



Reduplication as a trigger of intersubjectivity:

Mandarin Chinese ideophones and reduplication in the CHILDES corpora

Thomas Van Hoey
National Taiwan University
國立台灣大學

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

Chiarung Lu
National Taiwan University
國立台灣大學

Roadmap

Introduction

language acquisition and ideophones

Research question

how are ideophones and related
constructions acquired in Chinese?

Material

CHILDES — childesdb

Quantitative results

sun/moon; apple

Qualitative case studies

intersubjectivity

Discussion

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 facilitates learning (Goldberg & Caseniser 2008).
- Salience and prototypicality (cf. Geeraerts 2000; 2017) also prove an important factor.



Apple (píngguǒ 蘋果)

- 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.

constructional schema	mimetic	Japanese	meaning
ABAB	<i>korokoro</i>	コロコロ	'small thing rolling'
ABN	<i>koron</i>	コロン	'small thing rolling once'
ABri	<i>korori</i>	コロリ	'small thing rolling once'
...			

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

雾 茫茫

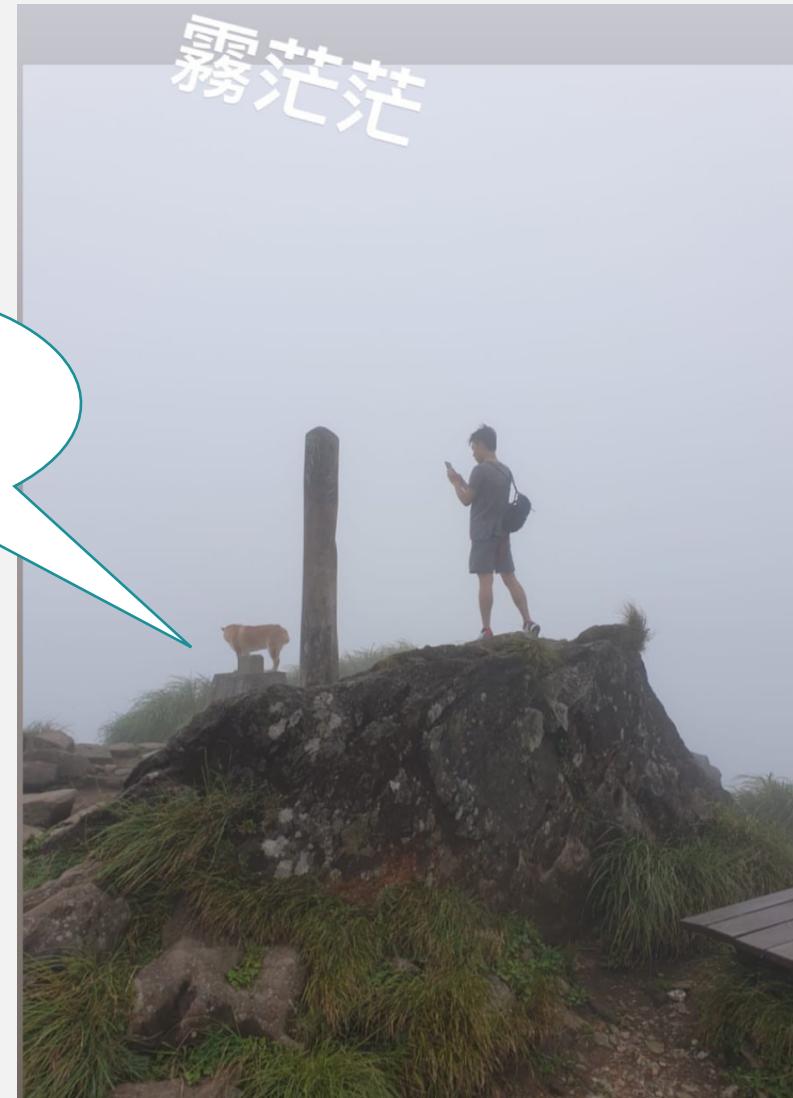
wù mángmáng
mist hazy.IDEO

“It's foggy.”

狗 吠 汪汪

gǒu fèi wāngwāng
dog bark woofwoof.IDEO

“The dog is barking woofwoof.”



Instagram: @nickprometheus31

Chinese ideophones

這 篇 文章 讀起來 凌亂。

zhè piān wénzhāng dú-qǐ.lái líng~luàn
DEM CL paper read-MID in.a.mess.IDEO

“This paper reads sloppily.”

摸起來 滑滑的， 冷冷的。

mō-qǐ.lái huá~huá=de léng~léng=de

touch-MID slippery.IDEO=LNK cold.IDEO=LNK

“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

CHILDES database

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 manner ([**CHAT**](#) ‘Codes for the Human Analysis of Transcripts’) and can be queried with CHILDES’s [**CLAN**](#) query builder.

CHILDES and childe-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 **childe-db**:

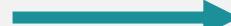
- Improve efficiency
- Reduce errors and inconsistencies
- Share scripts and improve reproducibility
- Track previous instances of CHILDES

CHILDES and childe-db



CHILDES

<https://childe.talkbank.org>



childe-db

<http://childe-db.stanford.edu>

version 2018.1



childesr

<https://github.com/langcog/childesr>

version 0.1.1



**MATERIAL
for this study**

CHILDES — childe-db: Chinese

Mandarin	Cantonese	Taiwanese
AcadLang	HKU-70	Tsay
Beijing	Lee/Wong/Leung	
Chang1	PaidoCantonese	
Chang2		
Context		
LiZhou		
TCCM		
Tong		
Xinjiang		
Zhou1		
Zhou2		
ZhouDinner		
ZhouNarratives		

CHILDES — childe-db: Chinese

Mandarin	Cantonese	Taiwanese
AcadLang	HKU-70	ㄔㄞ
Beijing	Lee/Wong/Leung	
Chang1	PaidoCantonese	
Chang2		
Context		
LiZhou		
TCCM		
Tong		
Xinjiang		
Zhou1		
Zhou2		
ZhouDinner		
ZhouNarratives		
XuMinChen		

CHILDES – childe-db: Chinese

	Mandarin	Cantonese	Total
Total utterances	435,452	281,371	716,823
Usable for age _{child}	237,887	281,371	519,258
Not usable for age _{child}	197,565	NA	197,565

