Reduplication as a trigger of intersubjectivity: Mandarin Chinese ideophones and reduplication in the CHILDES corpora

Thomas Van Hoey
National Taiwan University

Chiarung Lu
National Taiwan University

ICLC-15
15th International Cognitive Linguistics Conference
6-11 August 2019,
Kwansei Gakuin University, Nishinomiya,
Roadmap

Introduction
language acquisition and ideophones

Research question
how are ideophones and related constructions acquired in Chinese?

Material
CHILDES — childesdb

Quantitative results

Qualitative case studies
sun/moon; apple

Discussion
intersubjectivity

Conclusions
Introduction
Usage-based approach to acquisition

In Cognitive Linguistic approaches acquisition is mostly agreed as being usage-based and bottom-up.

Two big frameworks

- **Usage-Based theory** of language acquisition (e.g. Tomasello's e.g. 1992; 2003)
- **Emergentist approach** (e.g. MacWhinney & O'Grady 2015)
Chunking

There are competing form-meaning mappings (constructions), with chunks of language first learnt as a whole, only to be later analyzed in more discrete ‘words’.

I mean I can remember when I was very young, much + young + er, and I applied for a job they said, well, are + n’t + you planning to have children? Well, I mean, that’s none of + their + business.

20 choices, 35 words, 25 words in prefabs
(Bybee 2010:60)
Frequency, salience, prototypicality

Important factors include frequency, but also salience and prototypicality.

• Skewed frequency of input facilicitates learning (Goldberg & Casenhiser 2008).

• Salience and prototypicality (cf. Geeraerts 2000; 2017) also prove an important factor.

Apple (pingguo 苹果)
- relatively high token frequency
- relatively early and high conceptual frequency
- relatively early and high referential frequency
- prototypical structure (fruit vs. e.g. ‘apple of my eye’)
- easily identifiable shape and colour
Nouns first? Or verbs first?

Across many languages, nouns appear to be learnt earlier than verbs (Gentner 1982; Gentner & Boroditsky 2001; Tomasello 2003; Imai et al. 2008; Waxman et al. 2013).

But for verbs there are cross-linguistic differences. In ‘verb-friendly’ languages such as Chinese (Tardif 1996; Tardif 2006), Korean (Choi & Gopnik 1995; Kim, McGregor & Thompson 2000), and arguably Japanese (Ogura et al. 2006; Imai et al. 2008) nouns are often dropped, and verbs — relational items — get a somewhat privileged status.
Acquiring ideophones / mimetics

Imai & Kita (2014) have argued for a lexical bootstrapping hypothesis, which shows that 0;11 Japanese infants are already sensitive for some aspects of sound symbolism and mimetics.

Mimetics in Japanese are quite well-defined in terms of construction and often appear as an adverb, becoming part of the verb complex — relational items.

<table>
<thead>
<tr>
<th>constructional schema</th>
<th>mimetic</th>
<th>Japanese</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABAB</td>
<td>korokoro</td>
<td>コロコロ</td>
<td>'small thing rolling’</td>
</tr>
<tr>
<td>ABN</td>
<td>koron</td>
<td>コロン</td>
<td>‘small thing rolling once’</td>
</tr>
<tr>
<td>ABri</td>
<td>korori</td>
<td>コロリ</td>
<td>'small thing rolling once’</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Acquiring ideophones / mimetics

Cross-linguistically, the concept of ideophones is generally defined as “marked words that depict sensory imagery, and which belong to an open lexical class” (Dingemanse 2011; 2012; 2019)

Chinese also has a large number of ideophones (cf. Mok 2001; Lu 2006; Bodomo 2008; Meng 2012; Van Hoey 2015; Van Hoey & Thompson 2019), spanning onomatopoeia (sound ideophones) but also other modalities (visual and inner feelings being quite frequent).
Chinese ideophones

霧 茫茫
wù mángmáng
mist hazy.IDEO
“It’s foggy.”

