3. However, the number of children falling in pattern 3 is very small, so the tendency needs to be interpreted with caution.

Patterns		1	2	3
n		11	46	7
Age (year:month)	Mean (SD)	5:11 (5.9)	4:11 (7.5)	4:4 (2.9)
	Range	4;10-6;5	4;1-6;3	4;1-4;8
Maternal Education	Mean (SD)	4.5 (0.7)	3.9 (0.6)	4 (0.6)
	Range	3-5	3-5	3-5
Vocabulary	Mean (SD)	0.83 (0.07)	0.69 (0.13)	0.54 (0.10)
	Range	0.72-0.94	0.41-0.94	0.43-0.72
Categorization	Mean (SD)	0.74 (0.21)	0.72 (0.13)	0.64 (0.18)
	Range	0.25-1	0.33-0.92	0.33-0.92
ITALK	Mean (SD)	4.7 (0.20)	4.6 (0.22)	4.3 (0.39)
	Range	4.3-4.9	4-5	3.7-4.7
Comprehension	Mean (SD)	0.90 (0.08)	0.84 (0.10)	0.49 (0.09)
	Range	0.75-1	0.67-1	0.33-0.58
Production	Mean (SD)	0.44 (0.10)	0.07 (0.07)	0.02 (0.04)
	Range	0.33-0.58	0-0.25	0-0.08

Table 7. Characteristics of the three comprehension-production patterns. Pattern 1=high comprehension and relatively high production; Pattern 2=high comprehension and low production; Pattern 3=relatively low comprehension and low production. SD of age is in month. Maternal education was ranked: 1=middle school or below; 2=high school; 3=associate degree; 4=Bachelor's degree; 5= Master's degree or above. The ITALK score is the average of parent's and teacher's rating.

EXPERIMENT 2

In this experiment, we manipulated frequency of classifier input using a between-subject design, with 34 of the children being taught with a higher dosage of classifier input and 30 children being taught with a lower dosage of classifier input. Table 8 displays demographic information for the two frequency groups. The two frequency groups were comparable in age and maternal education. Independent samples t-tests showed that there were no significant group differences in their age ($t(62)=.28$, $p=.78$) and maternal education ($t(62)=.36$, $p=.72$).

Demographic Measures		Frequency Group	
		Lower-frequency Group	Higher-frequency Group
n		30	34
Gender		17:13	18:16
Age (year;month)	Mean (SD)	5;1 (8.4)	5;0 (8.9)
	Range	4;1-6;5	4;1-6;3
Maternal Education	Mean (SD)	4.0 (0.6)	4.1 (0.6)
	Range	3-5	3-5

Table 8. Demographic information for the two frequency groups in experiment 2. SD of age is displayed in month. Maternal education was ranked: 1=middle school or below; 2=high school; 3=associate degree; 4=Bachelor's degree; 5=Master's degree or above.

Since the main differences between experiment 1 and 2 are whether frequency and familiarity were manipulated, we display accuracy of comprehension and production for invented classifiers by frequency group and familiarity. See Table 9 and 10. As generalized linear mixed models were administered in the following, we did not conduct statistical analyses here. By observing Table 9, accuracy of comprehension and production was largely consistent between the lower and higher frequency groups. From Table 10, it appears that comprehension and production accuracy was highly similar in the two familiarity conditions.

Classifier Measures		Frequency Group	
		Lower-frequency Group	Higher-frequency Group
Comprehension	Mean (SD)	0.69 (0.22)	0.76 (0.23)
	Range	0.38-1	0.29-1
Production	Mean (SD)	0.20 (0.36)	0.20 (0.34)
	Range	0-1	0-0.96

Table 9. Comprehension and production accuracy by frequency in experiment 2.

Classifier Measures		Familiarity	
		Familiar Object Condition	Novel Object Condition
Comprehension	Mean (SD)	0.73 (0.45)	0.72 (0.45)
	Range	0-1	0-1
Production	Mean (SD)	0.19 (0.39)	0.20 (0.40)
	Range	0-1	0-1

Table 10. Comprehension and production accuracy by familiarity in experiment 2.

Two generalized linear mixed models with binary data were conducted to analyze the results of comprehension and production in experiment 2. Dependent variables were accuracy on comprehension and production respectively (0 indicates incorrect responses; 1 indicates correct responses). We entered frequency group (1 indicates lower frequency; 2 indicates higher frequency), familiarity condition (1 indicates the familiar object condition; 2 indicates the novel object condition), categorization scores (percentages of preference to similarity by classifiers from experiment 1), vocabulary (averaged accuracy of receptive vocabulary and expressive vocabulary scores from experiment 1), age group (1 indicates the younger group; 2 indicates the older group), classifier (1 indicates “dě”; 2 indicates “tá”), maternal education (from 3 to 5 - associate to Masters’ degree or above) and form (1 indicates form A and 2 indicates form B) as independent variables. On the basis of our hypotheses, we entered two 2-way interactions, including categorization*age and categorization*frequency.

Comprehension

In comprehension, categorization was a marginally significant predictor, $F(1, 1524)=3.00$, $p=.08$, odds ratio=0.012. Children who preferred to match objects that shared the same classifiers tended to perform better on the comprehension task in experiment 2. Per 1% increase in classifier-based categorization, there was 1.2% increase in the novel classifier comprehension. All the other predictors were not significant, including age group (F(1, 1524)=.14, *p*=.71, odds ratio=0.46), frequency (F(1, 1524)=.16, *p*=.69, odds ratio=0.42), vocabulary knowledge (F(1, 1524=1.25, *p*=.26, odds ratio=0.15), familiarity (F(1, 1524)=.04, *p*=.85, odds ratio=0.98), classifier (F(1, 1524)=1.19, *p*=.27, odds ratio=1.15), maternal education (F(1, 1524)=.12, *p*=.88, odds ratio between 3 and 5=1.40, odds ratio between 4 and 5=1.12) and form (F(1, 1524)=.55, *p*=.46, odds ratio=1.38). The interactions of categorization*age group (F(1, 1524)=.42, *p*=.52, odds ratio=6.48) and frequency*categorization (F(1, 1524)=.52, *p*=.47, odds ratio=8.91) were not significant.

