**Note: This is a post-print of the final accepted version of the Yang et al (2020). There will be some minor differences between this and the copy-edited version that occurred during the editing process.**

Four-year-old Mandarin-speaking children’s online comprehension of relative clauses.

Wenchun Yang¹, Angel Chan¹,²,³, Franklin Chang⁴,⁷, & Evan Kidd⁵,⁶,⁷

¹ The Hong Kong Polytechnic University
² The Hong Kong Polytechnic University – Peking University Research Centre on Chinese Linguistics
³ Research Centre for Language, Cognition, and Neuroscience
⁴ Kobe City University for Foreign Studies
⁵ Max Planck Institute for Psycholinguistics
⁶ The Australian National University
⁷ ARC Centre of Excellence for the Dynamics of Language

Addresses for correspondence:

Angel Chan
Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
Email: angel.ws.chan@polyu.edu.hk

Evan Kidd
Language Development Department
Max Planck Institute for Psycholinguistics
P.O. Box 310
6500 XD Nijmegen, The Netherlands
Email: evan.kidd@mpi.nl
Abstract

A core question in language acquisition is whether children’s syntactic processing is experience-dependent and language-specific, or whether it is governed by abstract, universal syntactic machinery. We address this question by presenting corpus and on-line processing data from children learning Mandarin Chinese, a language that has been important in debates about the universality of parsing processes. The corpus data revealed that the two different relative clause constructions differentially prefer syntactic subjects and objects. In the experiment, 4-year-old children’s eye-movements were recorded as they listened to the two RC construction types (e.g., Can you pick up the pig that pushed the sheep?). A permutation analysis showed that children’s ease of comprehension was closely aligned with the distributional frequencies, suggesting syntactic processing preferences are shaped by the input experience of these constructions.

Key words: Mandarin; Children; Relative Clauses; On-line Processing, Permutation Analysis.
Mandarin-speaking 4-year-olds’ online comprehension of relative clauses.

The processing of subject and object relative clauses (RCs) is a frequent topic of investigation in psycholinguistics because their study allows researchers to examine interactions between parsing procedures, input experience, and general memory processes. Chinese languages (e.g., Cantonese, Mandarin) are important in this literature because their typologically rare combination of Subject-Verb-Object (SVO) word order and head-final RCs allows opposing theories of syntactic processing to be tested (e.g., Jäger, Chen, Li, Lin, & Vasishth, 2015). Children’s on-line processing of RCs is an important though understudied source of evidence in this debate; assuming a degree of continuity between children and adults, developmental data both constrain and extend the explanatory reach of parsing models.

In the current paper we investigated 4-year-old Mandarin-speaking children’s processing of relative clauses (RCs). We take advantage of the fact that Mandarin has multiple RC constructions, as in (1) – (4).

(1) [RC] mai3 wan2ju4 de ma1ma

buy toy de mother

‘The mother that bought the toy’

(2) [RC] ma1ma mai3 de wan2ju4

mother buy de toy

‘The toy that mother bought.’

(3) [RC] mai3 wan2ju4 de na4 ge4 ma1ma

buy toy de that CL mother

‘The mother that bought the toy’

(4) [RC] ma1ma mai3 de na4 ge4 wan2ju4

mother buy de that CL toy

‘The toy that mother bought’
Sentences (1) and (3) are subject-extracted RCs because the RC modifies the grammatical subject (*the mother*), whereas sentences (2) and (4) are object-extracted because the RC modifies the grammatical object. Crucially, in (1) and (2) the head noun, introduced by the nominaliser *de*, is bare (henceforth ‘DE-RCs’). In contrast, the head noun in (3) and (4) is introduced by the demonstrative (DEM) + classifier (CL) combination (henceforth ‘DCL-RCs’). These two RC types have different discourse-functional properties (Chen, Ming & Jiang, 2015), and there is debate regarding their structural differences (Cheung & Li, 2015; Cheng & Sybesma, 2009), pointing to the conclusion that they constitute overlapping but partially-distinct constructions.

The typological uniqueness of Mandarin (SVO word order + head-final RCs) allows the predictions of different parsing theories to be teased apart. Structurally-oriented approaches assume that that the parser privileges structural information, and that subject RCs should be easier to process because they are less structurally complex than object RCs (Friedmann, Belletti & Rizzi, 2009; Lin & Bever, 2006). In contrast, Gibson’s Dependence Locality Theory (DLT) predicts that processing load increases with the number of unresolved arguments in a sentence. This predicts an object-RC advantage for Mandarin RCs, because the distance between the head noun and the gap is shorter than that for subject RCs. Finally, experience-based approaches argue that the parsing choices depend on frequency information that is accumulated across the course of a speaker’s developmental history with the language (e.g., Fitz et al., 2011; Hale, 2001; Levy, 2008; MacDonald, 2013; Vasishth et al., 2013).