狗 吠 汪汪
gǒu fèi wāngwāng
dog bark woofwoof.IDEO
“The dog is barking woofwoof.”
Chinese ideophones

This paper reads sloppily.

It feels slippery and very cold.

Ree Lin's pet snake Seysey
RQ:
How are ideophones acquired in Chinese?

We want to know **how ideophones and ideophonized constructions** are acquired in Mandarin Chinese.

What are some **factors** that can help their acquisition?

Does this differ from other languages?
Material and methodology
We are using the **CHILDES** database (MacWhinney 2000) “CHIld Language Data Exchange System”

Collaborative effort to exchange recorded and transcribed transcriptions between infants or children and adults, in order to study how input relates to output, viz. how language is learned.

The data is stored in a standardized manned (**CHAT** ‘Codes for the Human Analysis of Transcripts’) and can be queried with CHILDES’s **CLAN** query builder.
CHILDES and childes-db

While the CHILDES project (since 1984!) is very impressive, it is hard to master these idiosyncratic query languages. In this age of data-science, familiarity with R or python etc. should help us make use of CHILDES as well.

Sanchez et al. (2018) developed an R mirror of CHILDES, called childes-db:

• Improve efficiency
• Reduce errors and inconsistencies
• Share scripts and improve reproducibility
• Track previous instances of CHILDES
CHILDES and childes-db

CHILDES

https://childes.talkbank.org

childes-db

http://childes-db.stanford.edu
version 2018.1

childesr

https://github.com/langcog/childesr
version 0.1.1

MATERIAL for this study
<table>
<thead>
<tr>
<th>Mandarin</th>
<th>Cantonese</th>
<th>Taiwanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcadLang</td>
<td>HKU-70</td>
<td>Tsay</td>
</tr>
<tr>
<td>Beijing</td>
<td>Lee/Wong/Leung</td>
<td></td>
</tr>
<tr>
<td>Chang1</td>
<td>PaidoCantonese</td>
<td></td>
</tr>
<tr>
<td>Chang2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiZhou</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xinjiang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZhouDinner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZhouNarratives</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## CHILDES — childes-db: Chinese

<table>
<thead>
<tr>
<th>Mandarin</th>
<th>Cantonese</th>
<th>Taiwanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcadLang</td>
<td>HKU-70</td>
<td></td>
</tr>
<tr>
<td>Beijing</td>
<td>Lee/Wong/Leung</td>
<td></td>
</tr>
<tr>
<td>Chang1</td>
<td>PaidoCantonese</td>
<td></td>
</tr>
<tr>
<td>Chang2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LiZhou</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCCM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tong</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xinjiang</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZhouDinner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZhouNarratives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XuMinChen</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For this study 237,887 utterances of Mandarin Chinese are usable
Methodology

1. **Quantitative** overview of onomatopoeia, ideophones, and reduplicative constructions, using simple exploratory techniques from datascience (using the R language and mostly tidyverse packages)

2. **Qualitative** case studies

3. Discussion on markedness, depiction and intersubjectivity
Quantitative results
Onomatopoeia

Glossed as ‘on’ (onomatopoeia)

Mandarin onomatopoeia uttered by children and adults

<table>
<thead>
<tr>
<th>Age</th>
<th>Example utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1;2</td>
<td>嘎嘎 / 呱呱 / 吩</td>
</tr>
<tr>
<td>1;8</td>
<td>哐啪 / 吩 / 汪汪 / 吩 / 叨 / 嘀 / 吱 / 吱 / 嘀 / 嘀</td>
</tr>
<tr>
<td>2;11</td>
<td>那个 吩 吩 来了</td>
</tr>
<tr>
<td>3;0</td>
<td>有个 小 鸭子 吱嘎 吱嘎 吱</td>
</tr>
<tr>
<td>3;6</td>
<td>啪咚咚 吱咚咚</td>
</tr>
<tr>
<td>4;0</td>
<td>轮胎 破 掉 了 轮胎 吱 / 然后 他 就 拿 着 枪 砰 砰</td>
</tr>
<tr>
<td>4;6</td>
<td>这个 嘤 嘤 嘤</td>
</tr>
<tr>
<td>5;0</td>
<td>它 就 会 吱</td>
</tr>
<tr>
<td>5;6</td>
<td>咔嚓 咔嚓 / 眼 汪汪 的</td>
</tr>
<tr>
<td>6;0</td>
<td>怪 吱呼 带 顶 帽子</td>
</tr>
</tbody>
</table>
Onomatopoeia

