1. Introduction

Grammatical agreement is very common in human language. In many languages, functional categories have to agree on either syntactic features or semantic features with each other or lexical categories. For instance, in French, determiners and nouns must agree on gender features, e.g. un-masc chapeau-masc ‘a hat’. A number of acquisition studies provide cross-linguistic evidence that children at a young age are sensitive to grammatical feature agreement (e.g. Jolly & Plunkett 2008; Melançon & Shi 2015). French-learning 30-month-olds have access to representations of gender agreement between determiners and nouns which assist target object recognition (Melançon & Shi 2015). Similarly, English-learning 30-month-olds use plural morphology to aid novel word processing (Jolly & Plunkett 2008).

Unlike Indo-European languages where there are selection restrictions on grammatical feature agreement such as gender and case, classifiers in Mandarin Chinese have to agree with nouns in semantic features, such as shape, number, size or animacy (e.g. Tai 1994). In Mandarin Chinese, classifiers are employed to categorize or quantify nouns (e.g. Cheng and Sybesma 1998, Li and Thompson 1981). As a functional category heading a functional projection CIP illustrated in (1), the classifier is obligatory in the number-noun phrase. As (1) shows, the classifier tiao agrees with the noun scarf in shape.
Recent evidence suggests that adult speakers of classifier languages can process classifier-noun agreement under various conditions (e.g., Kwon et al. 2017; Hsu et al. 2014; Sakai et al. 2006; Zhang et al. 2012; Zhou et al. 2010). In a semantic prediction study, for instance, Kwon et al. (2017) used ERP to investigate Mandarin speaker’s processing of classifier-noun agreement in match and mismatch conditions (cf. 2a, b). They found the N400 effect elicited in the classifier position in the classifier-noun mismatch condition. This implies that their adult subjects already pre-activated the noun at least when they heard the classifier. These findings, together with findings from other studies (e.g., Altmann & Kamide, 1999, 2007; Delong et al. 2005; Marslen-Wilson W. D. 1975; Kwon et al. 2017; Traxler et al. 1997), point to the incremental and predictive nature of language processing, that is, the parser processes incoming words immediately as they are perceived and make predictions about the syntactic or semantic information of the upcoming words.

(2) a. Zhang Yimou zhidao de zhe bu dianying timing le
   ‘This movie that Zhang Yimou directed was nominated’

b. Zhang Yimou zhidao de zhe zuo dasha timing le
   ‘This building that Zhang Yimou directed was nominated’

The acquisition of classifiers in Mandarin Chinese has also attracted much attention. One of the acquisition issues is whether children have abstract representations of count-mass distinction in Mandarin Chinese. Previous experimental studies show that count-mass distinction is already represented in 3-year-old children’s grammar, but they disagree on how this distinction is represented (e.g., Chien et al. 2003; Huang 2009). Chien et al. (2013) holds that count-mass distinction is encoded lexically at the classifier level, while Huang (2009) proposes a syntax-first account. Another acquisition issue is whether children truly know the meaning of classifiers (e.g., Huang 2009; Li et al. 2008, 2010). Li et al (2010) shows that by age 3, children can use classifiers to recognize both known and novel objects/nouns and children under age 3 can also use the semantic content of several classifiers in object recognition. This indicates the emergence of children’s awareness of classifiers. In addition,
studies have investigated the acquisition age, order and the factors influencing children’s use of classifiers (Erbaugh 1986; Ding 1999; Fang 1985; Hu 1993; Wen 2013; Li et al. 2010; Ying et al. 1983; Yu 2013). Generally speaking, the general classifier ge is acquired the earliest before age 3, followed by specific shape classifiers such as tiao with scarce production at around 3 years of age. Children’s production of classifiers increases with age and is influenced by levels of cognitive development such as the ability in counting and reading (Fang 1985; Wen 2013; Li et al. 2010; Ying et al. 1983). However, very little is known about young children’s online processing of classifier-noun agreement. Answers to this question will not only provide a window to explore whether young children demonstrate incremental and predictive processing, but also bear directly to the issue of continuity of functional categories, i.e. whether functional categories are already available to infants and toddlers. Given that functional categories are much intertwined with grammar, it is hypothesized that functional categories form an inherent part of children’s early grammar (Hyams 1996; Poeppel & Wexler 1993) and they play a “skeleton” role throughout acquisition (Dye et al. 2019). Yang et al. (2018) shows that Mandarin-speaking two-year-olds have access to aspectual distinctions between perfective aspect marker le and imperfective aspect marker zhe and they rely on these aspect markers in recognizing ongoing and completed events on line. If we are able to show that Mandarin-speaking children of the same age encode semantic feature agreement between classifiers and nouns, it would provide further evidence, from the perspective of the acquisition of a language with no overt morphology, for the continuity hypothesis.

