Four-year-old Cantonese-speaking children’s online processing of relative clauses: a permutation analysis*

ANGEL CHAN
The Hong Kong Polytechnic University, Hong Kong, and The Hong Kong Polytechnic University – Peking University Research Centre on Chinese Linguistics

WENCHUN YANG
The Hong Kong Polytechnic University, Hong Kong

FRANKLIN CHANG
University of Liverpool, United Kingdom, ARC Centre of Excellence for the Dynamics of Language, Australia, and ESRC International Centre for Language and Communicative Development (LuCiD)

AND

EVAN KIDD
The Australian National University, Australia, ARC Centre of Excellence for the Dynamics of Language, Australia, and ESRC International Centre for Language and Communicative Development (LuCiD)

(Received 22 May 2016 – Revised 29 January 2017 – Accepted 21 April 2017)

ABSTRACT
We report on an eye-tracking study that investigated four-year-old Cantonese-speaking children’s online processing of subject and object
relative clauses (RCs). Children’s eye-movements were recorded as they listened to RC structures identifying a unique referent (e.g. “Can you pick up the horse that pushed the pig?”). Two RC types, classifier (CL) and $ge_3$ RCs, were tested in a between-participants design. The two RC types differ in their syntactic analyses and frequency of occurrence, providing an important point of comparison for theories of RC acquisition and processing. A permutation analysis showed that the two structures were processed differently: CL RCs showed a significant object-over-subject advantage, whereas $ge_3$ RCs showed the opposite effect. This study shows that children can have different preferences even for two very similar RC structures within the same language, suggesting that syntactic processing preferences are shaped by the unique features of particular constructions both within and across different linguistic typologies.

INTRODUCTION

The acquisition and processing of relative clauses (RCs) has received an enormous amount of interest in the psycholinguistic literature (e.g. Gibson, 1998; Kidd, 2011; MacDonald, 2013). This focus has been driven by the fact that RC processing is assumed to largely reflect syntactic processes (as opposed to other processes like ambiguity resolution, though see Gennari & MacDonald, 2008; Hale, 2006; Yun, Chen, Hunter, Whitman & Hale, 2015). Most research to date has been conducted on languages like English, German, and Hebrew, which are right-branching and have post-nominal (i.e. head-initial) RCs. The general conclusion from this body of research is that, with some qualifications (e.g. Diessel & Tomasello, 2000; Kidd, Brandt, Lieven & Tomasello, 2007), subject RCs like (1) are typically acquired earlier and are easier to process than object RCs like (2).

(1) The dog [that bit _ the bear].
(2) The dog [that the bear bit _].

Several explanations for this asymmetry exist. For instance, drawing upon formal grammatical theory, structurally oriented theories attribute the difference to the fact that object RCs are hierarchically more complex than subject RCs (e.g. Frazier, 1987; Friedmann, Belletti & Rizzi, 2009). Specifically, these models follow Chomskyan syntactic theory in assuming greater hierarchical distance between the head noun and the gap in object RCs in comparison to subject RCs. However, formal explanations fail to explain why object RCs are not more difficult than subject RCs in all instances (Kidd et al., 2007; Traxler, Morris & Seely, 2002), why processing difficulty can be attenuated following increases in exposure
through priming (Hutton & Kidd, 2011; Wells, Christiansen, Race, Acheson & MacDonald, 2009), or why difficulties in online processing of object RCs are not exclusively syntactic (Bornkessel-Schlesewsky & Schlesewsky, 2009; Weckerly & Kutas, 1999).

Several alternative models do not assume that the parser builds hierarchical phrase structure trees from which meaning is read, instead assuming a linear left-to-right parsing system which, depending on the specific theory, attributes the difficulty associated with object RCs to different phenomena. These include (i) differences in the linear distance between filler and gap (greater for object RCs; Gibson, 1998), (ii) differences in frequency (object RCs often contain infrequent non-canonical word orders and rarely occur with two animate NPs; Ambridge, Kidd, Rowland & Theakston, 2015; MacDonald & Christiansen, 2002), and (iii) cross-linguistic tendencies favouring relativization on subject over direct object NPs (Keenan & Comrie, 1977). While explaining many of the effects that formal accounts fail to explain, no single explanation appears to cover the full range of empirical facts (for a good discussion see Kim & O’Grady, 2016).

