Che-che ('car-car') and *chi-chi* ('eat-eat'): Reduplication in Mandarin Chinese child-directed speech

Mengru Han (mrhan@zhwx.ecnu.edu.cn)
Department of Chinese Language and Literature, East China Normal University,
500 Dongchuan Rd. Shanghai, 200241 China

Yan Gu (yan.gu@ucl.ac.uk)
Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK, Department of
Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Abstract

Recent studies have demonstrated that reduplication facilitates children's word segmentation and learning. Although reduplication is often considered a feature of child-directed speech (CDS), it remains unclear if it is indeed more frequent in CDS than in adult-directed speech (ADS). This study examines the production and perception of reduplication in the context of language acquisition by focusing on Chinese, which has a rich system of reduplication. We analyze the frequency of reduplicated types and tokens as a function of speech register (ADS/CDS) and children's age (18m/24m) in a corpus. Additionally, we conduct a survey to examine adults' perception of reduplications and determine their degree of child-directedness. Results indicate that there are more reduplicated types in CDS than in ADS. However, only the reduplicated tokens that are rated to be child-directed-specific occur more in CDS than in ADS. These findings provide insight into the nature of lexical input in language acquisition.

Keywords: Reduplication; child-directed speech; language input; Mandarin Chinese

Introduction

The mascot of the Winter Olympic Games Beijing 2022, *Bing Dwen Dwen*, is a cute little panda wearing a full-body shell made of ice (see a picture in International Olympic Committee, n.d.[1]). The name is derived from Mandarin Chinese (hereafter 'Chinese'), with *Bing* meaning 'ice' and *Dwen Dwen* being a reduplication of the morpheme *Dun* ('a bloke of stone or wood'), symbolizing being chubby and powerful and sometimes representing children. This is one of the few Olympic mascots that is named with a reduplicated form, likely due to the prevalence of reduplication in the Chinese language: reduplication has been widely used in classical Chinese poetry since the 11th century BCE (Jin, 2018). Expressing gratitude in Chinese is often done by saying the reduplicated form *xièxie* ('Thanks'). Approximately 10% of Chinese given names are reduplicated (van de Weijer et al., 2020), for example Yo-Yo Ma (cellist) and Youyou Tu (Nobel Prize winner). Also, reduplication is widely used between caregivers and children, as well as between intimate friends or lovers. Nevertheless, our understanding of Chinese reduplication is rather limited. Specifically, psycholinguistic studies on Chinese reduplication are scarce, especially in the context of child-directed language. Most studies of this area have been conducted on European languages, and there has been hardly any research in Chinese, a language with a complex system of reduplication. This research aims to analyze the production and perception of reduplication in Chinese child-directed speech (CDS), as well as the connection between them. Additionally, we explore the role of reduplication in children's lexical development.

Reduplication in Language, Cognition, and Acquisition

Reduplication is a morphological process in which meaning is expressed by repeating an entire word (full reduplication; e.g., Malay: *buku-buku* 'books') and *night-night* (English: 'good night')) or part of a word (partial reduplication; e.g., *huáliūliū* (Chinese: 'slippery'). Reduplication has been widely documented in languages around the world (Rubino, 2005). Since the 1960s, language acquisition researchers have taken an interest in this linguistic phenomenon. On the one hand, children's early language production is characterized by reduplication: Early babbling typically consists of a repeated sequence of CV syllables (Smith et al., 1989). Also, children's first words tend to appear in a reduplicated form rather than the standard form (Dressler et al., 2005; Ferguson, 1983). On the other hand, reduplication is often considered a feature of child-directed speech (CDS), a speaking style that caregivers use when addressing young children. Child-directed specific (CD-specific) words are often characterized by reduplication in many languages (Ferguson, 1964; Mazuka et al., 2008). For example, English has a set of reduplicated forms of common nouns specific to CDS, such as *choo-choo*, *night-night*, and *wee-wee*. Japanese mothers use a large number of reduplicated forms instead of the regular adult forms when addressing children, for example, *gohan* ('food') becomes *maNma* in CDS (Mazuka et al., 2008). In Chinese, there is also a set of words that are usually reduplicated in CDS but not ADS, such as *chēchē* ('car-car') and *chīchī* ('eat-eat'). Such evidence has also been found in a diversity of understudied languages such as Marathi, Comanche, and Gilyak, etc. (Ferguson, 1964).