2. The present study

We asked whether Mandarin-learning 2-year-olds were sensitive to classifier-noun agreement in an IPLP (Intermodal Preferential Looking Paradigm) comprehension experiment. IPLP is a widely used method in studies tapping syntactic, semantic and phonological knowledge of children younger than age three (Hirsh-Pasek and Golinkoff 1996). In a typical IPLP experiment, children are presented with two pictures or videos side by side and at the same time presented with a speech stimulus matching only one of them. Their looking behavior throughout the experiment is recorded for later analysis.

2.1. Participants

32 Mandarin-learning 30-to 32-month-olds (Mean age: 2;7;20) participated in the experiment. All participants whose main caregivers were speakers of Mandarin Chinese were recruited from Beijing through online advertisements.

2.2. Test stimuli and design

Speech stimuli included shape classifiers and count nouns which were familiar to the child. They were chosen from the Beijing Chinese Early
Language Acquisition corpus (BJCELA)\(^1\) and were among the most frequently used in the adult input. Table 1 lists the 4 classifiers and 8 nouns used in this experiment. The nouns were depicted in pictures used as visual stimuli.

**Table 1. Classifiers and nouns in the experiment**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Semantic feature</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang</td>
<td>Flat</td>
<td>Kapian ‘card’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tiezhi ‘paster’</td>
</tr>
<tr>
<td>Tiao</td>
<td>Long</td>
<td>Weijin ‘scarf’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Liyu ‘fish’</td>
</tr>
<tr>
<td>Kuai</td>
<td>Block-like</td>
<td>Dangao ‘cake’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jimu ‘brick’</td>
</tr>
<tr>
<td>Gen</td>
<td>Stick-like</td>
<td>Binggun ‘popsicle’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Qianbi ‘pencil’</td>
</tr>
</tbody>
</table>

All the speech stimuli were in the form of one + classifier + adjective + Noun. A neutral adjective caisede ‘colorful’ was inserted between classifiers and nouns to allow processing time of classifiers and at the mean time to avoid collocation effect. For each classifier, there was a classifier-noun match trial and a classifier-noun mismatch trial, as illustrated in Table 2. Zhang and tiao were in pairs, and kuai and gen formed another pair. The tiao trials are presented in Table 2 as an illustration. In the match trial, the classifier matches the spoken noun (e.g. tiao vs. weijin ‘scarf’) whereas it mismatches the spoken noun (e.g. tiao vs. kapian ‘card’) in the mismatch trial. As tiao formed a pair with zhang, the same nouns/objects (i.e. weijin ‘scarf’, kapian ‘card’) also served as stimuli in zhang trials, but zhang matched kapian ‘card’ and mismatched weijin ‘scarf’. Each child was presented with 8 trials, 4 classifier-noun match trials and 4 classifier-noun mismatch trials. In each trial, two side-by-side objects were displayed, one representing the noun that was named in the auditory stimulus, and the other a distractor object. All stimuli were created by cross-splicing different parts of recorded classifier-noun matching sentences to avoid producing unnatural classifier-noun mismatching utterances.