Production

In production, the main effect of classifier was significant, F(1, 1524)=18.52, *p*<.001, odds ratio=2.95. "Tá" yielded higher accuracy in production than "dě", and children were 2.95 times more likely to accurately produce "tá" than "dě". There were no significant main effects of age group (F(1, 1524)=.006, *p*=.94, odds ratio=0.61), frequency (F(1, 1524)=1.09, *p*=.30, odds ratio=0.58), vocabulary knowledge (F(1, 1524)=2.39, *p*=.12, odds ratio=0.001), familiarity (F(1, 1524)=.37, *p*=.54, odds ratio=1.16), categorization (F(1, 1524)=.41, *p*=.52, odds ratio=0.008), maternal education (F(1, 1524)=.05, *p*=.94, odds ratio between 3 and 5=1.76, odds ratio between 4 and 5=1.49) and form (F(1, 1524)=.38, *p*=.54, odds ratio=0.47). The two 2-way interactions were not significant, including categorization*age group (F(1, 1524)=.002, *p*=.97, odds ratio=1.40) and frequency*categorization (F(1, 1524)=1.16, *p*=.28, odds ratio=0.00).

It is common for children to use the general classifier "个 gè" to replace the targeted invented classifiers. Actually, 40 out of the 64 children used "个 gè" exclusively in their responses. Another strategy was to use one of the two invented classifiers in all test items. Between the two invented classifiers, children tended to use "tá" more often than "dě". Five children used "tá" throughout whereas only one child used "dě" throughout. This may be related to the significant main effect of classifier in production. It is likely because the rising tone of "tá" is louder and easier to produce than the falling and rising tone of "dě".

Comprehension and production

We explored the correlation between comprehension and production in experiment 2. Spearman's rho correlation test showed a significant correlation (r_s=.386, *p*=.002). Similar to experiment 1, children demonstrated better performance on comprehension than production. While comprehension accuracy was 72.5% on average, production accuracy was only 20%. From Figure 14, it seems that children who could produce the targeted invented classifiers also showed high accuracy in comprehension, indicating that comprehension preceded production. The accuracy of comprehension displayed a continuum, ranging from 0.3 to 1. However, the accuracy of production appeared to be mostly at two extremes, either near ceiling or down at floor. The six children who achieved 50% accuracy in production were the children who used one of the two invented classifiers exclusively in the production task.

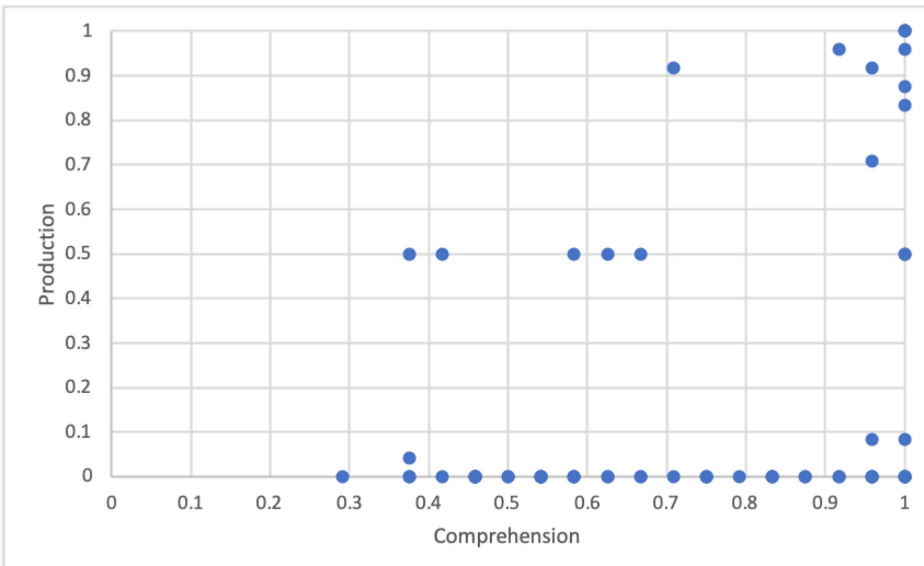


Figure 14. Correlation scatterplot between comprehension and production in experiment 2

Correlations between performance on experiment 1 and 2

We examined whether children's performance on real classifiers (experiment 1) was related to their performance on invented classifiers (experiment 2). Spearman's rho correlation test did not show a significant correlation regarding the performance on comprehension accuracy between experiment 1 and experiment 2 (r_s=.071, p=.579). However, there was a significant correlation on production accuracy between real and novel classifiers in the two experiments (r_s=.267, p=.03).

Chapter 4: Discussion

In the current study, we explored contributions of classifier-based semantic categorization, semantic relatedness, input frequency of classifiers and vocabulary knowledge to Mandarin-speaking children's learning of classifiers. Sixty-four monolingual Mandarin-speaking children between 4;1 to 6;5 completed two experiments. In experiment 1, we measured their comprehension and production of six real classifiers that are opaque in semantic categorization. Semantic categorization strategies and vocabulary knowledge were tested. Semantic relatedness was manipulated in the picture selection task. In experiment 2, we taught children two invented classifiers that were transparent in semantic categorization. We examined their comprehension and production of these invented classifiers, and input frequency of the two novel classifiers was manipulated across different participants. We included in experiment 2 the data of semantic categorization and vocabulary knowledge that was collected in experiment 1. The results of experiment 1 showed that for real classifiers, vocabulary knowledge predicted comprehension and production, whereas object categorization by shared classifiers did not. When children made comprehension errors, they were more likely to pair a classifier with a noun that should be paired with a semantically-related classifier than a semantically-distant classifier. In addition, age and maternal education predicted (or marginally predicted) the production accuracy of real classifiers. The findings of experiment 2 were that for invented classifiers categorization strategies marginally predicted comprehension, but input frequency and vocabulary were not significant predictors of both comprehension and production. Based on findings from the two experiments, the primary approach that Mandarin-speaking children take to learn classifiers depends on the transparency of the classifier semantic system. In both experiment 1 and 2, children on average demonstrated higher accuracy in comprehension than production. Below we discuss our findings in regard to each of the predictors of interest.