Recent studies conducted by Ota, Skarabela, and their colleagues have demonstrated the significance of

---

[1] https://olympics.com/en/olympic-games/beijing-2022/mascot

reduplication in early language acquisition. For example, Ota and Skarabela (2018) found that nine-month-old English-learning children are better at segmenting reduplicated words than non-reduplicated ones. Ota and Skarabela (2016) revealed that eighteen-month-old English-learning children are more successful in learning novel word-object pairings when the words are reduplicated. Also, Ota et al. (2018) observed that the amount of reduplicated words in the input is positively associated with children's vocabulary growth from nine to twenty-one months.

The robust facilitative effect of reduplication in language acquisition is not surprising considering its form and meaning in language and cognition. First, in a broad sense, humans tend to be more adept at learning patterns that involve repetition across different domains (Endress et al., 2007). Compared to random sequences (ABC; e.g., *mubage*), newborns are more sensitive to reduplicated sequences (ABB; e.g., *mubaba*) (Gervain et al., 2008). Reduplicated forms, such as audio syllables, letters, and musical notes, are easier for adults to learn (Endress et al., 2007; Marcus et al., 1999). Evidence also shows that reduplicated forms facilitate artificial grammar learning (Gomez et al., 2000). Second, some reduplicated forms are cognitively related to iconicity across spoken and sign languages (see a review in Perniss et al., 2010). The semantic function of reduplication can be iconic, indicating onomatopoeia, iteration, augmentation/ intensification, repetition, continuation, among others (Fischer, 2011). For example, the English onomatopoeia *choo-choo* has an association between the spoken word and the sound of a steam locomotive. Research has shown that humans are sensitive to iconicity in language, and that iconicity in both spoken and gestural communication is beneficial to language learning. For instance, studies by Laing (2017) and Motamedi et al. (2021) have demonstrated that onomatopoeia can help children to learn words. Other aspects of iconicity, such as sound symbolism and iconic gestures, can also assist children's word learning (Imai & Kita, 2014; Aussems & Kita, 2021). It should be noted that reduplicated forms are not necessarily iconic. A reduplicated form can also indicate diminution or affection, which is argued to be related to the frequent use of reduplication in CDS (Fischer, 2011).

Reduplication is commonly regarded as a significant feature of CDS across languages. CDS is a primary input source in children's first years of language development (Ramírez-Esparza et al., 2016). To examine the special role of CDS in language acquisition, researchers often compare CDS with adult-directed speech (ADS) on different linguistic dimensions. This line of research suggests that CDS typically has exaggerated prosody, more diminutives, hyperarticulated vowels, shorter utterances, more repetitions, more gestures, and is grammatically less complex than ADS (e.g., Campisi & Özyürek, 2013; Tippenhauer et al., 2020; see reviews in Golinkoff et al., 2015; Soderstrom, 2007;). However, to date, no study has been conducted to compare the use of reduplication between ADS and CDS in a systematic way, particularly in a language like Chinese where reduplication is more common than in most Indo-European languages (Rubino, 2005) and many reduplication forms are not used for a child-directed purpose. In addition, previous studies have often identified CD-specific reduplicated forms through observation of speech production (e.g., Skarabela et al., 2015). It is yet unknown which reduplicated forms are *perceived* to be CD-specific by native speakers.

From the typological point of view, the productivity of reduplication varies among languages. European languages use reduplication in a rather restricted way, often in onomatopoetic words and expressive/figurative speech (Hurch & Mattes, 2009; Nadarajan, 2006). In these languages, reduplication is usually found in CD-specific words but are not common in ADS. In some languages, however, reduplication has a broader range of use. For example, in Chinese, reduplication is widely used for phonological, morphological, syntactic, and pragmatic purposes (Wang, 2012). A quantitative study on Chinese corpus and dictionary found that a variety of word classes can be reduplicated, including many basic forms of nouns, verbs, adjectives, and onomatopoeias, as well as numbers, classifiers, and pronouns, suggesting that reduplication is a productive morphological method in Chinese (Xing, 2020). Interestingly, about 4% Chinese male and 17% female given names are in reduplicated forms (van de Weijer et al., 2020). Thus, it is not surprising that the Beijing 2022 Olympic Mascot *Bing Dwen Dwen* has a reduplication in its name. However, what remains unclear is whether there is a correlation between the high prevalence of Chinese reduplication in input and the language learning of children.