Therefore, despite intense research over several decades, a comprehensive account of RC acquisition and processing has remained elusive. Recently, researchers have begun to test these competing theories in typologically different languages that provide opportunities to tease apart predictions of theories in ways that investigating most European languages do not allow. For instance, studies of Basque have revealed a processing advantage in comprehension for object RCs in both children and adults (Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía & Laka, 2010; Gutierrez-Mangado, 2011). Data from other highly inflected languages show no subject–object asymmetry (e.g. Finnish: Kirjavainen, Kidd & Lieven, 2017; Kirjavainen & Lieven, 2011; Quechua: Courtney, 2006). However, most attention has been focused on East Asian languages such as Chinese (Mandarin and Cantonese), Japanese, and Korean, to which we now turn.

**RC acquisition in East Asian Languages**

RCs in East Asian languages are typologically very different from RCs in well-studied languages like English and German. Most notably, they are prenominal, such that the RC is placed before the head noun, as in the Japanese object RC example in (3) (from Ozeki, 2011).

(3) [papa kara moratta] yatu
Dad from received one
‘The one [I was given by Dad].’
The languages for which we have the most data are Japanese and Mandarin. In both cases the acquisition data are mixed and do not point to a uniform pattern across languages. In Japanese, the data suggest no subject advantage in acquisition. In a longitudinal study of children’s spontaneous speech, Ozeki and Shirai (2007; see also Ozeki, 2011) found that Japanese-speaking children produce subject, object, and oblique RCs at approximately the same rate from the onset of production, and their functions are very different from what has been described in well-studied languages like English and German (Brandt, Diessel & Tomasello, 2008; Diessel & Tomasello, 2000). In experimental work, Suzuki (2011) found a significant object advantage in five-year-old children, but once children’s knowledge of case marking was controlled, the difference was not significant. One source of this variability is a preference for shifting heavy RCs early in Japanese (Hakuta, 1981; Hawkins, 2004; Yamashita & Chang, 2001), which could increase the frequency of object RCs compared to languages without this early shifting bias, like English.

The data from Mandarin are more inconsistent still. In acquisition studies of both comprehension and production, both subject and object advantages have been found, in addition to null effects (for a review see Chan, Matthews & Yip, 2011). One potential reason for the inconsistency was discussed by Chan et al. (2011): Chinese (both Mandarin and Cantonese) possesses the typologically rare combination of SVO main clause word order and prenominal RCs (Dryer, 2005). This combination creates competing processing demands based on surface/linear structure and canonical word order, which favour object RCs, and the general prominence of subjects, which favour subject RCs. Consider Mandarin sentences (4) and (5):

(4) [RC __; qíní gōngjíjí] de lǎo2shú3i
    kiss chicken   PRT mouse
    V    O    S
    ‘The mouse that kisses the chicken’

(5) [RC xiǎo3yáng2 tuīi __;] de xiǎo3tú4i
    sheep    push    PRT rabbit
    S    V    O
    ‘The rabbit that the sheep pushes’

Sentence (4) is a subject RC and sentence (5) is an object RC. Mandarin RCs have a particle de that signals the relativized structure (PRT). Subject RCs have non-canonical VOS word order and, in (4), the verb and its object complement separate the head noun (‘mouse’) and the gap. In contrast, object RCs follow canonical word order and the linear distance between head noun (‘rabbit’) and gap is shorter. These features favour object RC processing and appear to significantly affect acquisition. Chen and Shirai (2015) report that object RCs are produced about 60% of the time by
children and adults (compared to 20% subject RCs), suggesting that children learning Mandarin prefer object RCs, and that they occur more frequently in the input. In contrast, the general prominence of subjects in nominative–accusative languages pull in the direction of subject-over-object RCs, a fact which is captured across numerous theoretical traditions in linguistics (e.g. Keenan & Comrie, 1977; O’Grady, 2011; Rizzi, 1990) and psycholinguistics (Bornkessel-Schlesewsky & Schlesewsky, 2009). Therefore, unlike in languages such as English, where all these cues favour subject RC processing, in Chinese the cues compete. This may explain the mixed acquisition results.