For this study 237,887 utterances of Mandarin Chinese are usable

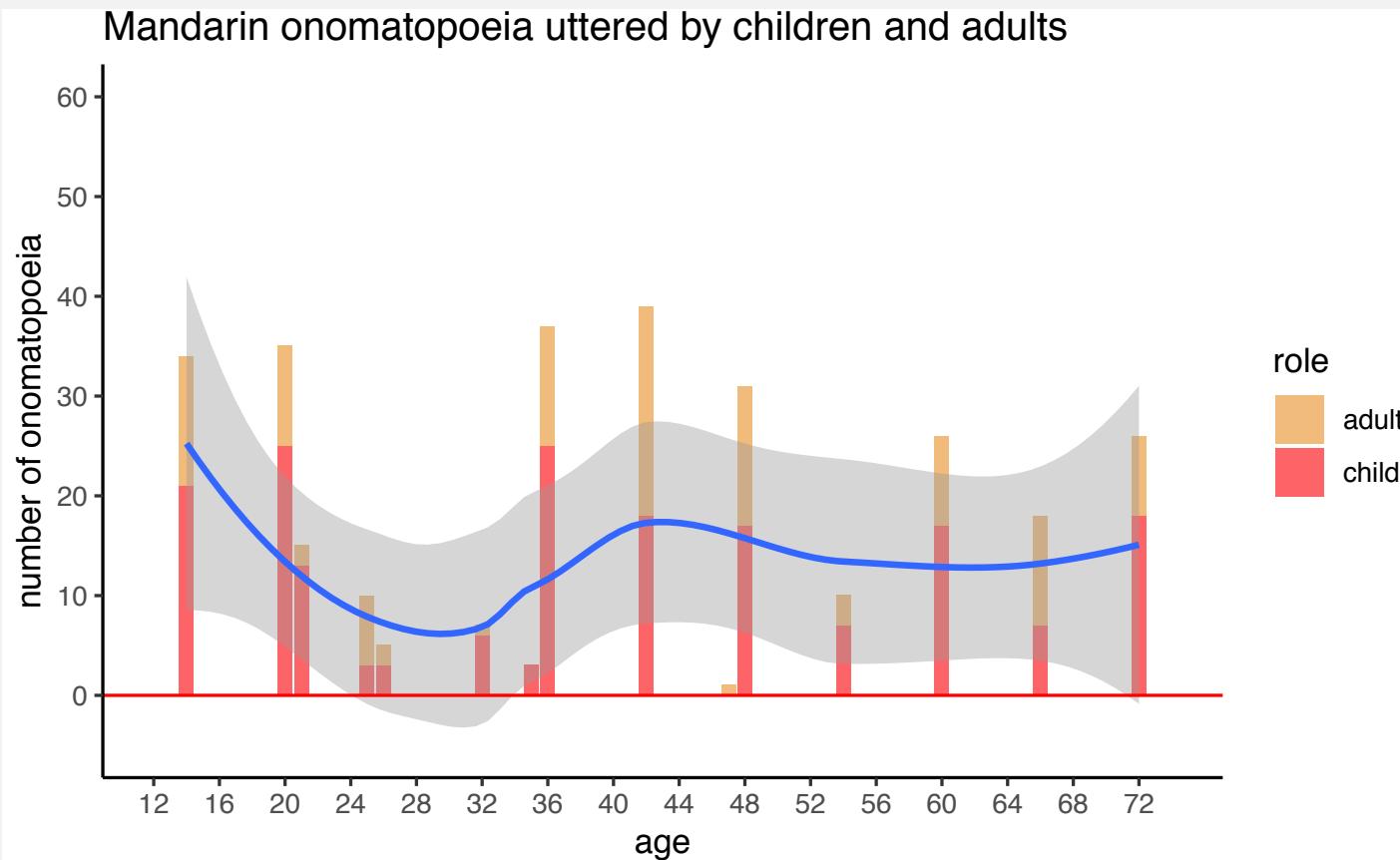
Methodology

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

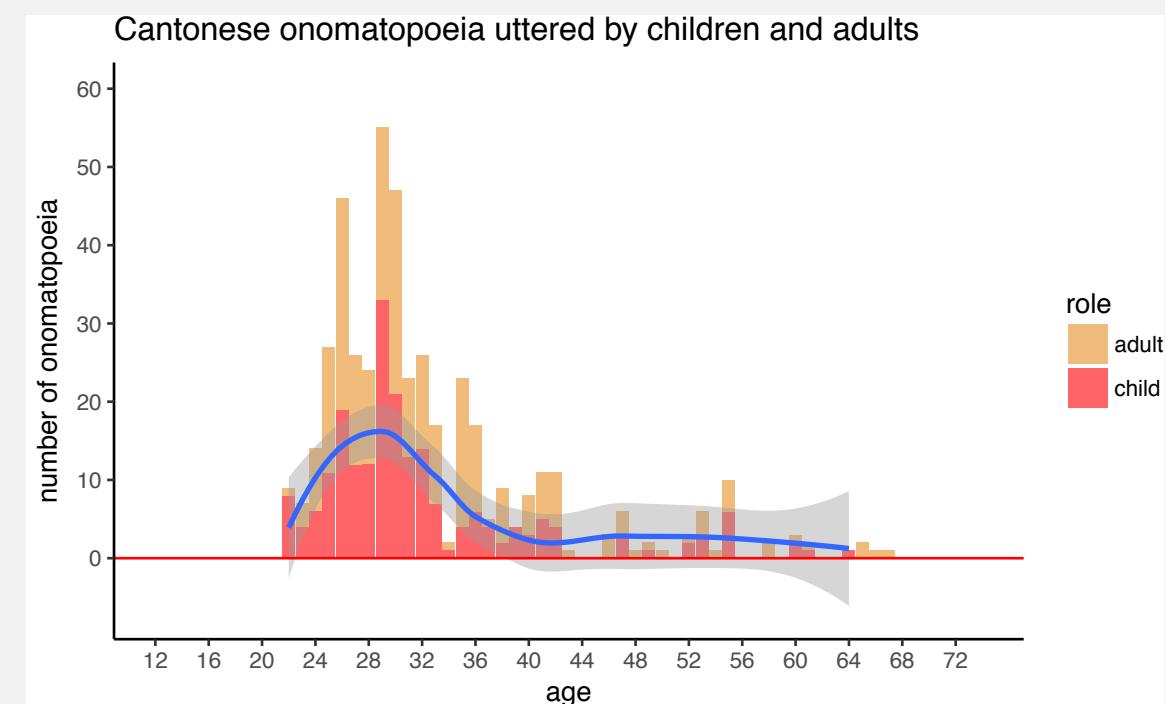
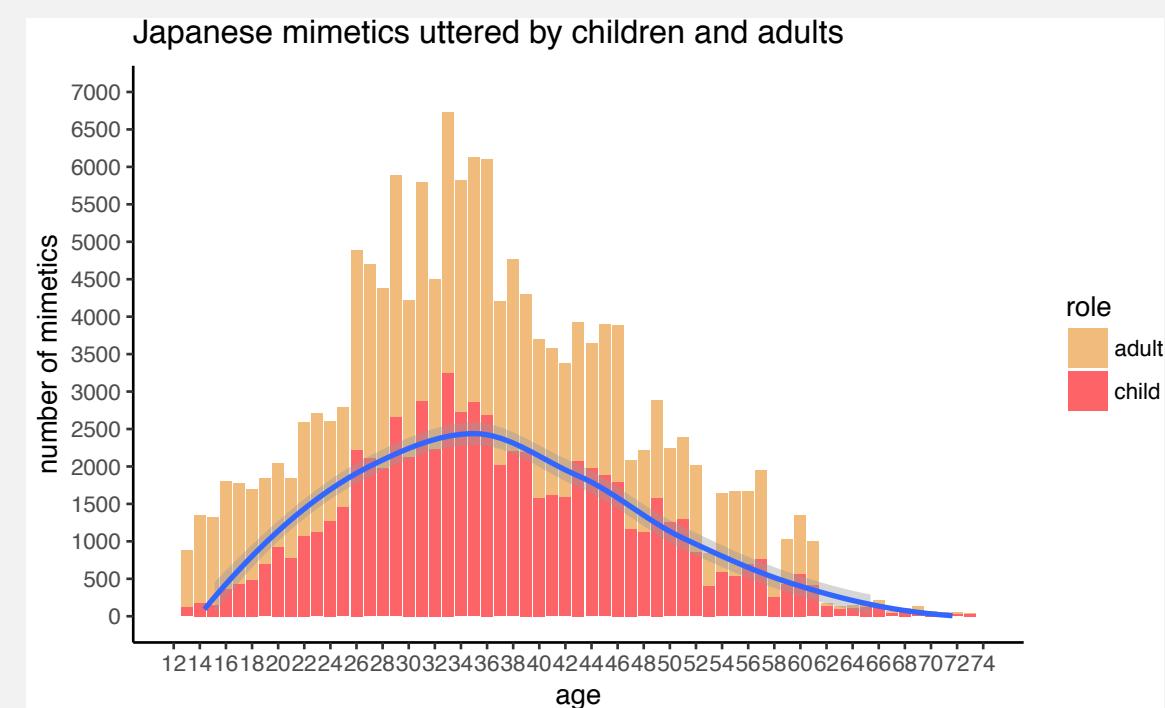