The weight of empirical evidence supports a subject RC advantage in Chinese (for a meta-analysis of adult data see Vasishth et al., 2013; for discussion of acquisition see Tsoi, Yang, Chan, & Kidd, 2019). The data are consistent with both the structurally-oriented and the experience-based approaches, since the *only* structure that has been tested – DE RCs – more
frequently modify subjects than objects\(^1\). However, corpus work has shown that the DCL has different distributional properties from the DE RCs. Chen et al. (2015) reported that DCL RCs more frequently modified objects (70% vs 16%), critically arguing that the DEM-CL sequence (i) signals the need to link an upcoming definite referent (i.e., the head noun) to a co-occurring referent within the RC, and (ii) most frequently this involves linking a grammatical object (i.e., the head noun) to the subject role within the RC. Thus the presence of the DEM-CL sequence is a probabilistically reliable cue to an object RC analysis.

The asymmetry in distribution of DCL- and DE-RCs is pertinent to theoretical debates concerning RC processing. Both the structural and linear-based accounts predict uniform processing of both RC construction types; the structural account predicts a subject advantage whereas the linear-based account predicts, all things being equal, an object advantage for Mandarin. However, experience-based accounts predict differential effects based on frequency-based expectations. Thus testing the two structures in parallel provides a means with which to tease apart the competing predictions of the different theoretical approaches. In the current research we: (i) present corpus data from child-directed speech to demonstrate that the two RC constructions differ in their distributional properties, and (ii) present the results of a visual world eye-tracking experiment that tested whether children’s on-line parsing decisions are guided by this distributional information.

**Study 1: Corpus Study of Child Directed Speech**

We extracted all morphologically tagged adult utterances from six Mandarin child language corpora in CHILDES (approximately 380,000 words in total) using the LUCID Toolkit (Chang, 2017): AcadLang corpus (Zhou doi:10.21415/T5SC9D), Chang1 & Chang 2

\(^1\) Note that the experience-based approach makes specific predictions about the location of the complexity effects on-line, whereas the structure-based approach does not. See Jäger et al. (2015) for more details. Note that the same authors also used the DEM-CL combination to introduce a RC, not a head noun, which has the effect of reducing a local syntactic ambiguity. Other adult processing studies (e.g. Lin & Garnsey; 2011; Wu, Kaiser & Vasishth 2018) also used DEM-CL before the RC. This structure differs from our DCL-RCs tested in terms of discourse-functional properties (Chen et al. 2015; Ming & Chen, 2010).
corpus (Chang, 1998), Tong corpus (Deng & Yip, 2018) and Zhou 1 (Zhou, 2001) & Zhou 2 (Li & Zhou, 2004). Our target level of analysis targeted ‘general RC-like’ sequences (following Vasishth et al. 2013). These structures were ‘RC-like because they had the same surface structure as Mandarin subject RC [V-N-de-(DEM)-(CL)-(N)] and object RC [N-V-de-(DEM)-(CL)-(N)], but due to the plurifunctionality of de as a modification marker may not be RCs. Thus they include both genuine RCs and sentences that have the same surface structure. This is the relevant grain of frequency because the identity of the structure is not clear until the sentence-final head noun. Table 1 lists the structural frequencies of subject and object RC-like sequences for both DE and DCL structures.

[insert Table 1 about here]

Overall, DE RC-like sequences were far more frequent than DCL RC-like sequences (2024 vs 37 tokens). Consistent with Vasishth et al.’s (2013) corpus study, subject RC-like DE sequences were 2.5 times more frequent than object-like DE RC sequences. For DCL RC-like sequences we see the opposite pattern: object RC-like sequences were 8.25 times more frequent than subject-like RC sequences.

The experience-based account predicts that children’s parsing decisions will be predicted by corpus frequencies. This makes the unique prediction that the children will have a subject preference for DE RCs, but an object preference for DEM-CL RCs. In contrast, competing theories make uniform predictions for the both structures. The structure-based approach predicts a subject advantage, under the assumption that universal parsing machinery prefers to relativise on subjects (Friedmann et al., 2009; Lin & Bever, 2006; Wagers, Borja, & Chung, 2018). In contrast, linear-based approaches such as Gibson’s (2000) DLT an object advantage.
Study 2: Online Study of Developmental Processing Preferences

We tested the predictions of competing predictions of the three approaches in an experimental study assessing the online comprehension of DE and DCL RCs by young Mandarin-speaking children.