Age Effect and Individual Speaking Style of CDS

One factor associated with the quantity and quality of CDS is children's age. As children age and develop more advanced language skills, there is a tendency for CDS to become more alike to ADS. For example, the prosody of CDS is characterized by an exaggerated prosody, including a higher pitch and a slower speaking rate (Soderstrom, 2007). Studies have shown that the pitch height of CDS reduces and becomes more similar to ADS as children grow older (Liu et al., 2009; Han et al., 2020). The speaking rate of Swedish CDS increases from 7 to 33 months (Sjons & Hörberg, 2016). As such, if reduplication is adapted to children's age and language development, we would expect that as children grow older, the difference in the frequency of reduplication will decrease, in accordance with the tendency of other adaptive features of CDS. Alternatively, if reduplication serves additional functions such as expressing affect, it may not change significantly within a small age range. In addition, if caregivers' use of reduplication in ADS and CDS is merely a reflection of their speaking style, we can expect that those caregivers who have more reduplication in ADS will likely have more reduplication in CDS, independent of the children's age.

The Current Study

The present research aims to investigate Chinese reduplication in the context of language acquisition, focusing on four aspects: (1) the reduplication in Chinese CDS and ADS; (2) the native speakers' perception of reduplication, and the relationship between the production and perception of reduplication; (3) the effect of age and individual speaking style on mothers' use of reduplication; and (4) the correlations between reduplication CDS and children's language outcomes.

To address the first three aspects, we study whether the distribution of reduplication differ as a function of speech register (ADS/CDS) and children's age (18/24 months), and how reduplication is perceived by native speakers. It has been assumed that reduplication is a unique characteristic of CDS and, thus, should be more common in CDS than ADS. However, no study has empirically tested this assumption, especially in Chinese, a language that makes an intensive use of reduplication. It is possible that, even though reduplication is normal in ADS, Chinese mothers nevertheless use more reduplication when talking to young children. Alternatively, it is also quite plausible that the overall frequency of reduplication does not differ between ADS and CDS, as it is already prevalent in ADS and the majority of reduplications are shared both in ADS and CDS. In that case, only the CD-specific reduplicated forms are expected to be more frequent in CDS than ADS.

For the age effect, we hypothesize that mothers will produce more reduplicated tokens and types in CDS when addressing 18-month-old children in comparison to 24-month-old children, which is in accordance with the results on mean pitch and speaking rate in previous studies (e.g., Liu et al., 2009; Sjons & Hörberg, 2016). Additionally, we explore the role of individual speaking style on the use of reduplication, and anticipate that mothers who use more reduplications in ADS will also use more reduplications in CDS.

Furthermore, based on findings from the corpus study, we aim to identify which Chinese reduplicated types are perceived to be CD-specific. As reduplication is prevalently used in Chinese language, we expect that the reduplicated forms that are perceived to be CD-specific may occur more in CDS than in ADS, however, the frequency of non-child-directed (NCD) specific reduplications will not differ between ADS and CDS.

As for the role of reduplication in language development, previous evidence suggests that reduplication facilitates word learning and the frequency of reduplication in CDS predicts children's vocabulary growth (Ota & Skarabela, 2016; 2018; Ota et al., 2018). However, as reduplication is a common feature of ADS and is used for various purposes in Chinese, we expect that the frequency of reduplicated forms specific to CDS, but not others, will predict children's language outcomes.

To address these aspects, we conducted an analysis of the frequency of reduplicated forms in a corpus of ADS and CDS, as well as a survey of native Chinese speakers' perceptions of these forms. These analyses will contribute to our understanding of the role of reduplication in language acquisition and cognition.

Method

Corpus Analysis of Reduplication in Chinese CDS and ADS

Corpus and Participants We analyzed the frequency of reduplicated forms in a corpus of semi-spontaneous ADS and CDS. This corpus is part of a larger cross-linguistic corpus of Dutch and Chinese ADS and CDS (Han, 2019).