SEMANTIC CATEGORIZATION

We hypothesized that Mandarin-speaking children who were more sensitive to semantic categorization by shared classifiers would demonstrate more mature classifier knowledge (Kuo & Sera, 2009; Sera et al., 2013). In the categorization task, we presented a contrast of semantic categorization on the basis of shared classifiers (e.g., pencil and bamboo) and thematic relations (e.g., pencil and eraser). We focused on children's preference to noun categorization by classifiers. This preference was then examined as a potential predictor of children's existing knowledge of real classifiers (experiment 1) and learning of invented classifiers (experiment 2). The role of classifier-based semantic categorization differed across the two experiments. While classifier-based categorization was not related to the comprehension and production of real classifiers, it was a marginally significant predictor for children's comprehension of invented classifiers.

Distinguishing real and invented classifiers

A major difference between real classifiers and invented classifiers is whether classifier-based semantic categorization was clear-cut and transparent. The semantic organization for real classifiers appears to be complex and opaque. As we mentioned in the introduction, "条 tiáo" can be paired with *snake*, *fish, tie*, *scarf*, *necklace* and *rope* that are long, narrow and flexible. In addition, "条 tiáo" can be paired with *news*, *clue, instruction* and *message* that are apparently not long, narrow and flexible. For another example, "只 zhī" is typically paired with animals. It is also paired with singletons in a pair, like *earring*, *sleeve*, *shoe* and *glove*. In Shanghai dialect, "只 zhī" is widely used in a variety of circumstances (e.g., *bucket*, *stool*, *spoon*), and it is hard to determine its core semantic representations. Our sample of children were from Beijing, where standard Mandarin is used. Although classifier usage may be less flexible compared to many dialects in southern areas of China, the extraction of typical semantic properties for different real classifiers is still challenging and may almost be a mission impossible. As a result, children appeared to be less likely to take a rule-based approach by summarizing core semantic features for different real classifiers.

The semantic representations of the two invented classifiers were very transparent compared to real classifiers. One invented classifier was associated with fluffy and curly-haired entities, and the other related to entities that were broken. Importantly, there were no exceptions to the semantic classification of the two novel classifiers. The learning environment for these novel classifiers thus was less noisy than that of real classifiers which has many exceptions to general semantic categorization. When the semantic categorization was clear and organized, Mandarin-speaking children appeared to rely more heavily on a rule-based approach, and we saw a marginally significant correlation between object categorization by classifiers and novel classifier comprehension. In the categorization task of the current study, we used real-life objects to index children's semantic categorization strategies. In the future, a categorization task including novel objects may be used to examine object categorization strategies, and novel object categorization may be more strongly related to novel classifier comprehension.

Although classifier-based categorization was marginally significant in novel classifier comprehension, it did not significantly correlate with novel classifier production. A possible reason is that most children were unable to produce the targeted novel classifiers. As reported in the results section, 40 out of 64 children used the general classifier "个 gè" exclusively in their production to replace the targeted novel classifiers. Six children used "dě" or "tá" throughout their production, which also suggested that they could not produce the two novel classifiers. Since the production of novel classifiers was too challenging for most children, data from this task was less sensitive to reflect their knowledge of the two novel classifiers than the comprehension task.

Comparing to Uchida and Imai (1999)

Our findings challenge Uchida and Imai (1990)'s proposal about the indispensable role of semantic categorization strategies in classifier development. Not as they proposed, the main effect of semantic categorization and the interaction between age and categorization were not significant in either experiment. Uchida and Imai (1999) appeared to overly emphasize the contribution of semantic categorization. Mandarin-speaking children did not appear to rely very heavily on semantic categorization, as the semantic categorization of Mandarin real classifiers is opaque.

However, it is worth noting that findings in Mandarin may not be applied to another classifier language. If semantic features of classifiers in that language is more transparent than Mandarin, categorization strategies may be related to classifier learning. The Japanese classifier system seems to be less opaque than the Mandarin classifier system. For example, while Japanese classifiers strictly distinguish animacy, Mandarin classifiers do not. In the same research review, the researchers reported that classifier comprehension and production accuracy was higher in Japanese-speaking children from age four to six than Mandarin-speaking children at the same age (Uchida & Imai, 1999). Transparency of the classifier system may explain the accuracy gap between Japanese-speaking and Mandarin-speaking children.

Comparing to Sera et al. (2013)

Our results also differed from Sera et al. (2013), in which the probability of categorizing objects by real classifiers was significantly higher in children at three, five and seven who demonstrated better real classifier comprehension. In our study, categorization preference by real classifiers was not significantly related to 4 to 6-year-old Mandarin-speaking children's comprehension and production of real classifiers. Note that Sera et al. (2013) did not consider vocabulary knowledge. If vocabulary was controlled, the correlation may not be significant in their study.

Another potential reason to explain the difference between our findings and their findings is related to the age difference. Sera et al. (2013) did not specify the correlation between semantic categorization and classifier accuracy in each age group. A stronger correlation would be likely to be found in older Mandarin-speaking children. According to Ibbotson (2013), language development is accompanied by improved cognitive categorization and pattern finding in older children, whereas the correlation was not as strong in younger children who pay more attention to idiosyncratic constructions. Despite the difference, averaged percentages of categorization preference by classifiers were similar across Sera et al. (2013) and our study. In both studies, categorization preference by classifiers was around 70% by Mandarin-speaking children.

Potential shape bias in the categorization task

There could be a shape bias in the categorization task we administered. On one hand, the target objects based on shared classifiers (e.g., bamboo) were perceptually similar to the reference objects (e.g., pencil). However, semantic features of most sortal classifiers are based on shape and size (Erbaugh, 1986), and the six sortal classifiers we selected were based on shape and size. It is thus hard to disentangle shape-based semantic categorization and classifier-based semantic categorization. In order to control for the shape bias, future studies may consider other classifiers that are not shape-based, but animacy-based (e.g., 只 zhī) and function-based (e.g., 把 bǎ).