Method

Participants

Thirty-six monolingual Mandarin-speaking children were recruited from a kindergarten in Beijing, China and aged from 4;3 to 4;9 (Mean=4;6, SD=0;1). Children were tested on both DCL and DE RC construction types in a within-subjects design. Since we were interested in children’s online sentence processing when they correctly interpreted the RC, children whose accuracy was too low to offer an accurate record of their eye movements were excluded. Following our similar work on Cantonese (Chan et al., 2018), the inclusion criterion was set to 50% overall comprehension accuracy for each sentence condition. As such, thirteen children who did not score over 50% accuracy on both RC construction types were excluded. Two additional children did not score over the threshold for one RC construction type each, one from the DCL condition and one from the DE RC condition; their data was included for the condition in which they scored above the threshold. The final sample consisted of twenty-three (N=23) children for each RC construction type. All participants were typically developing with no known language impairments (see supplementary materials S2 for comparison of included and excluded children).

Materials

Sixteen test sentences were constructed: eight in the DE condition and another eight in the DCL condition, with four subject RCs and four object RCs for each condition. The test sentences used animal names (e.g. dog, lion, zebra, bear, pig, monkey, cow, tiger, elephant, giraffe, horse, sheep, panda), and transitive verbs (e.g. chase, kick, wipe, tickle, lick, bump,
Online processing of Mandarin relative clauses

*bite, push, touch, feed* that are familiar to young children. All the sentences were pre-recorded by a female native Mandarin speaker (see Appendix for the test sentences). A SONY HDR-CX580VE digital camera was used to record children’s eye-movements.

**Procedure**

*Referent selection task*

The testing and data coding procedures followed those established in Chan et al. (2018). Four toy animals were placed on the table equidistant from a central camera (see Figure 1a). In each case there were two tokens of the head referent (e.g., two dogs, as in sentence 5) and two additional toys that served other roles in the experimental items (i.e., as referents for the RC-internal noun, or the unrelated distractor). A smiley face sticker was placed at the centre of the table just below the camera. Before placing each animal toy on the table, the experimenter asked the child to name each animal character to ensure that the child knew their names.

[insert Figure 1a and 1b about here]

Two background scenes consisting of one target scene [see (5a)] and one distractor scene [see (5b)] were presented to provide a felicitous discourse context for using a restrictive RC (5c) (Corrêa, 1995; Hamburger & Crain, 1982). The experimenter acted out each of the background scenes and then returned the animals back to their original positions before the next sentence played, as shown in Figure 1b.

(5) a. Ni3 kan4! Zhe4 zhi1 xiao3gou3 zai4 ti1 zhe4 zhi1 xiao3zhu1
    You look! This CL dog PROG kick this CL pig
    ‘Look! This dog is kicking the pig.’

b. Yi2! Ling4wai4 yi1 zhi1 xiao3gou3 zai4 tian3 zhe4 zhi1 xiao3zhu1
EXCL another one CL dog PROG lick this CL pig
‘The other dog is licking the pig.’\(^2\)

Attention getter:

Xian4z4ai qing3kan4yi1xia4 zhong1jian1 na4 ge4 xiao4lian3
Now please look-at centre that CL smiley-face
‘Now please look at the smiley face in the centre.’

c. Ni 3 ke3-bu4-ke3yi3 na2qi3
You can-not-can pick-up
#gang1cai2 ti1 xiao3zhu1 de na4 zhi1 xiao3gou3 ya
just-now kick pig de that CL dog SFP
‘Can you pick up # the dog that just kicked the pig?’
(#: pause)

An attention getter (‘now please look at the smiley face in the centre’) was played after the background scenes, directing the children’s attention away from the toys so that they were not biased by perceptual information when the test sentence began (e.g., by looking at one token of the head referent). Across trials the presentation of the target and the distractor animals in the background scenes was counterbalanced, and the location of the toys was pseudorandomized. From the child’s perspective, the target and the distractor were put horizontally or diagonally, but never put along the same vertical plane, which served to facilitate the accurate coding of eye-movements. Two practice trials that did not contain RCs were used to familiarize children with the experimental procedure. The entire experiment lasted approximately 25 minutes per child.

\(^2\) Note: EXCL = exclamative (e.g., ‘wow!’), PROG = progressive aspect marker.
**Eye-movement coding**

Children’s eye movements were recorded by the camera placed under the table. The eye movements were coded frame-by-frame using the visual editing program Sound Forge© (Magix Software GmbH). Looks to each of the four toys were recorded for each 40ms frame. Coding started from the onset of the first syllable of the RC and continued until 2400ms post RC onset. Data from six children (27.3% of final sample) were re-coded by a second trained coder for inter-coder reliability across the two sentence conditions. The results suggested excellent reliability (DE: \( r_s = .919, p < .001, k = .917 \); DCL: \( r_s = .933, p < .001, k = .933 \)).