Participants were Mandarin-Chinese-speaking mother-child dyads ($N = 40$) who were divided in two groups by the age of the child participant. There were twenty-one 18-month-old children (girls $N = 9$; mean age = 18;15 [months;days], age range = 17;21–18;27) and nineteen 24-month-old children (girls $N = 10$; mean age = 24;13, age range = 23;27–24;30). All children were typically developing and had no language impairment or hearing problems.

We created a storybook to elicit ADS and CDS. The book is 12 pages long. On each page of the picture book, a word is shown on the left side and an illustration including a depiction of the word is on the right side. No other script is provided beside the words. In this way, mothers are free to construct the story, and the only requirement is to include the words given on each page. Each participant mother told the story twice, namely in ADS and CDS. To elicit CDS, the mother was instructed to tell her child the story as they normally would at home. To elicit ADS, mothers told the same story to an adult (female, a Mandarin native speaker), while the child was not present. The order of the two speech registers was counterbalanced across participants. As mothers told the same story which provided no script in ADS and CDS, we may compare mothers' natural use of language, in this case reduplications, when the content was similar in the two speech registers.

Speech Data Analysis We transcribed mothers' speech data using an automatic speech recognition tool for Chinese developed by iflytek (https://www.iflyrec.com). As Chinese is written without spaces between successive characters and words, we used JiebaR (Qin & Wu, 2019), a Chinese text segmentation tool in the R environment (R Core Team, 2020) to segment the transcribed scripts into words. Both the transcription and segmentation were manually checked by two Chinese native speakers. The corpus consists of 25,280 word tokens (ADS: 8,739 word tokens; CDS: 16,541 word tokens). A native speaker of Chinese manually marked the reduplications in the transcription. Here, we included all reduplicated types such as full reduplication (e.g., *huahua* ('flower-flower')) and partial reduplication (e.g., *shui jiaojiao* ('sleep'). For each participant and each speech register, the number of unique reduplicated types and the number of times these reduplicated types occurred (reduplicated tokens) were calculated. The proportion of

reduplicated tokens was calculated by dividing the number of reduplicated tokens per speech register by the total number of tokens in this speech register per participant.

Vocabulary Test After each recording session, mothers filled in the Mandarin Chinese version of MacArthur-Bates Communicative Development Inventories (CDI; (Fenson et al., 2007); M-CDI (Tardif et al., 2009). The expressive vocabulary size was calculated for each participant and was used as a measure of children's language outcomes.

Survey

Based on the reduplicated forms found in the corpus, we conducted an online survey in Chinese native speakers to examine their perception of reduplication forms as to the degree of child-directness.

Participants The participants were 134 native Chinese adults (Mean age: 36.3 years (SD = 13.9); Female: N = 94). Five participants were eliminated from the analysis due to incomplete questionnaires, identical ratings across the entire questionnaire, or a failure to answer two test questions correctly.

Procedure We created an anonymous questionnaire on an online platform Tencent Survey. In the survey, we asked a question for each reduplicated type: is this reduplicated form more likely to be addressed to children or adults? Participants were instructed to rate, at a 5 point-scale, as to whether each reduplicated type found in the corpus were more likely to be addressed to children or adults: 1 = to children only; 2 = mostly to children; 3 = both children and adults; 4 = mostly to adults; 5 = to adults only). We divided the 158 reduplicated types in the corpus into 16 blocks, with ten reduplicated types in each block. In this way, we get a score of child-directedness for each reduplicated type. The order of the reduplicated types was randomized within each block. There were two filler questions after the tenth and the sixteenth block. In addition, we collected participants' demographic information such as age, education, and gender. The average time to finish the survey was 20.6 minutes. All participants received a financial compensation ranging from 1.05 CNY to 8.94 CNY (randomly assigned via a lottery).

Data Processing and Analyses Based on the survey, we calculated the mean child-directedness score (CD-score) for each reduplicated type across raters. For all reduplicated types, our survey yielded a mean CD-score of 2.41 (SD = 0.47, range: 1.4-3.4). Based on the CD-scores, we established a criterion that a CD-specific type has a mean CD-score of no more than 1.7 (half of the maximum score). In this way, the reduplicated types were further divided into 18 child-directed-specific (CD-specific) reduplicated types and 141 non-child-directed-specific (NCD-specific) ones.