Cantonese RC acquisition

The situation is mirrored in Cantonese, with some language-specific differences in RC formation that make it a particularly interesting language to study (Matthews & Yip, 2001). Cantonese has two common relativization strategies, as shown in (6) and (7). Both examples have a third person (3SG) pronoun subject in the RC (keoi5), which is not gender or human specific.

(6) [RC keoi5 gaan2 __] go2 lap1 tong2 i 3SG choose that CL candy.

‘The candy she chooses’.

(7) [RC keoi5 gaan2 __ i] ge3 tong2 i 3SG choose PRT candy.

‘The candy(ies) she chooses’.

Sentence (6) is a classifier RC (henceforth CL RC), so-called because it contains the demonstrative (DEM) go2 and an appropriate classifier (CL) before the head noun (lap1 in (6) is a classifier for grain-like things). CL RCs are commonly used in spoken Cantonese, and are relatively informal in register. A more formal relativization strategy that is similar to the structure in Mandarin Chinese is to mark the RC with the particle ge3, as in (7). Although the two can be used interchangeably in many cases, there is a semantic contrast between them: the CL RC entails specific reference, while the ge3 RC does not, and can be construed as quantifying over a larger set. The ge3 RC, therefore, is also not specified for number: both singular head and plural head readings are possible. For CL RCs with a plural head, for instance ‘the candies she chooses’, the classifier di1 for plural objects and kinds is obligatory.

The two Cantonese RC types also differ in another interesting way that is relevant to acquisition. There is an isomorphism between object classifier RCs and simple main clauses, such that the object classifier RC in (6) is identical in surface form to a SVO main clause, as in (8).
This structural overlap raises the possibility that children bootstrap into the syntax of RCs from their knowledge of simple transitives, predicting an early acquisition advantage for object RCs because children may use their early developed knowledge of canonical sentence patterns to acquire more complex syntactic patterns (Diessel, 2007). Chan et al. (2011) argue that during this process Cantonese-speaking children may analyze object classifier object RCs as internally headed RCs. Thus (6) can be analyzed as (9):

(9) \[ NP/S \text{ keoi}\text{5 gaan2 go2 lap1 tong2} \]
3SG choose that CL candy

‘The candy she chooses’.

Under the internally headed RC analysis, sentence (9) has the internal structure of a SVO clause, but behaves as a NP in terms of its external syntax. The internally headed RC analysis is represented by the notation NP/S in (9) above, indicating a constituent having externally the syntax of a NP but internally that of a clause (S). The internal structure is a SVO main clause, with the object, which is also the head noun, in situ. Hence the head ‘candy’ is internal to the RC. This internally headed analysis is only possible for Cantonese object classifier RCs since it is only in this case where there is complete surface identity with simple clauses and therefore ambiguity of analysis. In contrast, while object ge3 relatives have structural similarity (also SVO) they lack the surface identity relation because of the presence of the relative marker ge3.

Examples like (6) are attested in young children’s naturalistic speech (Yip & Matthews, 2007). These utterances are structurally ambiguous as they can be analyzed as head-final RCs (6) or internally headed RCs (9). Further suggestive evidence for the internally headed RC analysis comes from ill-formed child utterances in naturalistic speech (10) and experimental tasks (11) (examples from Chan et al., 2011; Yip & Matthews, 2007 where SFP is sentence final particle and PROG is progressive aspect marker).

(10) ngo5 sik6 jok6 aa3 \[NP/S ngo5 sik6 jok6\] haai6 leizek3 (Alicia 2;08·10)

I eat medicine SFP I eat medicine is this CL

‘I’m taking medicine. The medicine I take is this one.’

(11) Experimenter (in an elicited imitation task):
The giraffe that the zebra’s kicking hard is here.

Child (pointing to the particular giraffe that the zebra was kicking):

The giraffe that the zebra’s kicking hard is here.

The second clause, ngo5 sik6 joek6, in (10) and the first clause from the child, baan1maa5 daai6 lik6 tek3 gan2 coeng4geng2luk2, in (11) are functionally noun-referring expressions, but structurally ill-formed because the demonstrative go2 plus classifier or the particle ge3 would be required in order to be grammatical. They are, however, consistent with the analysis whereby the children were using SVO clauses [S ‘I take medicine’] and [S ‘zebra’s kicking giraffe forcefully’] as internally headed RCs [NP ‘I take medicine’] and [NP ‘zebra’s kicking giraffe forcefully’] to mean ‘the medicine I take’ and ‘the giraffe that the zebra’s kicking hard’.