**Results**

Children’s offline accuracy was above chance (> 60% for all conditions, where chance = 25%), with no significant differences across conditions (see Supplementary Materials S3).

**Eye-movement data**

We only analysed eye-movement data for those trials during which children chose the correct referent. Figure 2 depicts the proportion of looks to the target referent for DCL and DE RCs relative to the total looks to all four referents (chance looking to the target would be 0.25). To determine how the children’s preferences changed as a result of hearing the sentence, we subtracted the target codes at the start of the RC from the rest of the target codes. For DE RCs there were more fixations to the target during subject in comparison to object RCs after the offset of the head noun. The DE RCs show the opposite pattern.

[insert Figure 2 about here]
The data were analysed using a non-parametric permutation test, which follows a three-step procedure (see Chan et al., 2018 for a detailed tutorial). First, a series of linear regressions were conducted for every 40 ms time bin to predict looks to the target with extraction and RC construction type crossed. This yields the typical statistical information (e.g., beta coefficient, p-value) for the main effects of extraction and RC construction type, and an interaction for each time bin. We test here for the significance of the interaction because this is the most likely effect suggested by Figure 2 as well as similar past research on Cantonese (Chan et al., 2018). These interaction p-values are depicted at -0.1 on the y-axis of Figure 2; the significant time bins are black bars that rise above -0.1. The next step was to cluster the time bins based on whether the interaction p-value term in adjacent time windows was less than 0.05. This encodes the idea that effects in adjacent windows are not independent and hence should be treated as one processing effect. There is one cluster with significant interaction p-values starting from 2000 until 2400. The final step was to create a permutation distribution. To do this, we randomly permuted the extraction and RC construction type labels in the cluster, so that the original looking time is now randomly paired with extraction and RC construction type labels. We fit a regression model with extraction and RC construction type crossed on this randomized data and extracted the interaction t-value term. This procedure was repeated 1000 times, which yielded an exact distribution of sum t-values for the cluster, representing how likely this cluster would occur by chance. This revealed a significant interaction of extraction site and RC construction type for the cluster between 2000 ms and 2400 ms (sum t = 22.224, p<.001, shown as grey region on Figure 2).

3 For a justification of the use of the permutation test, and a comparison with other commonly used analytical approaches, see Supplementary Materials S4.
To understand this interaction, we created a separate permutation distribution for each RC construction type (DCL, DE). For DE RC constructions, subject extraction yielded more looks than object extraction in 2000-2400ms window ($t = 19.7, p < 0.001$); the opposite effect was observed for DCL RC constructions in the same window ($t=11.5, p < 0.001$). Overall, the permutation analysis yielded a single significant window, which provides clear evidence of an interaction of RC construction type and extraction site after the offset of the sentences in all conditions.

**Discussion**

Across two studies we found an asymmetry in the distribution of DCL and DE RC-like structures that matched 4-year-old monolingual children’s on-line processing of subject- and object-extracted DCL and DE RC constructions. The result is consistent with experience-based accounts of parsing (e.g., Fitz et al., 2011; Levy, 2008; MacDonald, 2013; Vasishth et al., 2013), but is inconsistent with accounts that predict a uniform pattern of responding across both RC constructions (Diessel, 2007; Gibson, 2000; Friedmann et al., 2009; Lin & Bever, 2006).

One question concerns which experience-based models are most compatible with the results? Many experience-based models use a pre-existing grammar and a corpus tagged by adult linguists to assign probabilities to structures (e.g., Levy 2008, Yun, Chen, Hunter, Whitman & Hale, 2015). These models would not be able to explain our results if they use a grammar that does not distinguish DE and DCL structures (e.g., as in Chen, Grove, & Hale, 2012). There is evidence to suggest that the two structures are at least partially distinct (Chen et al., 2015; Cheung & Li, 2015). Thus there are motivations for treating them differently, yet this can only happen if the experience-based model has an acquisition mechanism that can learn this distinction.
In contrast to models that implement existing grammars, connectionist models can learn typologically-different grammars (Hsiao & MacDonald, 2013) and encode statistical information about discourse functions and sentence structures (e.g., given/new, Fitz & Chang, 2017). These models also have production biases (Chang, 2009), which could provide an underlying basis for the frequency distribution that is encoded by the comprehension system (e.g., ERPs, Fitz & Chang, 2019). This approach instantiates the Production-Distribution-Comprehension account (MacDonald, 2013; Gennari & MacDonald, 2009), where frequency patterns in the input arise ultimately from ease of expressing different meanings within the constraints of the grammar and production system.