To better understand the link between the production and perception of reduplicated types, we examined the relationship between the number of reduplicated types and the CD-scores obtained from the survey. For each participant, we calculated the number of reduplicated types and the mean CD-scores of these reduplicated types per Speech Register. Then we examined whether the number of reduplicated types can be predicted by the CD-scores, Age, Speech register. In addition, we built models for the proportion of CD-specific reduplicated tokens and the proportion of Non-CD-specific (NCD-specific) reduplicated tokens, respectively.

We used R (R Core Team, 2020) and lme4 (Bates, Maechler & Bolker, 2014) for the analysis of distribution of replication. First, we fit a linear mixed-effects model to examine the effect of Speech Register (ADS/CDS), Age (18m/24m), as well as their interaction (independent variables) on the number of overall or CD-specific reduplicated types (dependent variables). We used AIC and the 'anova' function to compare model fitness. The effect of Age and the interaction between Age and Speech Register were dropped when they did not significantly improve the model. When the dependent variable was binary (reduplicated token frequency), we used a generalized linear mixed model (GLMM) to compare the differences in proportion.

In addition, Spearman's rank correlation was used to assess the relationship between the number of reduplicated types in ADS and CDS. Liner models were used to analyze the relationship between children's language outcomes (dependent variables) and caregivers' proportion of reduplicated tokens, the number of reduplicated types, and the number of CD-specific reduplicated types in CDS (independent variables).

Results

In total, we identified 1061 reduplicated tokens in the corpus, which were classified into 158 reduplicated types. **Table 1** presents the total number of utterances with reduplication, as well as the means (and SDs) of the numbers and proportions of reduplicated tokens and reduplicated types. Interestingly, regardless of speech register, approximately 20% of utterances in the corpus contained at least one reduplicated form, demonstrating its common use in Chinese language.

Table 1: The total number of utterances with reduplication and the means (and SDs) of the numbers (N) and proportions (%) of reduplicated (RED) types and tokens.

Measures			18 months		24 months	
			ADS	CDS	ADS	CDS
Utterances with RED		N	158	415	166	312
		%	21.18	20.85	21.96	18.92
RED type	All	N	4.81 (4.13)	9.29 (4.87)	5.89 (3.93)	8.37 (4.75)
	CD	N	0.38 (0.81)	1.05 (1.02)	0.32 (0.58)	1.11 (1.29)
RED token	CD	N	0.62 (1.60)	1.76 (2.02)	0.32 (0.58)	1.42 (1.87)
		%	0.03 (0.08)	0.05 (0.06)	0.01 (0.03)	0.03 (0.03)

NCD	N	6.90 (5.50)	17.9 (17.9)	8.42 (5.70)	15.6 (11.3)
	%	3.54 (1.59)	3.91 (1.44)	3.44 (1.45)	3.50 (1.70)

Number of Reduplicated Types

The results showed that there was a significant effect of Speech Register on the number of reduplicated types (β = 3.40, SE = 0.66, t = 5.11, $p < 0.001$). The results held when we focused on the CD-specific reduplicated types only: mothers used significantly more CD-specific reduplicated types in CDS than in ADS (β = 0.58, SE = 0.15, t = 3.73, $p < 0.001$). Together, these results indicate that mothers used significantly more reduplicated types as well as more CD-specific reduplicated types in CDS. As there was no significant effect of Age, these results held for both ages.

Table 2. Summary of model for the number of reduplicated types predicted by the CD-scores

Parameters	Estimate	SE	t-value	p
Fixed factors				
(Intercept)	21.79	6.98	3.12	0.003**
Age (24m)	-19.21	9.52	-2.02	0.048*
CD-score	-6.45	2.68	-2.41	0.019*
Speech Register (CDS)	3.76	0.69	5.43	<0.001***
Age(24m):Mean CD-score	-0.15	0.09	-1.61	0.048*

*Note. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*

Furthermore, we examined the relationship between the number of reduplicated types and the CD-scores. As summarized in **Table 2**, there was a significant effect of Speech Register and a significant interaction between Age and the Mean CD-scores (**Figure 1**), indicating that a smaller CD-scores (indicating more CD-specific) predicted a larger number of reduplicated types at 18 months (p = 0.022), but there was no such an effect for 24 months (p = 0.686).