On the other hand, the instructions we used in the categorization task may have an influence on children's selections. We followed Sera et al. (2013), in which the Mandarin instruction was "哪一张和这一张比较像?" (*which of these pictures is more similar to this one?*). The usage of "比较像" (*more similar*) in Mandarin may prompt the children to pay attention to perceptual similarity (i.e., shape). Previous research has shown that children's object categorization was sensitive to experimenters' instructions. For example, Diesendruck and Bloom (2003) presented Hebrew-speaking children a target novel object with a novel name (e.g., Patoo). In an object selection task, three novel objects were provided that were of the same shape, material or color to the initial novel object. Using a between-subject design, three groups of children were randomly assigned to the three types of instructions: 1) *which one of these is also a Patoo*? 2) *Which one of these is of the same kind like this?* 3) *Which one of these goes with this?* They found that the first two types of instructions yielded about 75% shape choice. Children hearing the last type of instruction chose a shape choice and a choice by other links equally often. While there is a natural tendency to attend to shape in children's early word learning, "go with" actually implies thematic relations in children's experiences (e.g., socks *go with* shoes) (Colunga & Smith, 2008). Note that the corresponding translation of "go with" (i.e., 在一起， 放在一起）in Mandarin may imply the sorting by co-occurrence. Future studies should continue exploring the influence of verbal instructions on the tendency of object categorization.

SEMANTIC RELATEDNDESS

We predicted that semantic relatedness would influence the learning of classifiers. In experiment 1, we manipulated semantic relatedness in the three objects of each test plate in the forced-choice selection task. Previous studies did not control for the semantic relatedness of selection objects, which yielded varied difficulty levels for different picture-selection trials (e.g., Li et al., 2010). We were the first to do so. The results showed that a majority of children's errors were selections of an object that should be paired with a semantically-related classifier to the target classifier (i.e., 81.2%). Only 18.8% of the errors were selecting an object that should be paired with a semantically-distant classifier. This suggests that the detection of obvious semantic differences is easier and happen earlier in classifier development (e.g., long vs. flat). However, it should take longer time for Mandarin-speaking children to distinguish more fine-grained differences (e.g., long and flexible vs. long and rigid).

In production, most errors were using the general classifier "个 gè" to replace the specific targeted classifiers (i.e., 84%). This finding is consistent with the "个 gè" generalization in many previous studies (e.g., Ying et al., 1983; Tse, Li, & Leung, 2007). Both semantically-related errors (1.8%) and semantically-unrelated errors (0.7%) were infrequent in children's classifier production. We focused on the nature of semantically-related substitutions (Table 6). Seven out of the 14 errors were using "四块面包" *four kuài (cube-like) bread* for "四片面包" *four piàn (slice) bread*. We suspect that these children may be more likely to see cube-like bread in their environment, and caregivers of these children may more frequently use "四块面包" *four kuài (cube-like) bread* in daily communication with children. The learning of classifiers may be heavily related to frequency and children's familiarity of the classifier-noun combination in real life.

The other semantically-related substitutions again revealed that children needed to make more fine-grained distinctions among semantically-related classifiers. "幅 fú", "片 piàn" and "张 zhāng" are all associated with flat objects but could not be paired with *wall*. "幅 fú" is specifically paired with *paintings*, "片 piàn" is typically paired with slices, and "张 zhāng" typically should be paired with *papers*. To gain the knowledge about these specific pairings, children may need to gain more language exposure. Also, they may need to memorize each classifier-noun combination and receive training about these combinations in their Chinese language classes. Similarly, other errors indicate that children need to make fine-grained differences for semantically-related classifiers. "颗 kē" and "块 kuài" cannot be paired with *rice*, as they typically are associated with larger cube-like objects. "颗 kē" should not be paired with *cake* as it encodes the feature of "round", and the cake we presented is not round.

In the previous section, we discussed the contribution of classifier-based categorization which was not a significant predictor in real classifier comprehension and production. Here, by exploring children's errors in comprehension and production, we found that they did distinguish semantic features and were influenced by semantic similarity. Thus, we refine our understanding of the role of semantic categorization, and we do not completely exclude semantic categorization as a predictor of real classifier learning. Our position is that children do take rule-based approach to categorize semantic features, but not as heavily as item-based approach, which is elaborated in the following two sections.

FREQUENCY OF INPUT

Based on the usage-based approach (Tomasello, 2003; Ibbotson, 2013), we hypothesized item-based learning and the significant contribution of input frequency to the learning of classifiers. In experiment 1, we investigated individual classifier-noun combinations for the same real classifiers and found some large accuracy gaps between the two combinations, suggesting item-based learning. In experiment 2, we manipulated input frequency of the invented classifiers but did not find a significant correlation. In the following, we will discuss the mixed evidence.

Item-based learning of classifier-noun pairs

In experiment 1, we explored comprehension and production accuracy of the two test trials for each real classifier. We found that there were some accuracy gaps between the two classifier-noun pairs (Figure 9 and 11). In production, the accuracy gap between "zhī-pencil" (19%) and "zhī-candle" (5%) was relatively large. The pairing of *zhī* and *pencil* should be more frequent and familiar to children than the paring of *zhī* and *candle*, given candles are not used in either the school setting or the home setting these days. The most striking accuracy gap between the two test trials is the production of "片 piàn". While the production accuracy for "piàn-leaf" combination was 42%, the production accuracy for "piàn-bread" was only 8%. We speculate that the "piàn-leaf" combination was specifically used at school in book reading or some class activities (e.g., making handcraft using leaves), potentially resulting in higher frequency in children's language input and thus higher accuracy in production.

Children learn certain classifier-noun combinations better than others, providing some evidence for an item-based learning approach in classifier acquisition. Remember that the usage-based model emphasizes idiosyncratic learning of concrete words and phrases (Tomasello, 2003) as a powerful language learning strategy, and children's learning outcomes are largely shaped by their language input (Cameron-Faulkner et al., 2003). Therefore, the idiosyncratic learning of classifier-noun combinations is likely to be related to varied frequency of exposure to the different classifier-noun combinations in these Mandarin-speaking children's language environment. As we mentioned in the last paragraph, higher accuracy of "piàn-leaf" than "piàn-bread" could be a result of the more frequent usage of "piàn-leaf" in children's school activities. The current findings improved our understanding and guided us to pay more attention to frequency of classifier-noun combinations in addition to frequency of classifiers in future studies.

The lack of frequency effect for novel classifiers

In experiment 2, we manipulated input frequency of the two invented classifiers. Children were randomly assigned to a lower frequency group and a higher frequency group. Results showed that there were no main effects of frequency in the comprehension and production of invented classifiers. For both frequency groups, the averaged comprehension accuracy was about 70%, and averaged production accuracy was around 20%. Also, there was no interaction between input frequency and categorization.