The current paper highlights the value of psycholinguistic research across lesser-studied languages and participant groups, particularly when those languages can help tease apart competing theoretical predictions. Our study contains many of the limitations typical of developmental research, including the likelihood of large variability amongst participants (Kidd, Donnelly, & Christiansen, 2018) and the use of fewer items and comparatively small sample sizes (Bergmann et al., 2018). Thus future research corroborating the findings is necessary to further gauge the role of the input in shaping the nature of language comprehension.
Acknowledgements

This research was supported by a research grant from the Hong Kong Polytechnic University to Chan (G-YBF9) and by an Australian Research Council grant to Kidd (CE140100041).

We thank Silvia Gennari, Shravan Vasishth, and two anonymous reviewers for helpful comments. Annotated R code for our analyses can be found at

https://github.com/franklinr/PermutationAnalysis
Appendix

Can you pick up [relative clause] head noun?

Subject DE RCs

1. 舔　斑马　的　狮子
   tian3 ban1ma3 DE shi1zi
   lick zebra DE lion
   ‘the lion that licked the zebra’

2. 撞　狗熊　的　老虎
   zhuang4 gou3xiong2 DE lao3hu3
   bump bear DE tiger
   ‘the tiger that bumped the bear’

3. 咬　小牛　的小象
   yao3 xiao3niu2 DE xiao3xiang4
   bite cow DE elephant
   ‘the elephant that bit the cow’

4. 推　长颈鹿　的　老虎
   tui1 chang2jing3lu4 DE lao3hu3
   push giraffe DE tiger
   ‘the tiger that pushed the giraffe’
Object DE RCs

1. 熊猫 舔的狮子
   
xiong2mao1 tian3 DE shi1zi
   
panda lick DE lion
   ‘the lion that the panda licked’

2. 大象 追的老虎
   
da4xiang4 zhu1i1 DE lao3hu3
   
elephant chase DE tiger
   ‘the tiger that the elephant chased’

3. 小猪 踢的小牛
   
xiao3zhu1 ti1 DE xiao3niu2
   
pig kick DE cow
   ‘the cow that the pig kicked’

4. 大象 撞的长颈鹿
   
da4xiang4 zhuang4 DE chang2jing3lu4
   
elephant bump DE giraffe
   ‘the giraffe that the elephant bumped’
Subject DCL RCs

1. 追 小狮子 的 那只小狗
   zhui1 xiao3shi1zi DE na4 zhi1 xiao3gou3
   chase lion DE that CL dog
   ‘the dog that chased the lion’

2. 踢 斑马 的 那只狗熊
   ti1 ban1ma3 DE na4 zhi1 gou3xiong2
   kick zebra DE that CL bear
   ‘the bear that kicked the zebra’

3. 擦 小猪 的 那只小猴
   ca1 xiao3zhu1 DE na4 zhi1 xiao3hou2
   wipe pig DE that CL monkey
   ‘the monkey that wiped the pig’

4. 挠 小猴 的 那只小牛
   nao2 xiao3hou2 DE na4 zhi1 xiao3niu2
   tickle monkey DE that CL cow
   ‘the cow that tickled the monkey’
Object DCL RCs

1. 小马 推的那只小狗
   xiao3ma3 tui1 DE na4 zhi1 xiao3gou3
   horse push DE that CL dog
   ‘the dog that the horse pushed’

2. 老虎 咬的那只狗熊
   lao3hu3 yao3 DE na4 zhi1 gou3xiong2
   tiger bite DE that CL bear
   ‘the bear that the tiger bit’

3. 小羊 摸的那只小猴
   xiao3yang2 mo1 DE na4 zhi1 xiao3hou2
   sheep touch DE that CL monkey
   ‘the monkey that the sheep touched’

4. 老虎 喂的那只牛
   lao3hu3 wei4 DE na4 zhi1 niu2
   tiger feed DE that CL cow
   ‘the cow that the tiger fed’
References


Supplementary Materials (Yang, Chan, Chang, & Kidd)

S1. Additional analyses of corpus data.

An anonymous reviewer rightly pointed out that one additional cue that influences RC acquisition and processing is head noun animacy (e.g., Kidd, Brandt, Lieven, & Tomasello, 2007; Mak, Vonk, & Schriefers, 2002). While animacy was not the focus of our research, we report here an analysis of noun animacy in a subset of our corpus data. Specifically, we coded both the head noun and the non-head NP for animacy (for RCs, the RC-internal NP, for RC-like structures, the non-head NP) in Mandarin adult child-directed speech from the following six corpora: AcadLang corpus (Zhou doi:10.21415/T5SC9D), Chang1 & Chang 2 corpus (Chang, 1998), Tong corpus (Deng & Yip, 2018) and Zhou 1 (Zhou, 2001) & Zhou 2 (Li & Zhou, 2004).