Proportions of Reduplicated Tokens

Subsequently, we analyzed the proportion of CD-specific reduplicated tokens and the proportion of NCD-specific reduplicated tokens, respectively, as a function of Age and Speech Register. The final model for the proportion of CD-specific reduplicated tokens revealed that there was a significant main effect of Speech Register (β = 0.90, SE = 0.32, t = 2.81, p = 0.005), indicating that the proportion of CD-reduplicated tokens was significant larger in CDS. The Age effect was not significant and was not included in the final model. By contrast, the model for the NCD-specific reduplicated tokens showed that neither Speech Register nor Age was significant (all p's > .05), and only Age was included in the final model (β = -0.15, SE = 0.09, t = -0.61, p = 0.11). These findings suggest that the proportion of CD-specific reduplicated tokens was larger in CDS compared to ADS, but the NCD-specific ones did not show the same pattern.
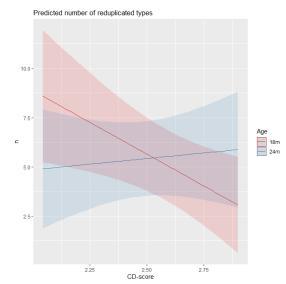

Figure 1: Predicted values of the number of reduplicated types as a function of the CD-score and children's age.

The Role of Speaking Style

Spearman's rank correlation test showed a positive correlation between the number of reduplicated types in ADS and CDS (r = 0.47, p = 0.002). A similar pattern was observed for the number of CD-specific reduplicated types (r = 0.31, p = 0.053). These findings suggest that mothers who use more reduplicated types and more CD-specific reduplicated types in ADS also tend to use more reduplication in CDS, indicating that the use of reduplication in CDS can reflect an individual speaking style.

Relating Reduplication in CDS to Language Outcomes

We further explored whether there was a correlation between mothers' use of reduplication and children's vocabulary size, controlling for children's age. There were no correlations between any of the predictors (the proportion of reduplicated tokens, the number of reduplicated types, and the number of CD-specific reduplicated types in CDS) and children's expressive vocabulary (all p's > 0.3).

Discussion

The present study conducted a corpus analysis to investigate the distribution of reduplication, tested native speakers' perception of reduplication, and examined whether caregivers' reduplication related to children's language outcomes. Firstly, we compared the frequency of reduplication in a semi-spontaneous speech corpus in which the content of ADS and CDS was highly similar. Our results indicated that mothers used more reduplicated types and more CD-specific reduplicated tokens in CDS than in ADS, while the NCD-specific ones did not differ between the two speech registers. This suggests that in Chinese, where CDS is

very common in ADS, mothers modify their lexical choices in CDS to better suit the addressee status. Given the evidence that reduplication can facilitate children's word segmentation and learning, as well as infants are sensitive to the repeated patterns in language (e.g., Gervain et al., 2008; Ota & Skarabela, 2016, 2018), our findings suggest that by using more reduplicated forms in CDS, mothers modify their lexical input to best serve children's language learning. It is possible that modifying the lexicon could be beneficial for language acquisition.

Previous studies have indicated a positive correlation between reduplication in CDS and children's vocabulary growth from 9 to 21 months (Ota et al., 2018); however, our research did not yield a statistically significant correlation. This may be because of the cross-sectional nature of our research and the fact that we focused on older age groups (18 and 24 months), thus making the correlational effect less pronounced. It is also possible that in a language in which reduplication is prevalent, the facilitative effect of reduplication is less robust compared to other languages. Our findings demonstrate a significant role of individual speaking style: mothers who use more reduplication in ADS also tend to use more reduplication in CDS. We intend to further explore how these individual differences in ADS and CDS affect children's language development.

Studies that quantify the proportion of reduplication in CDS in different languages are scarce. Ota et al. (2018) found that the proportion of reduplicated tokens in English CDS is approximately 2%. In our corpus, the overall proportions of reduplicated tokens in Chinese ADS and CDS are around 2% and 4%, respectively. This suggests that the proportion of reduplication in Chinese is almost double that of English CDS. At the utterance level, the difference is more striking: nearly one fifth of the utterances in Chinese CDS had a reduplicated form. These quantitative and qualitative differences in the lexical input between the two languages may lead to differences in children's language outcomes in typologically distinct languages. To better understand the nature of lexical input across languages, further research on typologically distinct languages is necessary.