One possible explanation to the lack of frequency effect is that there were individual variations regarding different styles of language acquisition (Nelson, 1973; Kohnert & Danahy, 2007). Children taking an analytical approach would be more sensitive to object labels, and classifiers are noun morphemes that are corresponding to semantic features of objects. The task demand thus matched their learning style. However, other children with an expressive learning style would be more sensitive to personal and social words and less sensitive to object labels. Their learning style thus did not match the task demand in experiment 2.

Another possible explanation is related to individual differences in executive functions (e.g., visual working memory, attention suppression). Research has shown that visual memory capacity differs across individuals. For example, Vogel and Machizawa (2004) found that among 12 college students, the mean visual capacity was 2.8 objects, but individual differences ranged from 1 to 4 objects. In our study, the two novel classifiers represented perceptual differences in shape, and visual working memory was perhaps required for children to store and make contrast between two types of visual input (i.e., curly-haired and broken). Individual differences with regard to children's visual working memory thus would contribute to variations of performance on novel classifier learning. Moreover, the ability to control attention has been shown to be involved in visual working memory (Gulbinaite, Johnson, de Jong, Morey, & van Rijin, 2014). Individuals with stronger visual working memory tended to be more capable of suppressing irrelevant information.

In the future, the above factors may be considered and controlled when comparing the lower and higher frequency groups. For example, an IQ test could be conducted, in which skills that support rule abstraction can be measured. An approach to bypass controlling these factors is to use a within-subject design. A child would receive higher frequency input for one novel classifier and lower frequency input for the other novel classifier. The two novel classifiers should be counterbalanced for their corresponding semantic properties and phonological forms.

In addition, the frequency gap could be enlarged between the higher-frequency group and the lower-frequency group. Currently, there were 42 times exposure to each novel classifier in lower-frequency group, and the higher-frequency group doubled the exposure in the lower-frequency group. Averaged comprehension accuracy was 69% in the lower-frequency group and 76% in the higher-frequency group. The difference of accuracy between the two frequency groups may be bigger after enlarging the frequency gap.

Finally, as noted in the previous section, in the future we should focus our attention on frequency of individual classifier-noun combinations. The kind of frequency we manipulated in experiment 2 (i.e., the frequency of individual classifiers) is a part of the total experience with classifiers and may contribute less to the success with classifier learning. Classifier-noun combinations are more idiosyncratic expressions that may account for more refined variations with regard to constructions with classifiers.

Generalization of classifier knowledge

We planned to control for exposure variations to the range of nouns that can be paired with the same classifier and included a novel object condition in addition to a familiar object condition. In experiment 1, due to the low adult agreement on one real classifier (i.e., 片 *piàn*), we excluded the novel condition in children's test. In experiment 2, as the semantic features of the two invented classifiers strictly aligned with the semantic categorization, we were able to include a novel object condition. Similar to previous studies (e.g., Li et al., 2010; Fang, 1985; Uchida & Imai, 1999; Culbertson & Wilson, 2013), there was no effect of familiarity. This indicates that if Mandarin-speaking children could learn the semantic features encoded by a novel classifier, they were able to generalize the knowledge of classifiers to novel objects.

VOCABULARY

Vocabulary has not been explored as a predictive factor in previous work. Vocabulary may have an influence on classifier acquisition in two different ways. First, many classifiers are used as nouns in Mandarin noun-noun compounds (e.g., 湖面 hú **miàn**-lake **surface**; 桌面 zhuō **miàn**-table **surface**; 墙面 qiáng **miàn**-wall **surface**). In these noun-noun compounds, 面 miàn means surface, which shares the same semantic property as the classifier 面 miàn.

More mature vocabulary knowledge should be related to more accurate understanding and usage of classifiers. Second, children's size of vocabulary could be related to their needs of semantic categorization. A child with larger vocabulary may have a stronger need to categorize words, so they can retrieve and access words promptly. The results showed that there was a significant main effect of vocabulary in both comprehension and production of real classifiers.

Classifier learning and noun learning

It is worth noting that only nouns were included in the vocabulary test, and there were no other classes of words (e.g., verb, adjective, adverb). To be more accurate about this correlation, noun acquisition was significantly correlated with real classifier acquisition. Children who demonstrated better performance on Mandarin nouns tended to achieve higher accuracy on Mandarin real classifier comprehension and production. The findings indicate that some similarities between classifier learning and noun learning.

The correlation between classifier learning and noun learning indicates that classifier learning could be item-based, as how children learn most nouns. This supports our position that to learn real classifiers, Mandarin-speaking children take an item-based approach primarily. An additional potential link between classifier learning and noun learning is that both are heavily shape based. Research has consistently shown a predominant tendency for children to extend a novel name to a novel object that shares a common shape (e.g., Gershkoff-Stowe, & Smith, 2004; Yee, Jones, & Smith, 2012). It was also found that the awareness of shape categorization is a reliable early index of children's noun learning. The majority of Mandarin classifiers has a cognitive base on shape (Zhang, 2007). The six classifiers we included in this study are associated with different shapes (i.e., long, flat, cube-like). Children's sensitivity to shape features may be a mechanism underlying both noun learning and classifier learning.

Despite the similarities of classifier learning and noun learning, the two were essentially different. While word learning is lexical learning, classifier learning falls in between the continuum of grammatical learning and lexical learning. Also, the rate of acquisition is different. Noun learning to start is largely fast mapping, which develops fast in early childhood (Hoff, 2013). Classifier learning cannot be fast mapping, and it grows slowly and gradually. It only begins to take off in the production modality when children reach school age. In this study, children's averaged accuracy in real/novel classifier production was only around 15%. Last but not least, semantic organization for classifiers is more complex and unintuitive compared to semantic organization for nouns. “Long, narrow and flexible” pertaining to “条 tiáo” is not a common combination of semantic properties.

Vocabulary knowledge as a reliable index of real classifier development

Vocabulary knowledge appeared to be a more reliable index of real classifier development than age. Vocabulary knowledge significantly predicted children’s accuracy of real classifier comprehension and production. Yet, a significant correlation was found between age and real classifier production but not real classifier comprehension. While averaged accuracy of real classifier comprehension was already high at 82%, averaged accuracy of real classifier production was low at 13%. Production task is more challenging thus more sensitive than comprehension task to reflect age-related changes in Mandarin-speaking children between ages four and six.