In Mandarin the head noun can be either expressed or null. Here we only analyse the animacy of expressed nouns since the animacy of unexpressed nouns is often difficult to determine from written transcripts. Table S1.1 reports the distribution of animate and inanimate nouns in general RC-like sequences; Table S1.2 reports the same distributions for genuine RC constructions only.

General RC-like sequences

Table S1.1
Animacy of the head and non-head NPs in subject and object RC-like sequences for DE and DCL construction types

<table>
<thead>
<tr>
<th></th>
<th>Non-head NP (animate)</th>
<th>Non-head NP (inanimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Animate Head</td>
<td>Inanimate Head</td>
</tr>
<tr>
<td>DE</td>
<td>SRC-like</td>
<td>3.7% (21/567)</td>
</tr>
<tr>
<td></td>
<td>ORC-like</td>
<td>2.8% (8/289)</td>
</tr>
<tr>
<td>DCL</td>
<td>SRC-like</td>
<td>0% (0/4)</td>
</tr>
<tr>
<td></td>
<td>ORC-like</td>
<td>0% (0/16)</td>
</tr>
</tbody>
</table>

RC structures
Table S2. Animacy of head noun and RC-internal NP in subject and object RCs for DE and DCL construction types

<table>
<thead>
<tr>
<th></th>
<th>RC-internal NP (animate)</th>
<th>RC-internal NP (inanimate)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Animate Head</td>
<td>Inanimate Head</td>
</tr>
<tr>
<td><strong>DE</strong> SRC</td>
<td>0% (0/79)</td>
<td>6.3% (5/79)</td>
</tr>
<tr>
<td><strong>ORC</strong></td>
<td>3.7% (8/215)</td>
<td>74.4% (160/215)</td>
</tr>
<tr>
<td><strong>DCL</strong> SRC</td>
<td>0% (0/0)</td>
<td>0% (0/0)</td>
</tr>
<tr>
<td><strong>ORC</strong></td>
<td>0% (0/14)</td>
<td>78.6% (11/14)</td>
</tr>
</tbody>
</table>

On the basis of the corpus data we make three crucial observations. Firstly, object RC-like and object RC structures have predominantly inanimate head nouns that contrast in animacy with the other NP, which is typically animate (about three-quarters of them: 73-78%). Thus, consistent animacy contrast cues are available to aid their interpretation in naturalistic speech. Secondly, animacy contrast cues less sharply define subject RC-like and subject RC structures; animacy configurations are more evenly distributed. Finally, in the DCL construction subject RC-like sequences are extremely rare, and subject RC structures are completely absent.

These data provide further support for the influence of children’s linguistic experience on their experimental performance. All nouns in the experiment were nominally animate (i.e., animals as represented by toy referents), and thus, all things being equal, the fact that the test sentences contained animate head nouns and no contrasting animacy cues should favour subject RCs analyses *in those structures where children have evidence for the analysis*; that is, in DE RCs. However, children rarely encounter a subject DCL RC or RC-like structure that follows a subject modifying pattern. Thus, while we should expect that children may experience difficulty with both subject and object DCL RCs with no animacy cues, we should expect to see greater difficulty with subject DCL RCs. This is the pattern of results we observed.
S2. Comparing the included to the excluded children.

Figure S2 compares the proportion of correct responses and head errors by RC construction type and extraction for the included and excluded children.

Figure S2. Proportion of correct responses (left panel) and head errors (right panel) by group (Included vs Excluded), structure (DCL versus DE) and extraction (subject versus object).

A head error occurs when children choose the RC noun as the head noun; when the errors occur in object RCs it suggests that children are pursuing a simple SVO transitive clause interpretation of the sentence. Consider sentence (1), a Mandarin DE object RC.

(1) [lao3shu3 qin1 __ ] de gong1ji1.

mouse    kiss    RL chicken.

‘The chicken that the mouse kisses’.

In this sentence, a head error constitutes choosing the mouse as the referent rather than the chicken, which preserves the thematic agent-patient role assignment of the structure but fails to indicate the child understood the sentence as a head-final RC structure. Children acquiring Chinese language tend to make many head errors in object RCs (Hu, Gavarro, Vernice, & Guasti, 2016; Kidd, Chan, & Chui, 2015; Tsoi, Yang, Chan, & Kidd, 2019).
We analysed the children’s performance on the task with group (i.e., included vs. excluded children) as a between-participants fixed effect using generalized linear mixed models (the lme4 package for Linear Mixed Effects, Bates & Maechler, 2010 in R version 3.5.2; R Core Development Team, 2018). For the correct responses, the final included significant fixed effects of extraction ($\beta = -1.63$, $se(\beta) = .37$, $z = -4.46$, $p < .001$), group ($\beta = .86$, $se(\beta) = .26$, $z = 3.33$, $p < .001$), and a significant extraction by group interaction ($\beta = 1.17$, $se(\beta) = .4$, $z = 2.94$, $p = .003$), and random intercepts for participants and items, and a random slope for group over items. The significant interaction was driven by the excluded group’s large subject-object asymmetry, or in other words, their comparatively poorer performance on the object RCs.