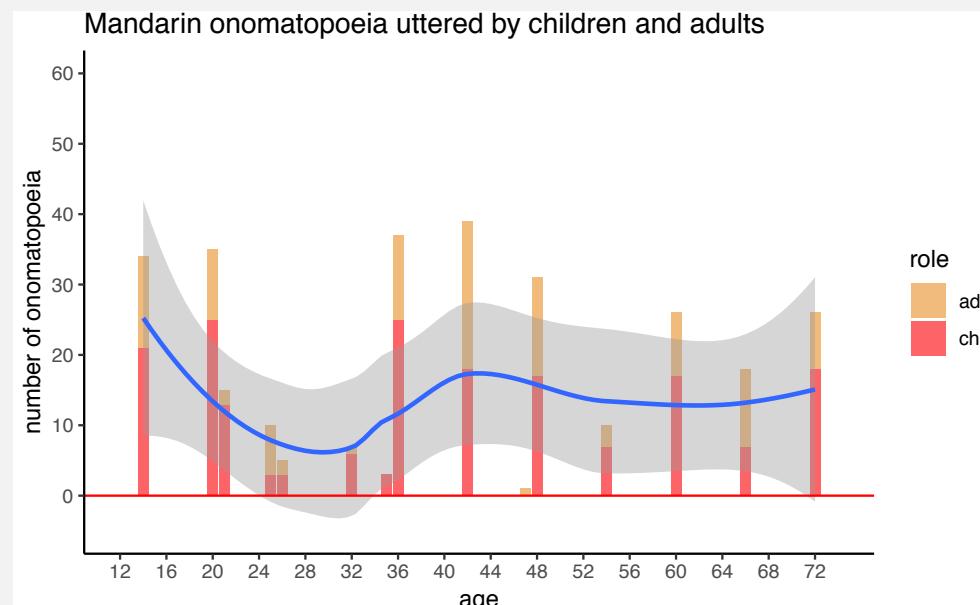
Glossed as 'on' (onomatopoeia)



Age	Example utterance
1;2	嘎嘎 / 呱呱 / 鸣
1;8	啪啪 / 鸣 / 汪汪 / 咩 / 叽叽喳喳叽叽喳
2;11	那个鸣鸣来了
3;0	有个小鸭子嘎嘎嘎嘎嘎
3;6	咚咚咚咚咚咚咚
4;0	轮胎破掉了轮胎啪 / 然后他就拿着枪砰砰
4;6	这个噔噔噔
5;0	它就会咚
5;6	咔嚓咔嚓 / 眼汪汪的
6;0	怪呼呼戴顶帽子