**Table 2. TIAO test trials in match and mismatch condition**

<table>
<thead>
<tr>
<th>Visual stimuli</th>
<th>Auditory stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1 Match</td>
<td><em>Kan! Yi tiao caisede weijin</em> Look a Cl_long colorful scarf</td>
<td></td>
</tr>
<tr>
<td>Trial 2 Mismatch</td>
<td><em>Kan! Yi tiao caisede kapian</em> Look a Cl_long colorful card</td>
<td></td>
</tr>
</tbody>
</table>

---

\(^1\) The corpus was built as part of a Chinese early language project led by Thomas Lee at the Chinese University of Hong Kong.
The child went through a picture parade in which all test pictures were presented one after another at the center of the screen with their names spoken. Children first received a teaching phase with two pictures on the LCD screen while listening to the audio stimulus matching one of the pictures, just as in Table 2. The only difference from the test stimuli was that the audio in the teaching phase ended with a question *zai nabian a* ‘on which side’ which was added to teach the child to look at the target picture.

Test phase followed immediately. The duration of each test trial was 7.5s. Figure 1 is the timeline of the stimulus presentation of a test trial. The trial started with two pictures appearing on the screen simultaneously in silence for 2s, providing children with enough time to process them without any sound. Then the test sentence was presented and lasted for 3.5s, followed by another 2s’ silence. All test trials were quasi-randomized and counterbalanced to avoid any potential patterns for children.

![Figure 1. Timeline of stimuli presentation during a test trial](image)

2.3. Procedure

Participants were tested individually in an acoustic booth. The child sat on their parent’s lap, facing an LCD screen. The parent listened to masking music from headphones, which was designed to prevent them from cuing the child. Loudspeakers on both sides of the screen played audio stimuli and a hidden camera below the screen recorded the child’s looking behavior for offline coding. The camera also sent simultaneous signals to a monitor next to the testing booth, where an experimenter, blind to all stimuli, observed the child and initiated each trial when the child looked at the center of the screen. After the test, the parent was asked to fill out a questionnaire on their children’s understanding of the words used in the test items.

2.4. Coding and Analysis

Children’s looking behavior in the test trials were coded offline by a blind researcher frame by frame at a rate of 50 frames/sec. Each frame was coded as left look, right look, center look or else where. Another researcher was asked to do a reliability check and the percentage agreement reached 92%. Trials that started with non-center looking were excluded for analysis, accounting for 5.8% of the total.
To assess whether children were sensitive to semantic feature agreement between classifiers and nouns, we chose three analysis windows: the silence window at the trial onset, the post-classifier window and the post-noun window, as illustrated in Figure 1. The 2s-silence window was included to examine if participants had any natural bias for a particular picture. It started from the trial onset to the onset of test stimuli (frames 0-100). The post-classifier window was chosen to see if children processed the classifier immediately after it was perceived and used the classifier feature to predict the upcoming noun. This was a very important window for us to examine children’s predictive processing. It started from 0.3s after the classifier onset and ended at the noun onset (frames 176-232). The post-noun window was also included to see how children responded when they heard the whole classifier-noun structure. It started from 0.3s after the noun onset and lasted 1s (frames 251-301).

For each window, we analyzed the proportion of looking to the noun-target (PLT). We assigned the noun picture labeled in the test stimulus as the target. We calculated the differential score (DS), which equals PLT in the post-classifier/post-noun window minus PLT in the silence window.

We made the following predictions. If children could represent classifiers and process classifier-noun agreement online, in the post-classifier window, DSs in match and mismatch trials should be different. Specifically, more looking to the target would be expected in match trials than in mismatch trials. In addition, in match trials, DSs should be above chance level, i.e. children should look more at the classifier-matching target. However, in mismatch trials, DSs should be below chance level, i.e. they should look at the classifier-agreeing distractor noun. In the post-noun window, while they would naturally look at the target picture named in both match and mismatch trials, their looking to the target would be impeded in mismatch trials as a result of their awareness of classifier-noun agreement. Predictions for individual classifiers follow the same pattern. If children understood the semantic content of a specific classifier and apply the feature in classifier-noun agreement, in both post-classifier and post-noun windows, DSs should be different in match and mismatch conditions. In the post classifier window, DSs should be above chance level in the match condition but below chance level in the mismatch condition.