The excluded children’s poor performance on object RCs was almost entirely explained by their tendency to make head errors; that is, to choose the RC subject as the target referent (e.g., choosing the lion in the bear that the lion pushed). We analysed the children’s head errors in the same manner as the correct responses. The final included a significant fixed effect of extraction ($\beta = 5.57$, $se(\beta) = .86$, $z = 6.51$, $p < .001$) and a significant extraction by group interaction ($\beta = -2.38$, $se(\beta) = 1.13$, $z = -2.1$, $p = .036$), and random intercepts for participants and items, and a random slope for group over items. In a mirror image to the correct responses, the excluded children made significantly more head errors on object RCs than the included children.

We interpret these results to indicate that the sole difference between the included and excluded children was the latter group’s tendency to pursue a simple main clause analysis of object RCs. There are several reasons as to why this might be the case, such as a failure to inhibit a more frequent (SVO transitive) structure (Woodard, Pozzan, & Trueswell, 2016) or a general preference for building minimal structure (Frazier, 1987; Gibson, 2000). Since the
included children’s data suggest frequency-sensitive syntactic processing, the former explanation may be more likely.
S3. Analysis of offline (accuracy) responses

Children were considered to have correctly interpreted the test sentence if they selected the correct token of the head referent. These offline responses were analyzed using generalized linear mixed effects models (the lme4 package for Linear Mixed Effects, Bates & Maechler, 2010 in R version 3.5.2; R Core Development Team, 2018). The fixed effects were: (i) sentence type (DE versus DCL), (ii) extraction (subject versus object), and (iii) their interaction. The random effects were participants and items. The maximal model had random slope for extraction under participants. Figure S3 presents children’s offline correct responses to the two types of RCs (standard error bars were created by computing standard error after removing the random effects, Hohenstein & Kliegl, 2013). Overall, children were above chance at selecting the correct referent, intercept $\beta = 0.78$, $z=5.7$, $p < 0.001$. There were no main effects or interactions of sentence type and extraction. The accuracy of subject RCs was numerically higher than object RCs in both DE and DCL conditions, but the differences for the final included children were not significant.

![Figure S3. Accuracy of offline responses for subject and object DCL and DE RCs](image-url)
S4. Comparison of permutation analyses and mixed models.

The standard approach to analysing eye-movement data is to use linear models like mixed models or ANOVA. This assumes that the DV is independent at each time sample, but typically this assumption is violated because there is often a high correlation between the behaviour at time \( t \) and \( t+1 \) (e.g., in our dataset, the correlation between target views on adjacent 40ms time windows is 0.93). One way to reduce this problem is to aggregate the data over time windows. Below we use a common approach, which is to aggregate over 200ms windows, but even with this choice we have a correlation of 0.74 between adjacent windows. Another approach would be to use autocorrelation analysis (\( acf \)) to find larger windows that are not correlated. However, there is still the problem that these windows are not necessarily aligned with the linguistic stimuli and may not reflect genuine parsing events, with the possibility that the same processing effect is split across two windows or that unnecessarily large windows contain genuine effects but are contaminated by unrelated noisy regions. Often researchers choose analysis windows based on prior research, but for studies like ours, for which there is no established research literature, this is not possible. To show these limitations of standard analytic approaches, we present two analyses using mixed models.

The first mixed model analysis was applied to all of the data from RC onset. First, the sum of fixations to the target referent relative to the other toys were calculated and binned into 200 millisecond windows, ranging from RC onset to 2400ms, for each participant across extraction type (subject RC, object RC) for both types of RC construction (DE, DCL). The sum of target fixations were transformed with the empirical logit.