OTHER ISSUES

Consistent with previous findings (e.g., Chien et al., 2003; Fang, 1985; Li et al., 2010; Ying et al., 1983), comprehension preceded production in classifier development. In both experiments in our study, children’s accuracy was much higher in comprehension than production. In Figure 13, we explored three comprehension-production patterns. The “good comprehension-relatively good production” group demonstrated the most advanced classifier understanding and usage. The “good comprehension-poor production” group showed emerging usage of specific classifiers. The “relatively poor comprehension-poor production” group struggled with both classifier comprehension and production. We then reported descriptive data of age, maternal education, vocabulary, classifier-based categorization preference and ITALK scores for the three pattern groups (Table 7). We found that age and vocabulary were more reliable indices of classifier development, which was consistent with results of the mixed model in experiment 1.

The two invented classifiers were very challenging for Mandarin-speaking children. In production, only seven children achieved an accuracy above 90%, 47 out of 64 children’s accuracy was below 10%, and accuracy of the remaining ten children fell between 10% and 90%. Children’s performance in our study was much lower than in Kohnert and Danahy (2007), in which TD Spanish-English bilingual children between 3;6 and 5;8 were tested. The researchers taught a novel morpheme indicating the whole-part distinction (wheel-whole vs.

wheel*ku*-part). For example, while children were looking at a picture of a wheel, the researcher verbalized "*This is a wheel*"; while children were looking at a partial wheel, the researcher said "*This is a wheelku*". Among the 20 children, eight children were able to produce the invented morpheme with more than 90% accuracy, seven children's accuracy was below 10%, and five children achieved an accuracy between 10% to 90%. See Figure 15 for pie charts to compare percentages of number of children who achieved different production accuracy.

We want to note important differences between the two learning tasks. Our learning task included two novel classifiers that were presented with 24 novel objects, whereas Kohnert and Danahy (2007) only included a novel morpheme which was illustrated using familiar objects. By presenting a familiar object and naming the object with a novel morpheme, the feature encoded by the novel morpheme was made very salient. The learning task in our study could be more complex with both novel objects and morphemes, which potentially yielded lower accuracy.

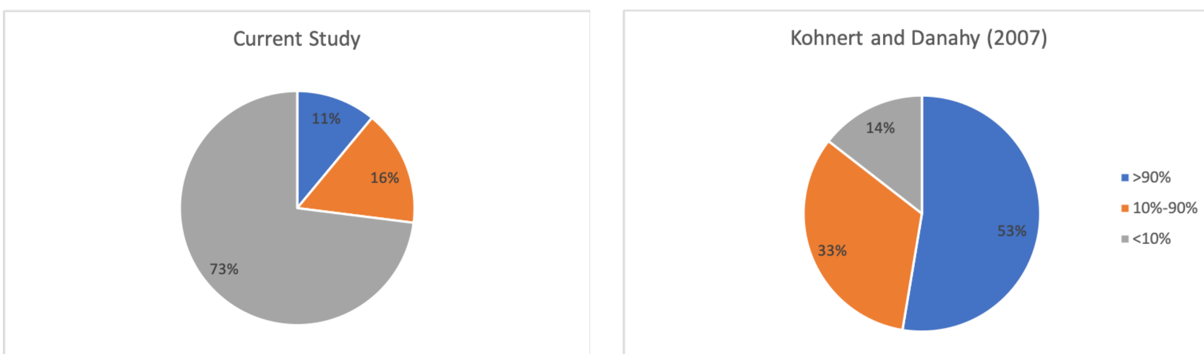


Figure 15. Comparison between the current study and Kohnert and Danahy (2007) with regard to percentages of number of children with different production accuracy

There was a main effect of classifiers in the production of the two novel classifiers. Children were more likely to produce "tá" than "dě". A strategy some children used was to use only one novel classifier in all the production trials. Five children used "tá" exclusively in their production, but only one used "dě" throughout. The second tone in "tá" is an easier tone to produce than the third tone in "dě". It has been found that the third tone is the last acquired tone for Mandarin-speaking children (before three years old), whereas the second tone is generally acquired earlier (at two years old) (So & Zhou, 2000). Children in our study were all above age four and should have acquired all the four tones in Mandarin. It is worth mentioning that in comprehension the same main effect was not present. This indicates that these children did not have a preference for the "broken" feature represented by "tá" over the "curly hair" feature represented by "dě". In future studies, the two novel classifiers (hence the two phonological forms) and their corresponding semantic features should be counterbalanced across participants. In addition, a better design would be to create novel classifiers that have the same tone.

Production accuracy on the two experiments was significantly correlated ($p=0.03$). If children demonstrated better performance on real classifier production, they were more likely to achieve higher production accuracy in invented classifiers. This shows that the two experiments had consistency in the content of examination. Particularly, the novel classifier learning task could indicate children's abilities and achievement of classifier learning. Within each experiment, comprehension accuracy and production accuracy were significantly correlated. If children achieved high accuracy in comprehension, they demonstrated relatively high performance in production.

Maternal education was entered into the mixed models as a potential confound, and there was a marginally significant correlation between maternal education and real classifier production. It is widely accepted that maternal language input heavily shapes children's language (e.g., Hoff, 2006). Here, parents who received higher education may use more specific classifiers when conversing with their children, so their children may be exposed to more advanced usage of classifiers.

CONCLUSION

In this study, we explored how Mandarin-speaking children learn classifiers. Based on the extant literature, we investigated potential predictors of classifier learning. The results showed that depending on the transparency of the to-be-learned classifiers, children take different approaches. Real Mandarin classifiers are opaque in the rules of semantic categorization. Mandarin-speaking children between ages four and six relied more heavily on item-based learning of classifier-noun combinations than classifier-based semantic categorization. When taught invented classifiers that are transparent in semantic categorization, Mandarin-speaking children seemed to rely more heavily on rules of semantic categorization than an item-by-item approach. The current findings drew our attention to the frequency of classifier-noun pairs in future studies.