3. Results

To test our predictions above, we conducted two statistical analyses: independent t-tests comparing DSs in each trial with chance level 0, and paired t-tests comparing DSs in match trials and mismatch trials. We did the two analyses for both aggregated and individual classifiers respectively.

Figure 2 depicts mean DSs of aggregated classifiers in the post-classifier window. As predicted, DS in match conditions was significantly above chance (M=.109, SE=.039, t (31)= 2.824, p=.008), which means that children predicted the classifier-matching target noun. In mismatch trials, DS was below the 0 chance level, though the difference was not significant (M=-.068, SE=.052,
This suggests that children also looked at the picture that matched the classifier before the noun was spoken. In addition, match trials differed significantly from mismatch trials \((p=.022)\), which provides even stronger evidence that children looked more at classifier-matching noun objects in both conditions.

\[
t(31)=-1.305, \quad p=.101
\]

Figure 2. Mean DS of aggregated classifiers in post-classifier window

Notes: *\(p < .05\). **\(p < .01\). 

Figure 3. Mean DSs of aggregated classifiers in post-noun window

Notes: *\(p < .05\). **\(p < .01\). ***\(p < .001\).

Apart from aggregated classifiers, we analyzed data of individual classifiers, hoping to see whether children’s performance was the same across classifiers. As Figure 4 shows, in zhang trials, DSs showed no significant difference from chance 0 in either match trials \((M=.092, SE=.091, t(31)=1.265, p=.216)\) and mismatch trials \((M=.081, SE=.091, t(31)=-2.133, p=.049)\). The results indicate that in mismatch trials, when they were supposed to look at the picture with the spoken name, their looking behavior was hindered due to the effect of classifier-noun agreement.
In tiao trials, DS was significantly above chance 0 in match trials (M=.265, SE=.097, t (26)= 2.738, p=.011), and below chance in mismatch trials (M=-.156, SE=.090, t (26)=1.740, p=.047). In kuai trials, in neither of the two conditions were the DSs significantly different from chance (Match: M=.033, SE=.103, t (24)= 0.317, p=.754; Mismatch: M= -.131, SE=.117, t (24)= -1.141, p=.265). In gen trials, DS was significantly above chance level in match trials (M=.157, SE=.087, t (27)= 1.804, p=.041), and below chance in mismatch trials (M= -.180, SE=.103, t (27)= -1.751, p=.046). Paired t-tests show that DSs were significantly higher in match conditions than in mismatch conditions in tiao (p=.005) and gen (p=.018) trials. Again, there was no noticeable difference in zhang (p=.645) or kuai trials (p=.246) when comparing DSs in match and mismatch trials. This indicates that children were able to identify the semantic features of tiao and gen, but not those of zhang and kuai.

As Figure 5 suggests, in the post noun window, no statistics indicate a significant difference between DSs and the chance level in zhang trials (Match: M=.129, SE=.079, t (23)= 1.622, p=.118; Mismatch: M=.079, SE=.085, t (23)= 0.923, p=.365). In tiao trials, DS was significantly above 0 chance level in match trials (M=.438, SE=.073, t (25)= 5.998, p=.000), and below the 0 chance level in mismatch trials (M=.224, SE=.091, t (25)=2.462, p=.021). In kuai trials, DSs bore no significant difference from chance in both conditions (Match: M=.122, SE=.090, t (31)= 1.353, p=.186; Mismatch: M=.028, SE=.073, t (31)= 0.383, p=.705). In gen trials, DS was significantly above 0 chance level in match trials (M=.342, SE=.092, t (23)= 3.709, p=.001), and marginally below the 0 chance level, in mismatch trials (M=.135, SE=.094, t (23)=1.443, p=.082). Paired t-test showed that DSs were significantly higher in match conditions than in mismatch conditions in tiao trials (p=.029) and the significance level was marginal in gen trials (p=.062). Again, there is no significant difference in zhang (p=.712) and kuai trials (p=.396). Taken together, these results show that children performed best with tiao, and worse with zhang and kuai.
4. Discussion and conclusions