Second, instead of identifying CD-specific words by observation, the current study used a survey to capture the CD-specific reduplicated forms in Chinese. Among the 158 reduplicated types found in the corpus, 18 were rated to be CD-specific by native speakers. These CD-specific words included various word categories such as nouns, verbs, and adjectives. It should be noted that this list of Chinese CD-specific reduplicated words is not exhaustive: this corpus was limited to a semi-spontaneous storybook telling task, and CD-specific words are not all reduplicated. More research is needed to identify the CD-specific words in Chinese as well as other languages. So far, there have been some cross-linguistic investigations of the acoustic and morphology of CD-specific words (e.g., Skarabela et al., 2015). However, these studies often identified CD-specific words by observation. We strongly suggest that a similar survey be conducted to evaluate the degree of child-directedness.

Third, we found an age effect in relation to the number of reduplicated types. Specifically, for the distribution of CD-specific reduplicated types, in comparison to ADS, mothers only used more CD-specific reduplicated types when addressing 18-month-old children, but not 24-month-old children. These findings suggest that mothers fine-tune their lexical choice according to children's ages.

This study used a storytelling task to elicit both CDS and ADS samples. While one reviewer suggested that the ADS sample may have resembled CDS due to the child-oriented content, the study still revealed significant differences in reduplication between CDS and ADS. Additionally, the prosodic analysis of this corpus revealed that mothers indeed modify their speech prosody in CDS as compared to ADS (Han et al., 2020; 2021). One advantage of this method is that it allowed for a comparison of lexical choices in speech when the content was similar. However, it is important to note that CDS can vary across different speech contexts, and future studies should consider a broader range of contexts to gain a more comprehensive understanding of CDS reduplication.

Reduplication is related to iconicity. Many of the reduplicated forms found in the corpus can be related to iconicity to some degree. In Chinese, reduplication of nouns can mean diminutively. However, only a subset of these forms is related to be iconic. It should be noted that not all the tokens that were rated as the most CD-specific are iconic. Future study can investigate the role of iconic-specific reduplication in Chinese language acquisition. Interestingly, several CD-specific forms are personification of common nouns, for example *yueliangpopo* ('Grandma Moon'), *yueliang gonggong* ('Grandpa Moon'), and *taiyang gonggong* ('Grandpa Sun'). The function of personification reduplication in is not yet clear.

Conclusion

This is the first study that systematically compared the use of reduplication in ADS and CDS, and assessed the child-directedness of reduplication from the perspective of native speakers' perception. By connecting production and perception of reduplication, this study revealed both quantitative and qualitative distinctions between ADS and CDS. We had three main findings: First, only about 10% of the reduplicated types in Chinese (in this corpus) are specific to CDS, while the rest are perceived as non-child-directed by native speakers. Second, there are generally more reduplicated types and more CD-specific reduplicated types in CDS compared to ADS, however, only the proportion of CD-specific reduplicated tokens, not the proportion of the NCD-specific ones, is greater in CDS compared to ADS. Third, mothers' use of reduplication in ADS and CDS are correlated, indicating a role of individual speaking style. These findings expand our knowledge of reduplication in the context of language acquisition, particularly in Chinese, an under-researched language in this field of research.

References

Aussems, S., & Kita, S. (2021). Seeing iconic gesture promotes first-and second-order verb generalization in preschoolers. *Child Development*, *92*(1), 124-141.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Dressler, W. U., Dziubalska-Kołaczyk, K., Gagarina, N., & Kilani-Schoch, M. (2005). *Reduplication in child language* (No. 28, p. 455). Walter de Gruyter.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*, *105*(3), 577–614. https://doi.org/10.1016/j.cognition.2006.12.014

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., & Reznick, J. S. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Ferguson, C. A. (1964). Baby Talk in Six Languages. *American Anthropologist*, *66*(6_PART2), 103–114. https://doi.org/10.1525/aa.1964.66.suppl_3.02a00060

Ferguson, C. A. (1983). Reduplication in child phonology. *Journal of Child Language*, *10*(1), 239–243.

Fischer, O. (2011). Cognitive iconic grounding of reduplication in language. *Semblance and Signification*, *10*, 55–81.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222–14227. https://doi.org/10.1073/pnas.0806530105

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339-344.

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition*, *28*(2), 253–263. https://doi.org/10.3758/BF03213804

Han, M. (2019) *The role of prosodic input in word learning: A cross-linguistic investigation of Dutch and Mandarin Chinese infant-directed speech*. Amsterdam: LOT Publications.