Onomatopoeia

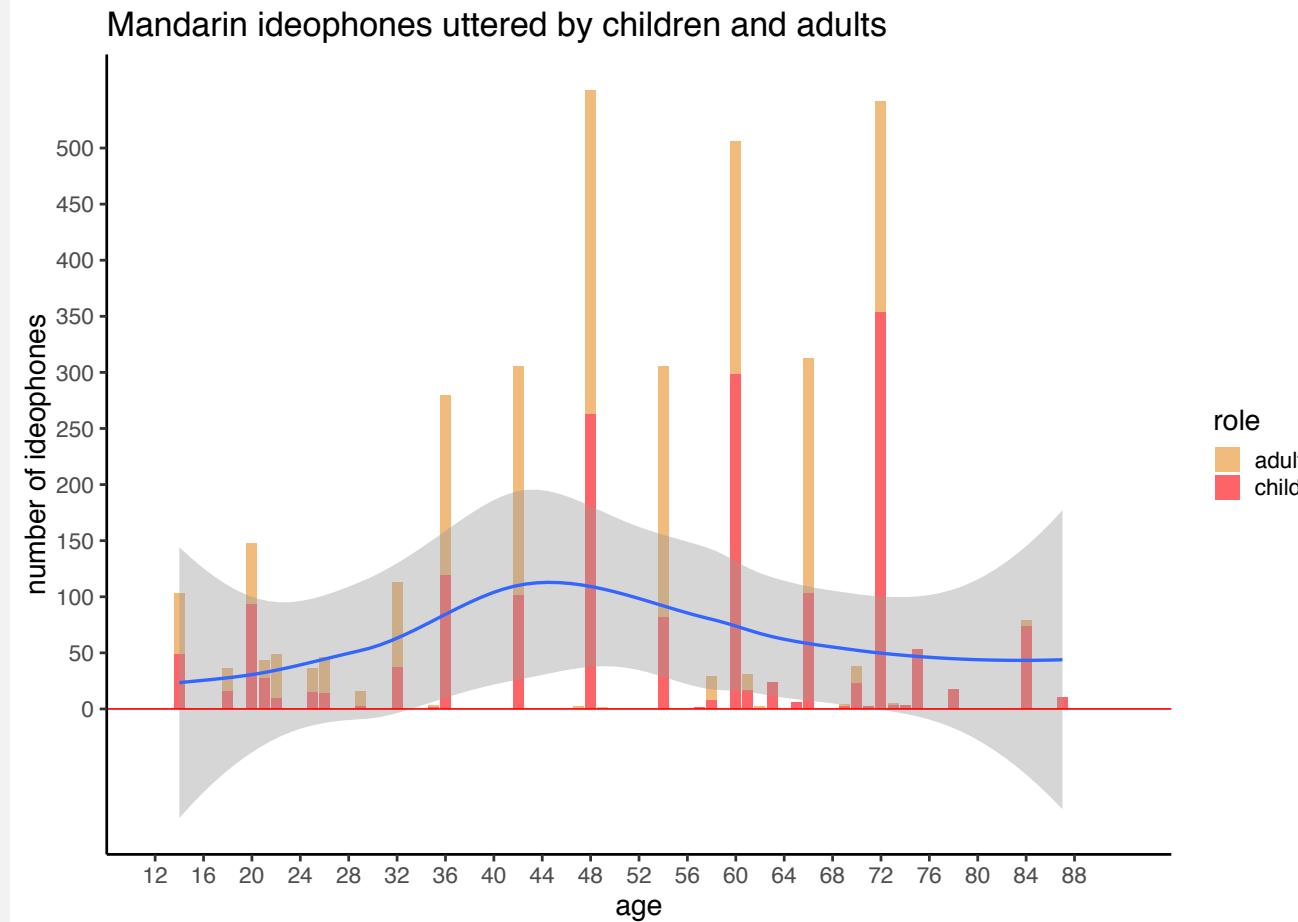
- Weird shape of plot
- Normally bell curve
- But evidence of early usage



Ideophones

A better result may be obtained by comparing the childe-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

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

Reduplicated items

lexicalized

pragmatic

ideophonized

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

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).

1.

[$\frac{\text{yuányuán} = \text{de}}{\text{圓圓} = \text{的}}$ | ROUND. IDEOZ = LNK]

2.

[$\frac{\text{yuán} = \text{de}}{\text{圓} = \text{的}}$ | ROUND = LNK]

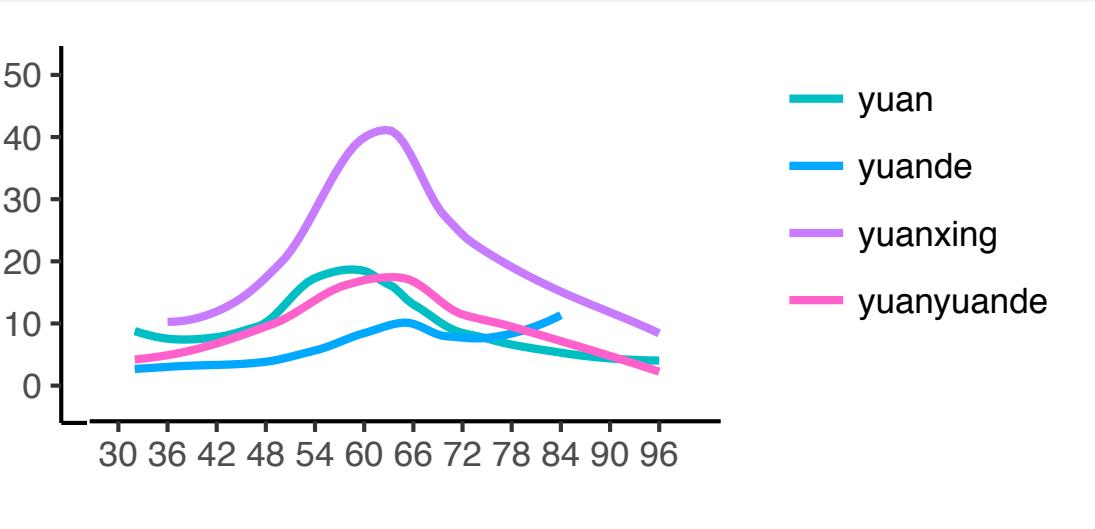
3.

[$\frac{\text{yuán} - \text{xíng}}{\text{圓} - \text{形}}$ | ROUND - SHAPE]

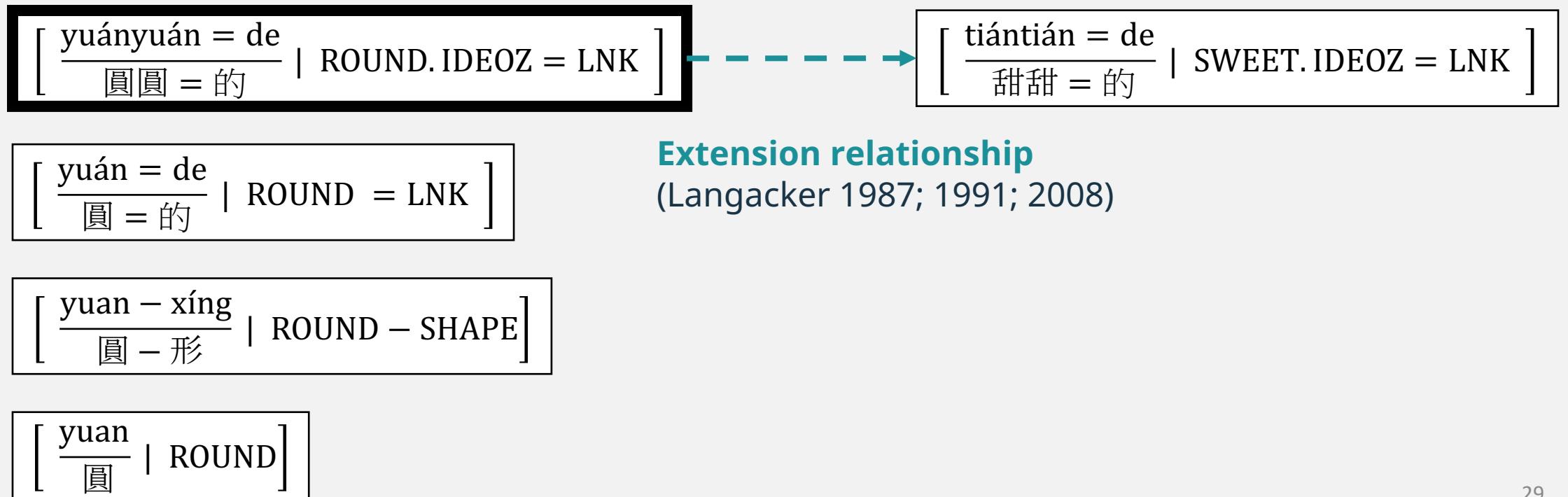
4.