To summarize, Cantonese has two relativization strategies, one that is commonly used in spoken discourse (classifier RCs) and the other which is more formal (ge3 RCs). For both RC types there is surface order overlap between object RCs and simple transitive sentences, both containing SVO word order. However, in object CL RCs there is complete surface identity between simple SVO sentences and object RCs, allowing for an internally headed object RC analysis. These specific features of Cantonese lead to some interesting predictions regarding RC acquisition and processing. If, following Chan et al. (2011), simple transitives serve as a path-breaking construction that allow children to bootstrap into the syntax of RCs (à la Abbot-Smith & Behrens, 2006), then we may observe a general object preference across the acquisition of all RC types. Furthermore, this advantage may be more pronounced for classifier RCs because: (i) they are generally more frequent in spoken Cantonese; and (ii) there is complete isomorphism with simple transitives, allowing them to be analyzed as internally headed RCs, which do not involve gaps or extraction, are structurally simpler, and hence may be easier to process, than externally headed RCs.1 Since ge3 RCs are structurally similar to Mandarin de RCs,

---

1 In Korean, for instance, which has both head-final RCs and head-internal RCs attested, both children and adult L2 learners have been shown to acquire head-internal RCs earlier than head-final RCs (Jeon & Kim, 2007).
their performance in Cantonese can help us to understand how the formal structure of these RCs work under different input conditions.

To date there has been very little published research on Cantonese RC acquisition. In a naturalistic study of three bilingual children, Yip and Matthews (2007) reported that all children produced object classifier RCs before or simultaneously with subject RCs. Chan et al. (2011) discussed two unpublished studies investigating comprehension and production of CL RCs only. The comprehension experiment used the picture selection methodology and revealed a significant object advantage. In contrast, the production experiment, which used the sentence imitation method, revealed a numerical but non-significant object advantage. In contrast, Lau (2016) reported on three- to five-year-old Cantonese-speaking children’s comprehension of CL RCs using picture selection, and reported a significant subject advantage. The difference between the Chan et al. and Lau studies was that, whereas Chan et al presented test RC structures in a supportive discourse context (thereby fulfilling the felicity conditions governing RC use; see Corrêa, 1995), the study reported in Lau did not. Lau also presented data from an elicited production experiment, which showed no overwhelming preference for either subject or object RCs.

In the only study that has simultaneously tested both classifier and ge3 RCs, Kidd, Chan, and Chiu (2015) tested monolingual (mean age = 6;3) and Cantonese–English bilingual children (mean age = 8;11, groups matched on Cantonese verbal ability) using picture selection. They reported a non-significant object advantage in the monolingual group for both RC types, and a significant subject advantage for the bilingual group, which was more pronounced for CL RCs (an effect attributed to cross-linguistic influence from English). Therefore, across the small set of naturalistic and experimental studies that have tested monolingual children, the data point to a general although weak object advantage for CL RCs, at least when test structures are presented in a felicitous discourse context, whereas the data for ge3 RCs are too preliminary to draw any firm conclusions.

Current research
In the current paper, we report on a study that investigated four-year-old monolingual Cantonese-speaking children’s online processing of subject and object CL and ge3 RCs. Children’s eye-movements to toy referents were recorded while they heard test sentences containing a RC. There are several advantages to studying children’s online processing. Most broadly, online data reveal complexity effects ‘in the moment’, giving a clearer indication of sentence difficulty than might be observed with offline tasks, where such effects can be obscured by post-interpretative processes. This is important in the context of East Asian RCs, where results comparing
subject and object RCs have been very inconsistent. With respect to Cantonese in particular, online data may reveal differences between the processing of classifier and ge3 RCs that are not evident in offline data (e.g. Kidd et al., 2015).