- Weird shape of plot
- Normally bell curve
- But evidence of early usage
Ideophones

A better result may be obtained by comparing the childes-db to the Chinese Ideophone Database (CHIDEOD, Van Hoey & Thompson 2019)

*Leaving out very frequent sentence particles like a 啊, ba 吧 etc.*
Reduplicated items

Closer investigation of the previous groups showed that reduplication occurs very frequently in onomatopoeia and ideophones.
This prompted our interest into reduplicated items in general.
Let us look at the top 10 reduplicated items:
### Reduplicated items

<table>
<thead>
<tr>
<th>item</th>
<th>pinyin transcription</th>
<th>meaning</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>爸爸</td>
<td>bàbà</td>
<td>dad</td>
<td>1930</td>
</tr>
<tr>
<td>妈妈</td>
<td>māmā</td>
<td>mom</td>
<td>1448</td>
</tr>
<tr>
<td>然后然后</td>
<td>ránhòu ránhòu</td>
<td>and then, and then</td>
<td>533</td>
</tr>
<tr>
<td>谢谢</td>
<td>xièxiè</td>
<td>thank you</td>
<td>433</td>
</tr>
<tr>
<td>他他</td>
<td>tātā</td>
<td>he he</td>
<td>352</td>
</tr>
<tr>
<td>毛毛</td>
<td>máomáo</td>
<td>1. name; 2. hairy</td>
<td>325</td>
</tr>
<tr>
<td>圆圆</td>
<td>yuányuán</td>
<td>round</td>
<td>287</td>
</tr>
<tr>
<td>哥哥</td>
<td>gēgē</td>
<td>older brother</td>
<td>255</td>
</tr>
<tr>
<td>看看</td>
<td>kànkàn</td>
<td>look for a short while</td>
<td>185</td>
</tr>
<tr>
<td>甜甜</td>
<td>tiántián</td>
<td>sweet</td>
<td>176</td>
</tr>
</tbody>
</table>
Reduplicated items

<table>
<thead>
<tr>
<th>item</th>
<th>pinyin transcription</th>
<th>meaning</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>爸爸</td>
<td>bàbà</td>
<td>dad</td>
<td>1930</td>
</tr>
<tr>
<td>妈妈</td>
<td>māmā</td>
<td>mom</td>
<td>1448</td>
</tr>
<tr>
<td>然后然后</td>
<td>ránhòu ránhòu</td>
<td>and then, and then</td>
<td>533</td>
</tr>
<tr>
<td>谢谢</td>
<td>xièxiè</td>
<td>thank you</td>
<td>433</td>
</tr>
<tr>
<td>他他</td>
<td>tātā</td>
<td>he he</td>
<td>352</td>
</tr>
<tr>
<td>毛毛</td>
<td>máomáo</td>
<td>1. name; 2. hairy</td>
<td>325</td>
</tr>
<tr>
<td>圆圆</td>
<td>yuányuán</td>
<td>round</td>
<td>287</td>
</tr>
<tr>
<td>哥哥</td>
<td>gēgē</td>
<td>older brother</td>
<td>255</td>
</tr>
<tr>
<td>看看</td>
<td>kànkàn</td>
<td>look for a short while</td>
<td>185</td>
</tr>
<tr>
<td>甜甜</td>
<td>tiántián</td>
<td>sweet</td>
<td>176</td>
</tr>
</tbody>
</table>

Mostly interested in these items, because they often occur in a construction ‘XXde’ (XX的)
The XXde construction

Different choices to be made on the speaker’s side (onomasiological choice).