\[
elogit(p, n) = \log\left(\frac{p + 0.5}{n - p + 0.5}\right)
\]
A linear mixed model was fit with a factorial combination of the fixed effects of window, extraction type, and RC construction type (all centred). The maximal model contained random intercept for subjects with no random slopes, as well as a random intercept for items with no random slopes (Barr, Levy, Scheepers, & Tily, 2013). There was a main effect of window ($\beta=0.22$, $SE=0.01$, $\chi^2(1)=430.15$, $p<0.001$) and an interaction of extraction type and RC type ($\beta=0.47$, $SE=0.16$, $\chi^2(1)=8.87$, $p=0.0029$). Crucially, the three-way window X structure X extraction interaction, the equivalent of which was significant in our permutation analysis, did not meet conventional significance levels ($\beta=0.079$, $SE=0.04$, $\chi^2(1)=3.62$, $p=0.057$). To illustrate further limitations of the mixed model approach, we conducted posthoc tests to understand the three-way interaction. Since every posthoc is an additional test, these tests give you an extra chance of finding a significant effect. Therefore, it is necessary to adjust for the number of comparisons, which we did automatically using the multcomp library (Hothorn, Bretz, & Westfall, 2008). When we did this analysis, the effect of extraction type at region 11 ($z=2.35$, $p = 0.0188$) was significant. Thus, while the permutation analysis found a significant interaction of extraction type and RC type from 2000-2400ms, the mixed model only found an effect at 2200-2400 ms. This is in part because this 400ms window was divided into two windows in the mixed model analysis and these were treated as independent effects. Finally, we point out that in this procedure we only found significant posthoc effects for the DE RC construction. This is in part because the posthocs required adjustments for multiple comparisons and this makes it more difficult to detect a significant effect.

The second mixed model analysis was the same except we began analysing the data from the first possible disambiguation point - the offset of the *de* in each sentence type. The empirical logit was equated at the offset of *de* to allow us to see how participants responded to the disambiguating input. The maximal model had a random slope for extraction type for
participants and no slopes for items. There was a main effect of window, $\beta=0.24$, SE=0.03, $\chi^2(1)=72.92$, $p<0.001$, but no other effects were significant. This model assumes that processing behavior is purely bottom up based on the input. In contrast, the previous mixed model included all of the data from RC onset and that analysis also takes into account the predictions and structural preferences that occur before the $de$. Since Chinese RCs are head-final, the listener accumulates a lot of important information prior to $de$ that will no doubt inform parsing decisions, and this is reflected in the different outcomes of the two analyses.

In summary, the use of linear mixed models to analyse our data was less ideal than using the permutation test because (i) the mixed models analysis violates crucial assumptions of the test, (ii) analysis windows place arbitrary limits on processing effects, and (iii) multiple comparisons require $p$-value adjustments. The non-parametric permutation test overcomes both the artificial selection of time windows and problems with dependent observations. Furthermore, there are no posthoc comparisons that must be adjusted for multiple comparisons. As such, it is better suited to the analysis of time-course data yielded in eye-tracking and EEG studies (see Groppe, Urbach & Kutas, 2011 for an overview; Maris, 2012; Maris & Oostenveld, 2007; Eklund, Nichols, & Knutsson, 2016). Specifically, instead of assuming windows a priori, the permutation test creates clusters based on adjacent time points with significant effects and we assume that these clusters represent a processing component (e.g., comprehension of the RC which cause participants to look at the target). In the cluster, the individual time points are correlated, but they are collapsed together into one sum-$t$ value for the whole cluster. The significance of the cluster is tested by randomly permutating the labels for that cluster to create an exact permutation distribution, which represents how likely it would occur by chance. See Chan et al. (2018) for a detailed tutorial about the approach.
References


https://doi.org/10.1017/S0142716419000079

https://doi.org/10.1016/j.jecp.2015.08.005


Zhou, J. AcadLang Corpus. doi:10.21415/T5SC9D. Retrieved from  
### Tables & Figures

**Table 1.** Frequencies of DE and DCL RC-like Sequences Attested in Child-Directed Speech.

<table>
<thead>
<tr>
<th></th>
<th>Subject RC-like</th>
<th>Object RC-like</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE</td>
<td>1456</td>
<td>568</td>
</tr>
<tr>
<td>DCL</td>
<td>4</td>
<td>33</td>
</tr>
</tbody>
</table>
Figure 1a. The layout of the toy props and the hidden digital camera to capture eye movements of the participant in the visual world eye-tracking task.

Figure 1b. Experimenter acting out the background sentences and played the recorded test sentences. E.g.: ‘The dog is kicking the pig. The other dog is licking the pig. Can you pick up the dog that just kicked the pig?’
Figure 2. Average target proportions of looks for the DE (top panel) and the DCL (bottom panel) RCs are shown by solid/dashed lines. Onsets/offsets for different sentence units are shown by the size of the rectangles at the top left (solid for subject RCs, dashed for object RCs). Small grey/black bars near -0.1 are \(p\)-values for individual time bins. The large shaded grey bars represent the time-windows identified by the permutation analysis as significant.