There are a broad range of theoretical models that make predictions about RC acquisition and processing, and we do not have space here to do all of them justice. We instead contrast theoretical approaches that are relevant to the specific case of Cantonese. First, several models from a variety of traditions predict a universal subject preference across all languages (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2009; Friedmann et al., 2009; Keenan & Comrie, 1977; Vasishth, Chen, Li & Guo, 2013; Yun et al., 2015), and so predict a general subject preference for CL and ge3 RCs. In contrast, several different models predict either a no preference or an object preference. For instance, O’Grady (2011) attributed the aforementioned mixed results for Chinese RCs to a conflict between a general preference to relativize on subjects and differences in filler–gap distances between subject and object RCs, which are shorter in the case of object RCs. The implication is that these two influences on RC interpretation pull in opposite directions, and may therefore neutralize any potential asymmetry. Usage-based approaches predict earlier acquisition of object RCs in Cantonese because of their similarity to simple transitive sentences (Diessel, 2007). For example, Fitz, Chang, and Christiansen (2011) found that substructure similarity between different RC constructions influenced the ease of learning the constructions over development. This predicts a difference between the CL RCs and the ge3 RCs, since simple transitives do not include the ge3 marker.

Method

Participants

Seventy (N = 70) typically developing monolingual four-year-old Cantonese-speaking children were recruited from preschools in Hong Kong. Sentence type was tested between-participants (i.e. CL versus ge3). Since we were interested in children’s online processing of sentences when they correctly interpreted the RC, we excluded children whose accuracy was too low to provide an accurate record of their eye-movements. We set the inclusion criterion to 50% overall comprehension accuracy. For CL RCs this meant that eighteen out of thirty-seven children were excluded, and fifteen of thirty-three children excluded for ge3 RCs (final N = 37). The final sample for the CL condition therefore consisted of nineteen

2 Note that this is our interpretation of O’Grady (2011), who does not explicitly make this prediction.
children aged between 4;3 and 4;9 (Mean = 4;7, SD = 0;2), and for the ge3 condition consisted of eighteen children aged between 4;3 and 4;9 (Mean = 4;7, SD = 0;2). No child possessed any known language and/or cognitive impairment.

Part of the large drop-out rate is no doubt due to the age of the children we tested. Online studies of syntactic processing in children aged younger than five years are still vanishingly rare, but are important because they better capture the intersection between the acquisition of knowledge and the implementation of parsing routines. Therefore, if we are to capture the emergence of structural parsing routines we will inevitably face problems like participant attrition. Our attrition rate (47%) suggests that we are capturing RC processing at an age where there is significant variation amongst Cantonese-speaking children. Therefore, although our results only represent those four-year-old children who have fairly good competence with transitive RC structures, they are likely to reflect online processing of RCs as relatively newly mastered forms.

Materials
Eight CL and eight ge3 relative clause constructions served as target sentences: four subject-extracted and four object-extracted (see Appendices A and B). Each sentence contained common nouns and verbs familiar to the children so as not to confound syntactic processing with lexical processing or gaps in vocabulary. The nouns denoted farm or zoo animals (bear, cow, dog, elephant, giraffe, horse, lion, monkey, panda, pig, tiger, zebra). These mapped onto a set of toy animals which served as referents for the NPs in the target trials. Eight transitive action verbs were used (bite, bump, chase, feed, kick, lick, push, tickle, wipe). A digital camera was used to record children’s eye-movements, which recorded an image every 33 ms.

Procedure
Referent selection task. We used a modified version of Brandt, Kidd, Lieven, and Tomasello’s (2009) referent selection task (see also Rahmany, Marefat & Kidd, 2014). In the task, children are introduced to four animals that are placed on a table in four locations equidistant from a central video camera that protrudes from a hole cut in the table (see Figure 1). There were two experimenters. One monitored the camera to ensure that it recorded the child’s face and also played each prerecorded item from a laptop. The other experimenter was responsible for placing the toy referents in prespecified locations on the table. As each animal was placed on the table, the experimenter elicited the name of each toy from the child to both ensure that the child knew the toy’s label and to
maintain the child’s interest in the task. Typically, children correctly named the toy, but on the rare occasion they provided a label that was different to what was used in the audio the experimenter corrected the child. An example trial is shown in (12).