In the case of round objects later on are talked about by ROUND-SHAPE but earlier on by ROUND, \( \text{ROUND.IDEOZ}=\text{LNK} \), and \( \text{ROUND}=\text{LNK} \).

1. \[ \begin{array}{c|c}
\text{圆圆} = \text{的} \\
\text{yuányuán} = \text{de} \\
\text{ROUND} = \text{LNK}
\end{array} \]

2. \[ \begin{array}{c|c}
\text{圆} = \text{的} \\
\text{yuán} = \text{de} \\
\text{ROUND} = \text{LNK}
\end{array} \]

3. \[ \begin{array}{c|c}
\text{圆 − 形} \\
\text{yuán − xíng} \\
\text{ROUND} = \text{SHAPE}
\end{array} \]

4. \[ \begin{array}{c|c}
\text{圆} \\
\text{yuán} \\
\text{ROUND}
\end{array} \]
The XXde construction

Extension relationship
(Langacker 1987; 1991; 2008)
The XXde construction

\[
\begin{align*}
XX = \text{de} & \quad | \quad \text{ADJECTIVE. IDEOZ = LNK} \\
XX = \text{的} & \quad | \quad \text{ROUND. IDEOZ = LNK} \\
\end{align*}
\]

Elaborative relationships
(Langacker 1987; 1991; 2008)

\[
\begin{align*}
\text{yuányuán} = \text{de} & \quad | \quad \text{ROUND. IDEOZ = LNK} \\
\frac{圆圆 = \text{的}}{} & \\
\end{align*}
\]

\[
\begin{align*}
\text{yuán} = \text{de} & \quad | \quad \text{ROUND = LNK} \\
\frac{圆 = \text{的}}{} & \\
\end{align*}
\]

\[
\begin{align*}
\text{yuan} - \text{xíng} & \quad | \quad \text{ROUND - SHAPE} \\
\frac{圆 - 形}{} & \\
\end{align*}
\]

\[
\begin{align*}
\text{yuan} & \quad | \quad \text{ROUND} \\
\frac{圆}{} & \\
\end{align*}
\]
Reduplication: ideophonized constructions

XXde uttered by children and adults

role
adult
child
Reduplication: ideophonized constructions

4 main periods

1. (0;0 – 2;0)
   input from adult
2. (2;0 – 3;9)
   rising out from child
3. (3;9 – 6;3)
   child usage peak
4. (6;3 – ...)  
   child post-peak
Qualitative case studies
Reduplication: ideophonized constructions

4 main periods

1. (0;0 – 2;0)
   input from adult

2. (2;0 – 3;9)
   rising out from child

3. (3;9 – 6;3)
   child usage peak

4. (6;3 – ...)  
   child post-peak
The sun is round and the moon crooked (stage II)

A 這 什 么 颜 色
A 圆 圆 的 红 太 阳
C 我 正 准 备 画 圆 圆 的 红 太 阳 呢
A 好 呦
A 这 太 阳 画 得 真 漂 亮
C 好 啦
A 好 啦
A 这 画 了 几 个 太 阳 呀
A 一 个 太 阳 两 个 太 阳 三 个 五 个 太 阳
A 你 画 个 小 鱼
A 好 不 好 呀
C 这 是 什 么 笔 呀
A 这 是 水 彩 笔
A 圆 圆 的 太 阳 弯 弯 的 月 亮
A 画 个 弯 弯 的 月 亮
A 弯 弯 的 月 小 小 的 船
C 弯 弯 的 月 小 小 的 船

What colour is this?
A round-round red sun
I’m just about to draw a round-round red sun
Okay
You drew this sun really beautifully
Okay
Okay
How many suns did you draw here?
One sun, two suns, three suns, five suns
Let’s draw a small fish
Can you do it?
What kind of brush is this?
This is a watercolour brush
A round-round sun and a crooked-crooked moon
Let’s draw a crooked-crooked moon
A crooked-crooked moon, a small-small boat
A crooked-crooked moon, a small-small boat
Apples are red-red, hard-hard, fragrant-fragrant and sweet-sweet (stage III)

Apple skin is red-red
And the meat inside is white-colour
Yes
The stem is black-black
What does it feel like?
Hard-hard
Good, what does it smell like?
Fragrant-fragrant
And if you bite it?
Sweet-sweet
Very good. And if you bite the apple but don’t eat and put if over there, what will happen?
Become black
Very good
What is going on?