[$\frac{\text{yuán}}{\text{圓}}$ | ROUND]

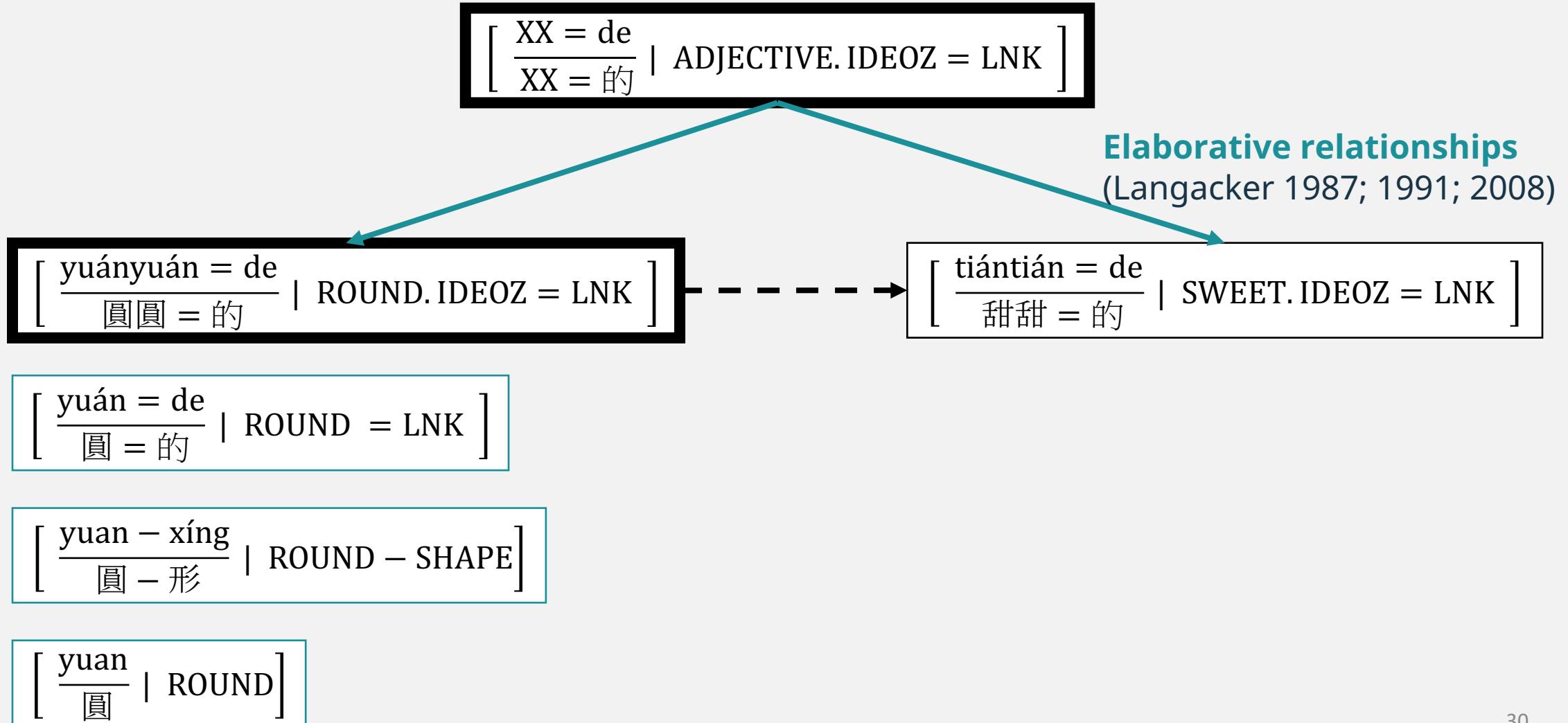
In the case of round objects later on are talked about by ROUND-SHAPE but earlier on by ROUND, **ROUND.IDEOZ=LNK**, and ROUND=LNK.