(12) a. tai2 haai5 nei1 zek3 hung4jan2 teoi1-gan2 nei1 zek3 daai6zoeng6 wo3 look PRT this CL bear push-PROG this CL elephant SFP ‘Look! This bear is pushing the elephant.’

b. ji2 ling6ngo6i6 jat1 zek3 hung4jan2 zau6 tek3-gan2 nei1 zek3 daai6zoeng6 EXCL another one CL bear then kick-PROG this CL elephant ‘The other bear is kicking the elephant.’

ji4gaai1, tai2haa5 go3 haai1haa1siu3 gung1zai2 aai
now look at CL smiley figure SFP ‘Now look at the smiley face.’

c. nei5 ho2-m4-ho2ji5 ling1hei2 you can-not-can pick.up

# tau4 sin1 sek3 daai6zoeng6 go2 zek3 hung4jan2 aai3 just.now kiss elephant that CL bear SFP ‘Can you pick up # the bear that just kissed the elephant?’
(#: pause)
A target trial began with two background scenes ((12a) and (12b)), the function of which was to create a felicitous discourse context in which the RC in the critical sentence (12c) uniquely identified one referent from a set (Corrêa, 1995; Hamburger & Crain, 1982). Therefore, both background scenes described activities in which two tokens of the same type, which in (12) is a bear, were participants in transitive actions with another animal on the table. As the background scenes were played, one experimenter acted out the scenes and returned the animals back to their locations before the next sentence played (target trials were played as one continuous audio file). After the two background scenes, the children heard the attention getter ji4gaa1, tai2hhaa5 go3 ha1hha4siu3 gung1zai2 aar 'now look at the smiley face', which served to divert their attention away from the toy referents to a smiley face sticker in the centre of the table just below the camera. This was important because it meant that children’s subsequent looks to the toy referents while they heard the test sentence (12c) would reflect processing of that sentence rather than perseverative looking attributable to background scenes. The order of mention of the target referent in the background scenes was counterbalanced across trials, with half in the first background scene and half in the second. The location of the toys was pseudo-randomized across trials, with one restriction: the two tokens of the head referent were never placed along the same vertical plane (from the child’s perspective). That is, while there were trials in which the two tokens of the head occurred on the same horizontal plane (either in front of or behind the camera), or diagonally across the line of the camera, they were never placed such that one was directly behind the other. This was because the eye-movements were coded offline (à la Snedeker & Trueswell, 2004), and organizing the toys in this manner ensured more accurate eye-movement coding because looks to the target versus distractor toy required children to make saccades or head movements. The children’s choice of toy referent provided offline indications of their final interpretations of the sentence. The entire experiment lasted approximately 20 minutes per child.

Eye-movement coding. Children’s faces were recorded, which enabled coding of their eye-movements to different locations on the table. The children’s individual recordings were digitised to avi files and were coded using the visual editing program Sound Forge©. The program shows the visual display (i.e. recording of child’s face) and a separate audio track as a wav file. The wav file enables the location of critical points in the target sentences, and the video allows frame-by-frame coding of eye-movements to the four locations on the table. Each frame was 33 ms. Coding began at the beginning of the RC. Since RCs in Cantonese are prenominal, this meant that we coded the entire RC. Although it is possible to identify the
target before hearing the head noun, because it can be predicted from the preceding verb and NP argument, we report looking behaviour until 2400 ms post RC-onset because this is, to our knowledge, the first eye-tracking study of RC processing in any Chinese language, and as such we do not have specific hypotheses regarding the location of any statistical effects in the eye-movement record. The data of three children in each sentence condition (15.8% of final sample for CL condition and 16.67% of final sample for ge3 condition) were re-coded by a second trained coder for inter-coder reliability, which was high (CL: \( r_s = 0.923, p < 0.001; \kappa = 0.937, p < 0.001; \) ge3: \( r_s = 0.945, p < 0.001; \kappa = 0.944, p < 0.001 \)).

**RESULTS**

*Offline responses*

Children’s offline responses give an initial indication of the relative difficulty of each sentence type, which is likely to be reflected in the online looking patterns. Figure 2 shows children’s offline comprehension accuracy for CL and ge3 subject and object RCs.

Figure 2 shows that the children performed similarly on CL subject and object RCs. For ge3 RCs, children performed much better on subject than on object RCs. Response (correct = 1) was predicted using Generalized Linear Mixed Models (GLMM; Jaeger, 2008) using the lme4 package for Linear Mixed Effects (Bates & Maechler, 2010) in R (version 3.2.2; R Core Development Team, 2014). Sentence type (CL versus ge3), extraction (subject versus object), and their interaction were entered as fixed effects. Random effects for participants and items were included and there was a random slope for RC type (Barr, Levy, Scheepers & Tily, 2013).