Now we go back to our motivations for this study. Classifiers are a challenging linguistic feature for monolingual Mandarin-speaking children with DLD and bilingual Mandarin-speaking children. To guide future directions in these populations, we studied predictors of classifier learning in TD Mandarin-speaking children. The findings help us generate hypotheses about potential difficulties pertaining to classifier learning (e.g., whether difficulties with classifiers are related to weaknesses in vocabulary; whether children with DLD have more difficulties distinguishing fine-grained semantic differences than TD children). As an interface phenomenon of syntax and semantics, the semantic aspect is more challenging than the syntactic construction. Assessment and intervention of classifiers in Mandarin-speaking with DLD should focus on the semantic aspect rather than the syntactic structure.

Appendices

Appendix 1.

Categorization Response Sheet (Form A)

Instruction: 小朋友，我们现在要来看一些图片。每次你都会先看到上面有一张图（指上面的图），然后看到下面的两张图（指下面的两张图）。你来告诉我（停顿）“下面哪一张和上面这张更像”，好吗？我们来试一试吧！We are going to look at some pictures. Every time, you will see one picture on top, and then two pictures below. You will need to tell me which one below is more similar to top picture?. Let’s have a try!
Notes: 1) write **1** if a child chooses the left picture below, and write **2** if a child chooses the right picture below; 2) no corrections should be made in practice and real test.

Practice

No.	Classifier-Reference Picture	Responses	Notes
P1	个gè-apple		
P2	朵duǒ-flower		
P3	把bǎ-comb		

Test

No.	Classifier-Reference Picture	Responses	Notes
1	块kuaì-soap		
2	支zhī-pencil		
3	粒lì-pill		
4	支zhī-candle		
5	片piàn-bread		
6	条tiáo-scarf		
7	片piàn-leaf		
8	面miàn-mirror		
9	粒lì-rice		
10	条tiáo-necklace		
11	面miàn-wall		
12	块kuài-cake		

Appendix 2.

Experiment 1 Production Response Sheet (Form A)

Prompts: “有多少？” (*How many*), “两什么？” (*two what*), “多少书” (*how many books*?)
Notes: 1) Experimenters should strictly follow the prompts, and no additional prompts can be included; 2) No classifier should be included in prompts.

Practice

No.	Target Classifiers	Responses	Notes
P1	本 běn		
P2	把 bǎ		

Test

No.	Target Classifiers	Responses	Notes
1	片 piàn		
2	粒 lì		
3	片 piàn		
4	条 tiáo		
5	块 kuài		
6	面 miàn		
7	粒 lì		
8	条 tiáo		
9	块 kuài		
10	支 zhī		
11	面miàn		
12	支zhī		

Appendix 3.

Experiment 1 Comprehension Response Sheet (Form A)

Instruction: " 米 老 鼠 说他想要一___something." (*Mickey Mouse said that he wanted one___something.*)
Notes: 1) Write **1** if a child chooses the left picture, write **2** if a child chooses the middle picture, and write **3** if a child chooses the right picture; 2) In order for children to understand the task, corrections may be made in practice. No corrections should be made in testing items.

Practice

No.	Target Classifiers	Response	Notes	Target
P1	本 běn			3
P2	把 bǎ			1

Test

No.	Target Classifiers	Responses	Notes
1	片 piàn		
2	粒 lì		
3	片 piàn		
4	条 tiáo		
5	块 kuài		
6	面 miàn		
7	粒 lì		
8	条 tiáo		
9	块 kuài		
10	支 zhī		
11	面miàn		
12	支zhī		

Appendix 4.

Experiment 2 – Production Response Sheet (Form A)

Prompts: "有多少？" (*How many*), "两什么？" (*two what*), "多少 something" (*how many something*?)
Notes: 1) Experimenters should strictly follow the prompts, and no additional prompts can be included; 2) No classifier should be included in prompts.

Practice

No.	Target Classifiers	Responses	Notes
P1	tá		
P2	dě		

Test

No.	Target Classifiers	Responses	Notes
1	dě		
2	tá		
3	dě		
4	tá		
5	dě		
6	dě		
7	tá		
8	dě		
9	dě		
10	tá		
11	dě		
12	tá		
13	tá		
14	tá		
15	tá		
16	dě		
17	dě		
18	dě		
19	dě		
20	tá		
21	tá		
22	tá		
23	dě		
24	tá		

Appendix 5.

Experiment 2 – Comprehension Response Sheet (Form A)

Instruction: "米老鼠说他想要一___something." (*Mickey Mouse said that he wanted one___something.*)
Notes: Write **1** if a child chooses the left picture, write **2** if a child chooses the right picture; 2) In order for children to understand the task, corrections may be made in practice. No corrections should be made in testing items.

Practice

No.	Target Classifiers	Response	Notes	Target
P1	dě			1
P2	tá			2

Test

No.	Target Classifiers	Responses	Notes
1	dě		
2	tá		
3	dě		
4	tá		
5	dě		
6	dě		
7	tá		
8	dě		
9	dě		
10	tá		
11	dě		
12	tá		
13	tá		
14	tá		
15	tá		
16	dě		
17	dě		
18	dě		
19	dě		
20	tá		
21	tá		
22	tá		
23	dě		
24	tá		

References

Allan, K. (1977). Classifiers. *Language, 53*(2), 285.

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge, United Kingdom: Cambridge University Press.

Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.). (2015). *Assessing multilingual children: Disentangling bilingualism from language impairment*. Bristol, United Kingdom: Multilingual Matters.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67-82.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modelling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*(41), 17284-17289.

Baron, A., Bedore, L. M., Peña, E. D., Lovgren-Uribe, S. D., López, A. A., & Villagran, E. (2018). Production of Spanish Grammatical Forms in US Bilingual Children. *American Journal of Speech-Language Pathology, 27*(3), 975-987.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*(5), 425-455.

Bybee, J. (2007). *Frequency of use and the organization of language*. New York, United States: Oxford University Press.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*, 843-873.

Cheung, H. (2009). Grammatical characteristics of Mandarin-speaking children with specific language impairment. In S. Law, B. Weekes, & A. M.-Y. Wong (Eds.), *Language Disorders in Speakers of Chinese*. Bristol, United Kingdom: Multilingual Matters.