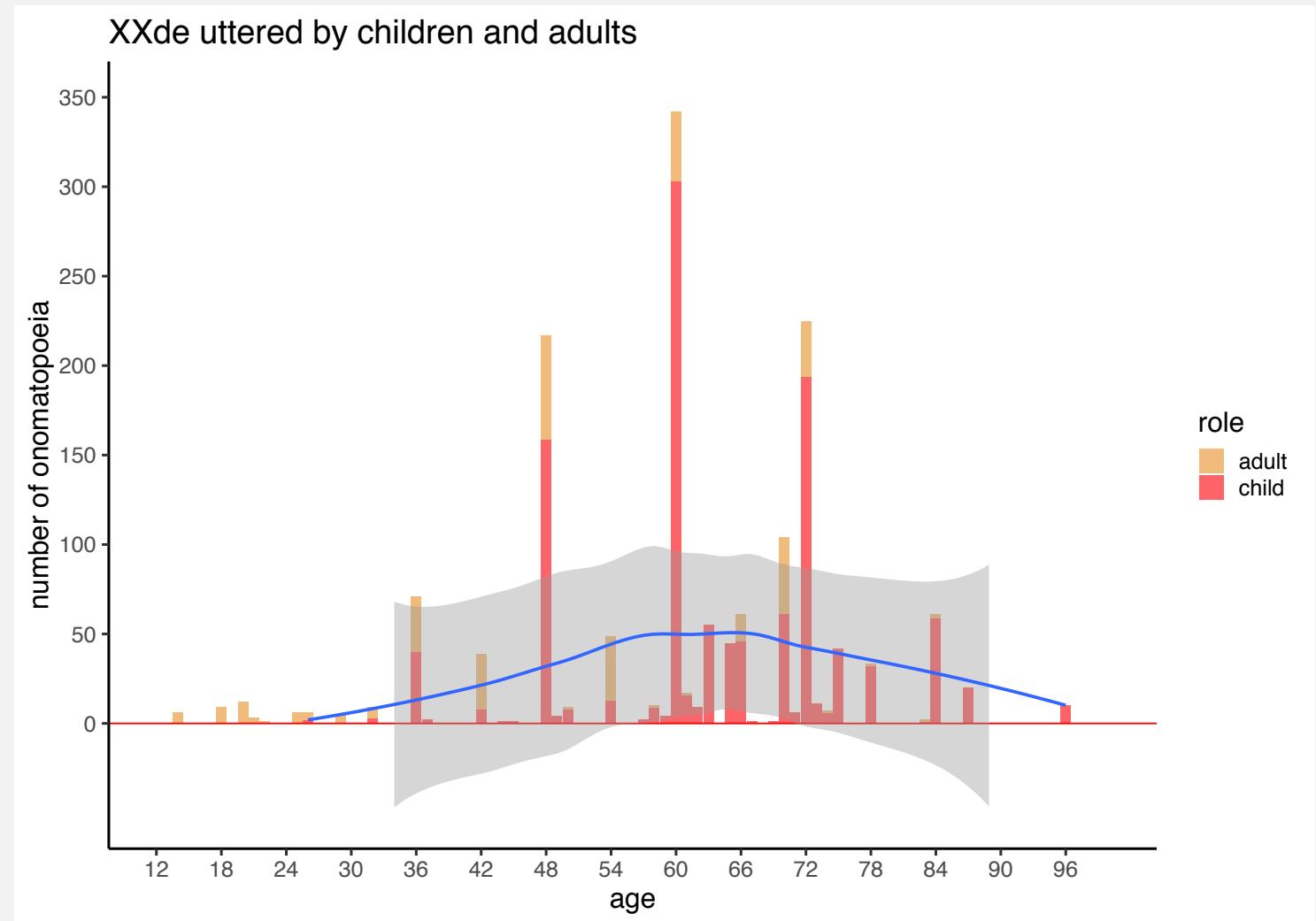
The XXde construction



The XXde construction



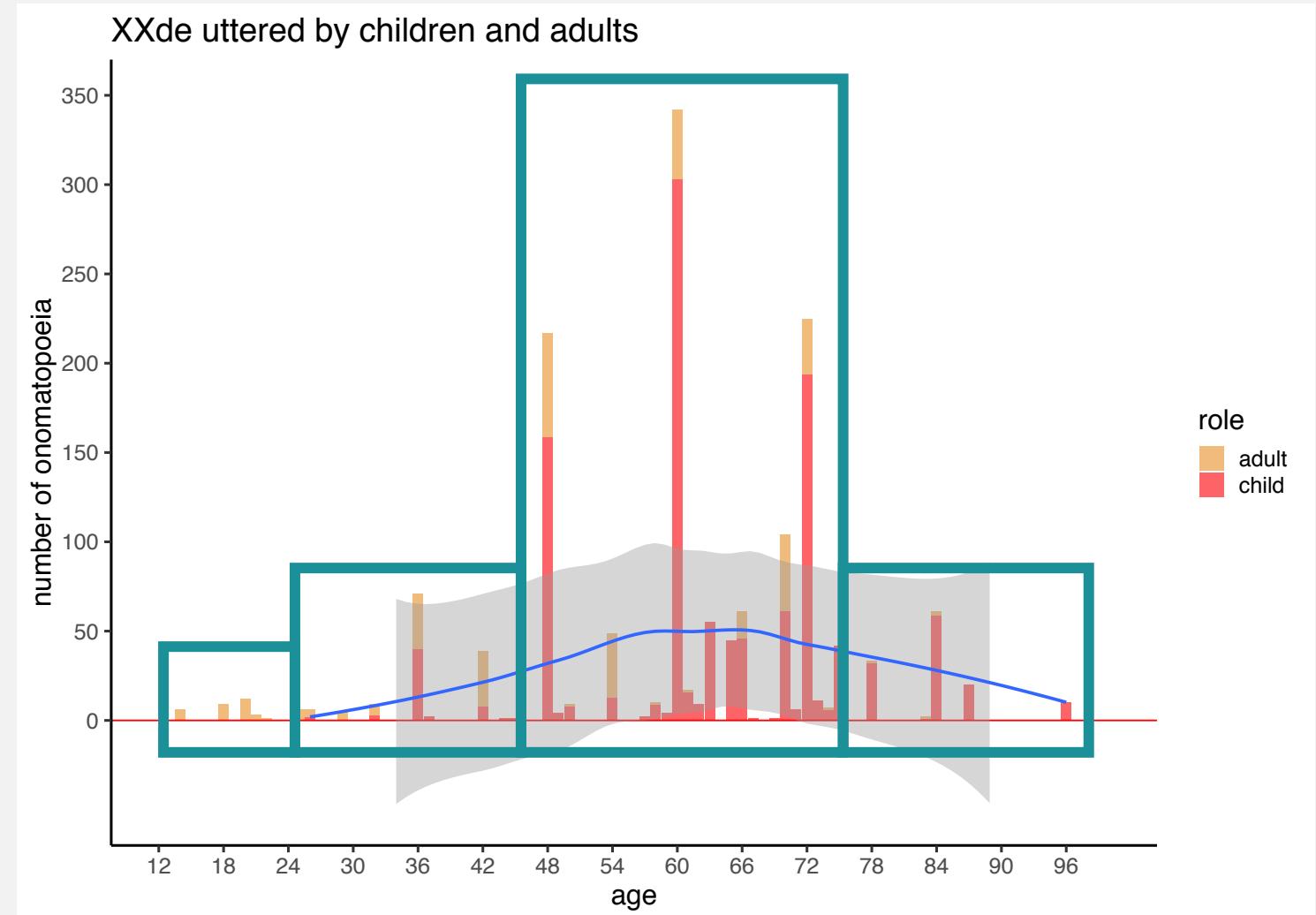
Reduplication: ideophonized constructions