In this paper, we asked whether 30-month-old Mandarin-speaking children are sensitive to classifier-noun agreement. We conducted an IPLP experiment to examine children’s looking behavior under two conditions, classifier-noun match and mismatch. In the experiment, we presented children with test stimuli involving shape classifiers *tiao*, *gen*, *zhang* and *kuai*. If they knew that classifiers agree with nouns in semantic features, they would use knowledge of classifier-noun agreement to guide their looking behavior as soon as they perceive classifiers. The experiment yielded several findings. First, results showed 30-month-old children’s representations of classifiers. This can be manifested by children’s looking preference to classifier-matching objects in the post-classifier window in both match and mismatch conditions. Second, children are sensitive to classifier-noun agreement. This is evidenced by their looking behavior in the post-noun window. Though children looked more to labeled object nouns in both match and mismatch conditions, their PLT was significantly lower in mismatch condition than that in match condition, which proved that children were aware of the fact that classifiers and nouns disagreed on semantic features in mismatch conditions. Third, results of individual classifiers suggest that children’s sensitivity of classifier-noun agreement varies across classifiers. This can be revealed by their differentiated performance in the four classifiers. They performed better in *tiao* and *gen* trials in both the post-classifier and post-noun windows, but not in *zhang* and *kuai* trials. These findings have important implications for our understanding of classifiers in early Mandarin grammar and use.

Our findings suggest that 30-month-old Mandarin-learning children have access to abstract representations of classifiers as a functional category. If they didn’t have any abstract representation of classifiers, they would probably
ignore classifiers and would not show any special preference to the target or the
distractor picture. Yet, what our experiment shows is that in the post classifier
window, they started searching for semantically-related noun objects once
classifiers were heard instead of waiting for semantically-related nouns. This
means that the occurrence of classifiers immediately activates the classifier and
its associated nouns. The findings are also consistent with Li et al (2010)’s
findings that 2-year-olds demonstrate an early use of classifiers. Our findings
provide further evidence for early availability of functional categories and for
the continuity of functional categories.

Our findings are also indicative of 2-year-old children’s sensitivity to shape
feature agreement between classifiers and nouns. Like their French-learning
counterparts (Melançon & Shi 2015), they demonstrate abstract representations
of feature agreement between functional category and lexical category. These
representations cannot result from the rote memory of particular classifier-noun
chunks, because in our experiment, we inserted an adjective modifier *caisede
‘colorful’, which blocked the collocation of classifiers and nouns. What’s more,
children started to look at classifier-matching object once classifier occurred
which also indicates that they do not treat classifier-noun as an inseparable pair.

Third, young Mandarin-speaking children also process classifiers
incrementally and predicatively during online comprehension. They
immediately parse and interpret the classifier when they hear it rather than
waiting to analyze it until the end of sentence. They also use classifiers to
predict the semantic information of the noun before it is presented. The
prediction effect is instant and online.

Fourth, for specific shape classifiers, children perform better in *tiao* and *gen
trials than in *zhang* and *kuai* trials. The results echo previous findings in
classifier acquisition (Fang 1985; Wen 2013; Li et al. 2010; Ying et al. 1983)
that children’s understanding of classifiers does not come all at once. They
understand some particular classifiers earlier. Presumably, features inherent in
*tiao* and *gen* are one-dimensional and are easier for children to acquire than
multi-dimensional features associated with *zhang* and *kuai* (Huang 2009).

In conclusion, 30-month-old Mandarin-speaking children are sensitive to
classifier-noun agreement, indicating an early access to abstract representations
of classifiers. They process classifiers and nouns incrementally and predictively
during online comprehension. Yet, their understanding of the semantic content
of specific classifiers differs.

**References**

Altmann, Gerry T.M. & Kamide, Yuki. (1999). Incremental interpretation at verbs:
Restricting the domain of subsequent reference. *Cognition, 73, 247–264.*

Altmann, Gerry T.M. & Kamide, Yuki. (2007). The real-time mediation of visual
attention by language and world knowledge: Linking anticipatory (and other) eye
movements to linguistic processing. *Journal of Memory and Language, 57,
502–518.*