Chien, Y., Lust, B., & Chiang, C. (2003). Chinese children's comprehension of count-classifiers and mass-classifiers. *Journal of East Asian Linguistics, 12*(2), 91-120.

Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: The self-organization of skilled noun learning. *Developmental Science*, *11*(2), 195-203.

Culbertson, J., & Wilson, C. (2013). Artificial grammar learning of shape-based noun classification. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 2118-2123). Austin, TX: Cognitive Science Society.

Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development*, *74*(1), 168-178.

Erbaugh, M. S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In C. Craig (Eds.), *Noun Classes and Categorization* (pp399-436). Amsterdam and Philadelphia: John Benjamins.

Erbaugh, M. S. (2006). Chinese classifiers: Their use and acquisition. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics* (Vol. 1, pp. 39–51). Cambridge, United Kingdom: Cambridge University Press.

Fang, F. X. (1985). 4-6 岁儿童掌握汉语量词水平的实验研究[An experiment on the use of classifiers by 4-6-year-olds]. *心理学报, 4*, 384-352.

Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, *75*(4), 1098-1114.

Gulbinaite, R., Johnson, A., de Jong, R., Morey, C. C., & van Rijn, H. (2014). Dissociable mechanisms underlying individual differences in visual working memory capacity. *Neuroimage*, *99*, 197-206.

Hao, Y., Bedore, L. M., Sheng, L., & Peña, E. D. (2018). Narrative skills in two languages of Mandarin–English bilingual children. *International Journal of Speech-Language Pathology*, 1-11.

Hao, Y., Sheng, L., Zhang, Y., Jiang, F., Villiers, J. d., Lee, W., & Liu, X. L. (2018). A narrative evaluation of mandarin-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 61*(2), 345-359.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counselling and Development, 34*, 177–189.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*(1), 55-88.

Hoff, E. (2013). *Language development*. Wadsworth, CA: Cengage Learning.

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology*, *4*, 255.

Kaushanskaya, M., Gross, M., Sheena, E., & Roman, R. (2017). Novel Morpheme Learning in Monolingual and Bilingual Children. *American Journal of Speech-Language Pathology*, *26*(2), 301-315.

Kohnert, K., & Danahy, K. (2007). Young L2 learners' performance on a novel morpheme task. *Clinical Linguistics & Phonetics*, *21*(7), 557-569.

Kuo, J. Y. C., & Sera, M. D. (2009). Classifier effects on human categorization: The role of shape classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, *18*(1), 1-19.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1-27.

National Bureau of Statistics of China (2010). Tabulation on the 2010 population census of the People's Republic of China. http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm. accessed 8th December, 2017.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1-2, Serial No. 149).

Justice, L. M., Bowles, R. P., Pence Turnbull, K. L., & Skibbe, L. E. (2009). School readiness among children with varying histories of language difficulties. *Developmental Psychology*, *45*(2), 460.

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology*, *19*(1), 51-65.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd edition). Cambridge, Massachusetts: MIT press.

Li, P., Barner, D., & Huang, B. H. (2008). Classifiers as count syntax: Individuation and measurement in the acquisition of Mandarin Chinese. *Language Learning and Development, 4*(4), 249-290.

Li, P., Huang, B., & Hsiao, Y. (2010). Learning that classifiers count: Mandarin-speaking children's acquisition of sortal and mensural classifiers. *Journal of East Asian Linguistics, 19*(3), 207-230.

Li, C., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Li, X. X., & To, C. K. S. (2017). A review of phonological development of mandarin-speaking children. *American Journal of Speech-Language Pathology, 26*(4), 1262-1278.

Peña, E. D., Gutierrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. (2018). *Inventory to Assess Language Knowledge (ITALK) (English and Spanish bilingual edition)*. Baltimore, MD: Brookes Publishing.

Peña, E. D., Gutierrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. (2018). *BESA: Bilingual English-Spanish Assessment Manual*. Baltimore, MD: Brookes Publishing.

Salehuddin, K., & Winskel, H. (2009). An investigation into Malay numeral classifier acquisition through an elicited production task. *First Language, 29*(3), 289-311.

Sera, M. D., Johnson, K. R., & Kuo, J. Y. (2013). Classifiers augment and maintain shape-based categorization in mandarin speakers. *Language and Cognition, 5*(1), 1-23.

Sheng, L., Lam, B. P. W., Wang, D., & Zheng, L. (2019). Receptive and Expressive Vocabulary Screener of Mandarin. In progress.

So, L. K. H., & Zhou, J. (2000). *Putonghua Segmental Phonology Test*. China: Nanjing Normal University.

Stokes, S. F. & So, L. K. H. (1997). Classifier use by language-disordered and age-matched Cantonese-speaking children. *Asia Pacific Journal of Speech, Language and Hearing, 2*, 83-101.

Swisher, L., & Snow, D. (1994). Learning and generalization components of morphological acquisition by children with specific language impairment: Is there a functional relation? *Journal of Speech, Language, and Hearing Research, 37*(6), 1406-1413.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1245-1260.

Tomasello, M. (2003). *Constructing a language*. Cambridge: Harvard University Press.

Tse, S. K., Li, H., & Leung, S. O. (2007). The acquisition of Cantonese classifiers by preschool children in Hong Kong. *Journal of Child Language, 34*(3), 495-517.

Uchida, N., & Imai, M. (1999). Heuristics in learning classifiers: The acquisition of the classifier system and its implications for the nature of lexical acquisition. *Japanese Psychological Research, 41*(1), 50-69.

U. S. Census Bureau. (2011). Language use in the United States:2011. https://www.census.gov/prod/2013pubs/acs- 22.pdf. accessed 16th December, 2016.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748.

Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, MA: MIT Press.

Yee, M. N., Jones, S. S., & Smith, L. B. (2012). Changes in visual object recognition precede the shape bias in early noun learning. *Frontiers in Psychology*, *3*, 533.

Ying, H. C., Chen, G. P., Song, Z. G., Shao, W. M., & Guo, Y. (1983). 4-7岁儿童掌握量词的特点[Characteristics of 4-7-year-olds in mastering classifiers], *心理科学通讯, 26*, 24-32.

Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics, 16*(1), 43–59.

中国社会科学院语言研究所词典编辑室 (2013). *现代汉语词典[Modern Mandarin Dictionary]*. 北京：商务印书馆.