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

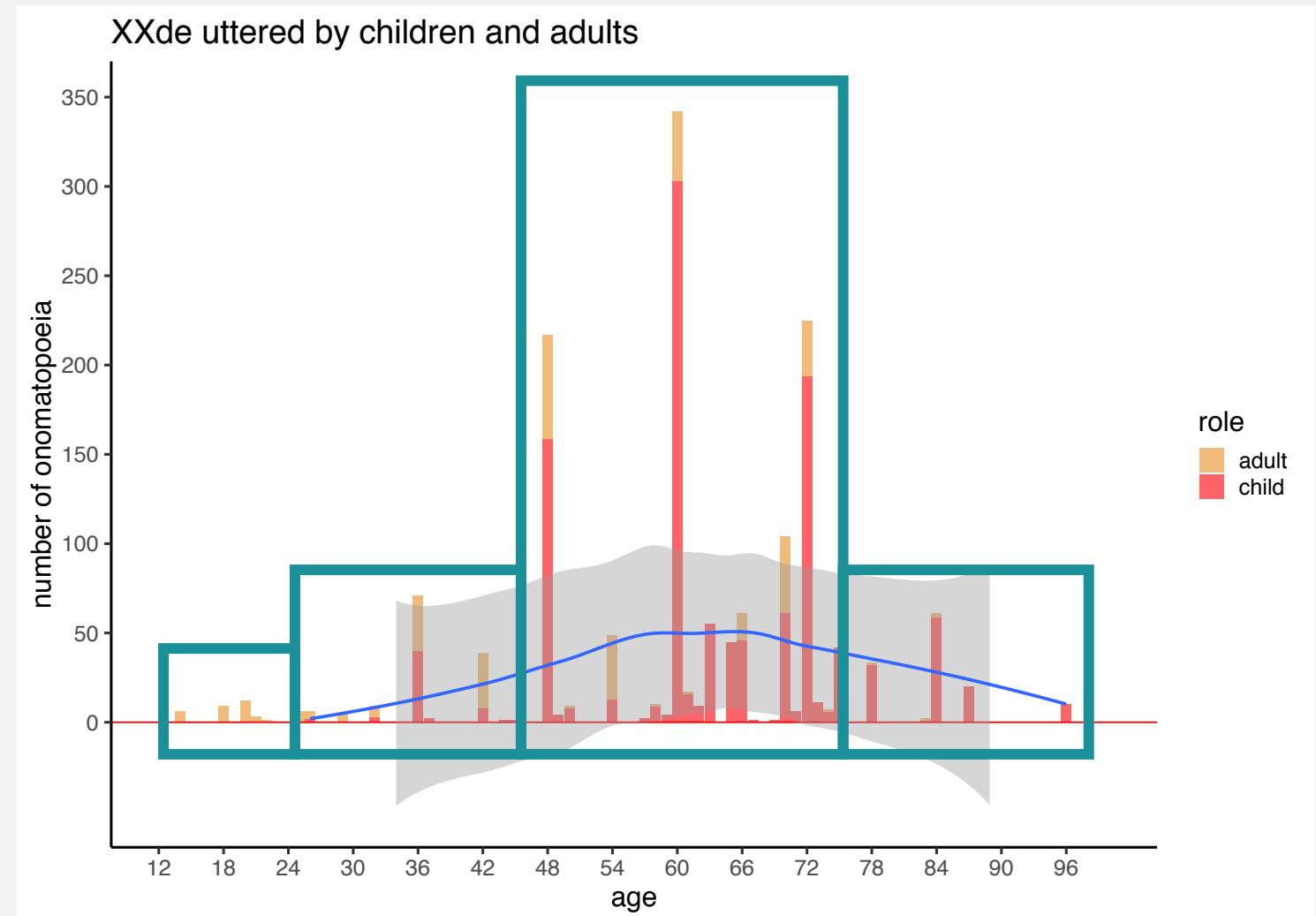
4 main periods

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 这是什么颜色
A 圆圆的 红太阳
C 我正准备画圆圆的 红太阳呢
A 好哟
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 月儿 小小的 船
C 弯弯的 月儿 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)

C	苹果的皮是 红红的	Apple skin is red-red
C	而且里面的肉是白色的	And the <i>meat</i> inside is white-colour
A	嗯	Yes
C	它的把子 黑黑 的	The stem is black-black
A	摸上去	What does it feel like?
C	硬硬 的	Hard-hard
A	好闻 上去	Good, what does it smell like?
C	香香 的	Fragrant-fragrant
A	咬在嘴巴 里面	And if you bite it?
C	甜甜 的	Sweet-sweet
A	很好 那咬了一口苹果 不吃放在那边它就会怎么样	Very good. And if you bite the apple but don't eat and put if over there, what will happen?
C	黑 掉	Become black
A	很好	Very good

AcadLang, age (4;2)
Transcript_id: 11450
Utterances: 1-13

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

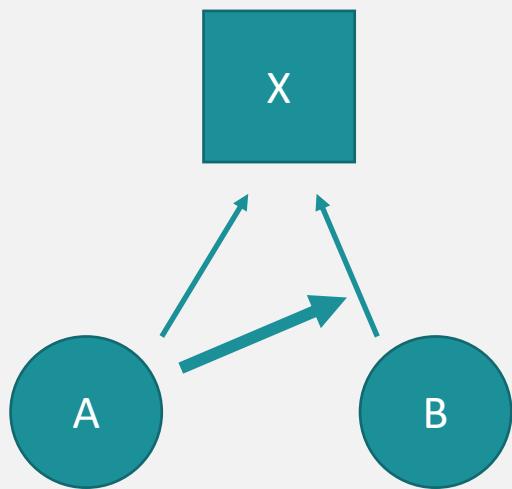
Proto-mimesis

- 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

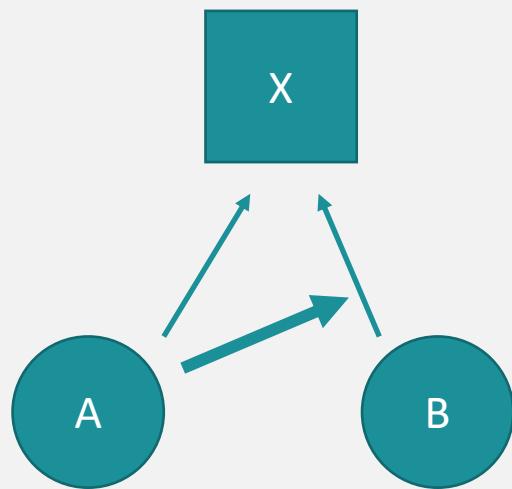
Proto-mimesis

- neonatal imitation
- (simple) empathy
- mutual attention

2nd order

Dyadic mimesis

- cognitive empathy
- shared attention
- understanding other's intentions



3rd order

Triadic mimesis

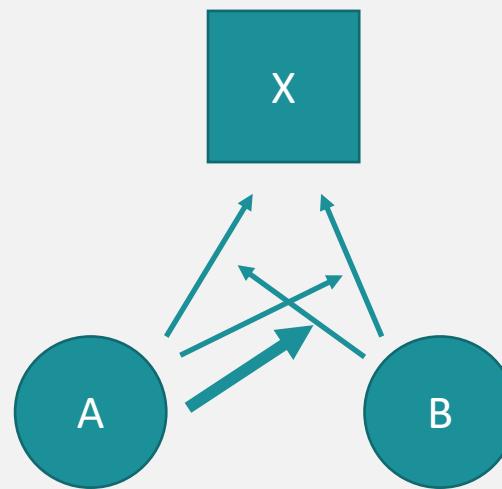
- joint attention
- haven and understanding intentions