Around age 1;0 Mandarin acquiring infants already use a number of onomatopoeia / ideophones.

At around age 2;0 simple but real dialogues can occur between infants and adults, in which they get output that contains reduplicated constructions.

In these dialogues there is a certain object that the child is asked about and a more contentful conversation emerges.

This is possible because of the joint attention the child and the adult have towards the object.
Intersubjectivity: in three stages

1st order
Proto-mimesis
- neonatal imitation
- (simple) empathy
- mutual attention

Zlatev (2008:227)
Intersubjectivity: in three stages

1st order
Proto-mimesis
- neonatal imitation
- (simple) empathy
- mutual attention

2nd order
Dyadic mimesis
- cognitive empathy
- shared attention
- understanding other’s intentions

Zlatev (2008:227)
Intersubjectivity: in three stages

1st order
Proto-mimesis
- neonatal imitation
- (simple) empathy
- mutual attention

2nd order
Dyadic mimesis
- cognitive empathy
- shared attention
- understanding other’s intentions

3rd order
Triadic mimesis
- joint attention
- haven and understanding intentions

Protolanguage
- semantic conventions

Language
- (false) belief understanding

Zlatev (2008:227)
Reduplication as a trigger for intersubjectivity

**Ideophones** are
- marked
- words
- that depict
- sensory imagery,
- which belong to an open lexical class

**Reduplicative constructions** are
- marked
- constructions
- that depict
- sensory imagery,
- which belong to an open lexical class

Based mostly on the criteria of MARKEDNESS and DEPICTION, we believe that the function of these reduplicative constructions is multifold:
1. Draw attention to the object
2. Scaffold the language for the child
3. Depiction of what the object looks, feels, smells, tastes... like
   ("ideophones are the next best thing to having been there." Levinson, quoted by Dingemanse 2011:299)
Reduplication as a trigger for intersubjectivity

While reduplication are is thus an important factor facilitator for intersubjectivity, it is neither sufficient nor necessary — other means, e.g. gesture, exist to convey these language elements.

Important to note, is that

1. There is an onomasiological choice made at every scaffolding moment (from the adult’s perspective)
2. This XXde construction persists well into adult language, so it is not (just) motherese/parentese.
Conclusions
Ideophones are starting to be learned early on, especially onomatopoeia

Onomatopoeia / ideophones are acquired from about the age of 1;0 onwards.

In Cantonese and Japanese they frequency shows nice bell-curves, but in Mandarin it did not.

This is perhaps due to gaps in the data collection, or due to less salient constructions that would fit the idea of onomatopoeia / ideophones.

Still, the peaks seems to be around 3;0 and 4;0 — later than Japanese (2;8) or Cantonese (2;6).
Reduplication can act as a signal to trigger intersubjectivity through markedness and depiction.

We have attempted to show that, based on corpus material, reduplication can act as a trigger for intersubjectivity — most notably to let the acquiring child ‘second-hand’ experience the sensory imagery of the object.

This is mostly limited to basic level items, or items that the child is referentially familiar with (sun, moon, animals, fruit…).

Furthermore, we acknowledge that this work is still “in its child’s shoes” (Dutch phrase), so more research is needed to investigate the interplay more comprehensively.
Evaluation of the material and methodology

As for material and methodology, the childes-db as a mirror of CHILDES works well:
+ the data is open,
+ Scripts will (soon) be available at github.com/simazhi
- Many utterances are ‘lost’ because of lack of child_age
- Typical for corpora: you have to make do with what you have — we did not collect the data, but recycle it.
+ That means we did not manipulate it either
Thank you!
ご清聴ありがとうございました！
References


References


References