Han, M., de Jong, N. H., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word-learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language, 47*(1), 85–111.

Han, M., de Jong, N., & Kager, R. (2021). Language specificity of infant-directed speech: Speaking rate and word position in word-learning Contexts. *Language Learning and Development*, *17*(3), 221–240.

Hurch, Bernhard & Veronika Mattes. 2009. Typology of reduplication: The Graz database. In Martin Everaert, Simon Musgrave & Alexis Dimitriadis (eds.), *The use of databases in cross-linguistic studies*, 301–328. Berlin: Mouton de Gruyter.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130298. https://doi.org/10.1098/rstb.2013.0298

Jin, B. (2018). *Reduplication in Classical Chinese Poetry* [Unpublished Doctoral Dissertation], Leiden University.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, *161*, 32-45.

Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language*, *36*(4), 909-922.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science (New York, N.Y.)*, *283*(5398), 77–80.

Mazuka, R., Kondo, T. & Hayashi, A. (2008). Japanese mothers' use of specialised vocabulary in infant-directed speech: Infant-directed vocabulary in Japanese. In N. Masataka (Ed.) *The Origins of Language: Unravelling Evolutionary Forces*, Tokyo: Springer.

Motamedi, Y., Murgiano, M., Perniss, P., Wonnacott, E., Marshall, C., Goldin-Meadow, S., & Vigliocco, G. (2021). Linking language to sensory experience: Onomatopoeia in early language development. *Developmental Science*, *24*(3), e13066.

Nadarajan, S. (2006). A crosslinguistic study of reduplication. *Journal of Second Language Acquisition and Teaching*, *13*, 39–53.

Ota, M., & Skarabela, B. (2016). Reduplicated words are easier to learn. *Language Learning and Development*, *12*(4), 380-397.

Ota, M., & Skarabela, B. (2018). Reduplication facilitates early word segmentation. *Journal of Child Language*, *45*(1), 204-218.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science*, *42*(6), 1974-1999.

Perniss, P., Thompson, R., & Vigliocco, G. (2010). Iconicity as a General Property of Language: Evidence from Spoken and Signed Languages. *Frontiers in Psychology*, *1*. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00227

Qin, W., & Wu, Y. (2019). *JiebaR: Chinese Text Segmentation*.https://CRAN.R-project.org/package=jiebaR

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interactions on later language development in Spanish–English bilingual infants. *Child Development*, *88*(4), 1216-1234.

Rubino, C. (2005). Reduplication: Form, function and distribution. *Studies on Reduplication*, *28*, 11–29.

Sjons, J., & Hörberg, T. (2016). Articulation rate in child-directed speech increases as a function of child age. In *Fonetik 2016, KTH Royal Institute of Technology, Stockholm, Sweden, 8-10 June, 2016*.

Skarabela, B., Ota, M., Fazekas, J., & Wihlborg, L. (2015). Do baby-talk words reflect biomechanical constraints on speech production? In The Scottish Consortium of ICPhD (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Skarabela, B., Ota, M., Fazekas, J., & Wihlborg, L. (2015). Do baby-talk words reflect biomechanical constraints on speech production? In The Scottish Consortium of ICPhD (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow.

Smith, B. L., Brown-Sweeney, S., & Stoel-Gammon, C. (1989). A quantitative analysis of reduplicated and variegated babbling. *First Language*, *9*(6), 175–189. https://doi.org/10.1177/014272378900900605

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501-532.

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*(5), 1115–1144.

Tippenhauer, N., Fourakis, E. R., Watson, D. G., & Lew-Williams, C. (2020). The scope of audience design in child-directed speech: Parents' tailoring of word lengths for adult versus child listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2163.

Van de Weijer, J., Ren, G., van de Weijer, J., Wei, W., & Wang, Y. (2020). Gender identification in Chinese names. *Lingua*, *234*, 102759.

Wang, F. (2012). *A typological study on the multi-functionality of reduplication* [重叠多功能模式的类型学研究] [Unpublished doctoral dissertation], Nankai University, China.

Xing, H. (2000). Quantitative analyses of the structure of Chinese reduplication. [汉语词语重叠结构统计分析]. *Language Teaching and Linguistic Studies,* (1), 32-37.