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 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 childe-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

- Bodomo, Adams. 2006. The structure of ideophones in African and Asian languages: The case of Dagaare and Cantonese. In John Mugane, John P. Hutchison & Dee A. Worman (eds.), *Selected Proceedings of the 35th Annual Conference on African Linguistics*, 203–213. Somerville, MA: Cascadilla Proceedings Project.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge ; New York: Cambridge University Press.
- Choi, Soonja & Alison Gopnik. 1995. Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language* 22(03). doi:[10.1017/S0305000900009934](https://doi.org/10.1017/S0305000900009934).
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Nijmegen: Radboud University Nijmegen dissertation.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives* (Iconicity in Language and Literature, ILL 16), 13–33. Amsterdam : Philadelphia: John Benjamins Pub. Co.
- Geeraerts, Dirk. 2006. *Words and other wonders: Papers on lexical and semantic topics* (Cognitive Linguistics Research 33). Berlin: Mouton de Gruyter.
- Gentner, Dedre. 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report* 257.
- Gentner, Dedre & Lera Boroditsky. 2001. Individuation, relativity, and early word learning. In Melissa Bowerman & Stephen C. Levinson (eds.), *Language acquisition and conceptual development* (Language, Culture and Cognition 3), 215–256. Cambridge, UK; New York: Cambridge University Press.
- Goldberg, Adele E. & Devin Casenhisser. 2008. Construction learning and second language acquisition. In Peter Jake Robinson & Nick C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 197–215. New York: Routledge.

References

- Imai, Mutsumi, Etsuko Haryu, Hiroyuki Okada, Lianjing Li & Jun Shigematsu. 2006. Revisiting the noun-verb debate: A cross-linguistic comparison of novel noun and verb learning in English-, Japanese-, and Chinese-speaking children. In Kathy Hirsh-Pasek & Roberta M. Golinkoff (eds.), *Action meets word: how children learn verbs*, 450–476. Oxford ; New York: Oxford University Press.
- Imai, Mutsumi & Sotaro Kita. 2014. The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). 1–13. doi:[10.1098/rstb.2013.0298](https://doi.org/10.1098/rstb.2013.0298).
- Imai, Mutsumi, Sotaro Kita, Miho Nagumo & Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109(1). 54–65. doi:[10.1016/j.cognition.2008.07.015](https://doi.org/10.1016/j.cognition.2008.07.015).
- Kim, Mikyong, Karla K. McGregor & Cynthia K. Thompson. 2000. Early lexical development in English- and Korean-speaking children: language-general and language-specific patterns. *Journal of Child Language* 27(2). 225–254. doi:[10.1017/S0305000900004104](https://doi.org/10.1017/S0305000900004104).
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar 1: Theoretical prerequisites*. Stanford, California: Stanford University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar 2: Descriptive application*. Stanford, California: Stanford University Press.
- Langacker, Ronald W. 2008. *Cognitive grammar: a basic introduction*. Oxford ; New York: Oxford University Press.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, Brian & William O'Grady (eds.). 2015. *The handbook of language emergence* (Blackwell Handbooks in Linguistics). Chichester, West Sussex, UK ; Malden, MA: Wiley Blackwell.
- Meng, Chenxi. 2012. *A description of ideophonic words in Mandarin Chinese*. Leiden: Leiden University Research Master in Linguistics.
- Mok, Waiching Enid. 2001. *Chinese sound symbolism: A phonological perspective*. Hawai'i: University of Hawai'i PhD dissertation.
- Ogura, Tamiko, Philip S. Dale, Yukie Yamashita, Toshiki Murase & Aki Mahieu. 2006. The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language* 33(01). 1. doi:[10.1017/S0305000905007270](https://doi.org/10.1017/S0305000905007270).

References

- Sanchez, Alessandro, Stephan Meylan, Mika Braginsky, Kyle MacDonald, Daniel Yurovsky & Michael Frank. 2018. childe-db: A flexible and reproducible interface to the Child Language Data Exchange System. doi:[10.17605/osf.io/93mwx](https://doi.org/10.17605/osf.io/93mwx) (19 June, 2018).
- Tardif, Twila. 1996. Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology* 32(3). 492–504. doi:[10.1037/0012-1649.32.3.492](https://doi.org/10.1037/0012-1649.32.3.492).
- Tardif, Twila. 2006. But are they really verbs? Chinese words for action. In Kathy Hirsh-Pasek & Roberta M. Golinkoff (eds.), *Action meets word: how children learn verbs*, 477–495. Oxford ; New York: Oxford University Press.
- Tomasello, Michael. 1992. *First verbs: a case study of early grammatical development*. Cambridge [UK] ; New York, NY, USA: Cambridge University Press.
- Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Cambridge, Mass: Harvard University Press.
- Van Hoey, Thomas. 2015. *Ideophones in Middle Chinese: A typological study of a Tang dynasty poetic corpus*. Leuven: KU Leuven Master thesis.
- Van Hoey, Thomas & Arthur Lewis Thompson. 2019. Bridging phonology, meaning, and written form across time: Introducing a database of Chinese literary ideophones. Oral presentation presented at the ILL 12 [International Symposium on Iconicity in Language and Literature], Lund: Lund University. <https://konferens.ht.lu.se/en/ill-12/>.
- Waxman, Sandra, Xiaolan Fu, Sudha Arunachalam, Erin Leddon, Kathleen Geraghty & Hyun-joo Song. 2013. Are nouns learned before verbs? Infants provide insight into a long-standing debate. *Child Development Perspectives* 7(3). 155–159. doi:[10.1111/cdep.12032](https://doi.org/10.1111/cdep.12032).
- Wickham, Hadley. 2017. *tidyverse: Easily install and load the "Tidyverse"* (v. 1.2.1). <https://CRAN.R-project.org/package=tidyverse>.
- You, Weilun 游韋倫. 2015. *Nichū ryō gengo ni okeru giongo no imi to imi kakuchō: Furēmu imi-ron no kanten kara no apurōchi* 日中両言語における擬音語の意味と意味拡張 —フレーム意味論の観点からのアプローチ— [The semantics and semantic extensions of onomatopoeia in Japanese and Chinese: A Frame Semantics approach]. Kobe: Kobe University PhD dissertation.
- Zlatev, Jordan. 2008. The co-evolution of intersubjectivity and bodily mimesis. In Jordan Zlatev, Timothy P. Racine, Chris Sinha & Esa Itkonen (eds.), *The shared mind: Perspectives on intersubjectivity* (Converging Evidence in Language and Communication Research v. 12), 215–244. Amsterdam ; Philadelphia: John Benjamins Pub. Co.