The results from the mixed model revealed no reliable main effects of either sentence type or RC extraction, and no interaction. This is due to the large variability across children in their ability to correctly respond in each condition, as can be seen by the large standard error bars that overlap with the means of the other conditions. There was, however, a significant intercept (\( \beta = 0.826, \ z = 6.26, p < 0.001 \)), which shows that accuracy at selecting the correct referent was significantly above chance overall.

*Online data*

Standard approaches for analysing eye-tracking data involve dividing the data into separate windows (e.g. 200 ms) and looking for interactions of time-window and experimental conditions. These approaches are most effective when dealing with a population where previous research has shown that differences tend to occur within the windows used in the analysis. These conditions are often met with eye-tracking studies of adult speakers of well-studied languages like English, but they are less likely to
be appropriate when dealing with developmental data in less-studied languages like Cantonese. When it is not known where effects will appear, post-hoc tests with adjustments for multiple comparisons are needed and these will be less sensitive than when appropriate windows are known beforehand. Recently, non-parametric permutation tests have been found to be appropriate for analysis of data where analysis regions were not known a priori (for detailed overviews see Groppe, Urbach & Kutas, 2011; Maris, 2012; Maris & Oostenveld, 2007). Eklund, Nichols, and Knutsson (2016) showed that these techniques yield target familywise error rates of 5% over 3 million random task group analyses of fMRI resting state data, showing that this approach is robust over noisy data. It has also been applied successfully to study noisy data in studies of infant word processing (e.g. Dautriche, Swingley & Christophe, 2015; Von Holzen & Mani, 2012).

Although different theories make distinct predictions regarding the relative complexity of Cantonese subject and object RCs, they do not make predictions regarding the precise temporal location of processing difficulty in the eye-tracking record. While adults can be consistent in the amount of time that they take to process a particular structure, children will vary in the exact location of this difficulty depending on their point in development. Since Cantonese RCs are head-final, and since RC-internal word order differs across subject and object RCs, we wanted to cast a wide net and analyze eye-movements throughout the entire RC and beyond, rather than simply at the disambiguation point (i.e. the head noun). This enabled us to not only identify any differences in processing across different structures, but also identify how word order differences between structures within the RC affect the identification of the head referent.

Fig. 2. Offline comprehension accuracy for subject and object CL and ge3 RCs (standard error bars for data with random effects removed; Hohenstein, 2013).
Before describing our permutation analysis in detail, we first describe the rationale for the analysis. To avoid any assumptions about windows, we use the time bins at the rate provided by the eye-movement coding (i.e. every 33 ms) and we apply a test statistic comparing subject and object RC for each time bin (any test statistic can be used, and here we use the t-test from a regression model). This provides a list of the observed bins with significant subject/object differences ($p < .05$, unadjusted for multiple comparisons). We then cluster adjacent bins with significant test statistics together. This captures the fact that adjacent time-windows are not independent, but rather are likely to reflect a single processing event. For example, if we have a difference at 1000 ms and later at 1066 ms between subject and object RCs, it is likely that this difference is due to same underlying process. In contrast, mixed-models analyses make the incorrect assumption that all datapoints are independent.

The next step is to create a sample of 1000 experiments. For each experiment, we take the data for each time bin, permute the subject/object labels without replacement, and apply the test statistic to predict the actual looking data using the permuted labels. By permuting the labels, we remove any link between the labels and the eye-tracking data, and hence these 1000 tests give us a distribution under the null hypothesis. The left-most panel in Figure 3 shows the 95% confidence interval for the observed data in the $ge_3$ condition at time bin 1914 ms, where there is a strong subject preference (error bars do not overlap). When the labels are permuted, as in the other three panels, then the difference can become weaker (Exp. 1), disappear (Exp. 2), or go in the opposite direction (Exp. 3).

Since each experiment independently permutes each time bin, we need to sum together the results for each experiment for each cluster (we call this the sum t-distribution). Smaller clusters have smaller sum t-values that can increase values in the centre of the distribution, so it is more conservative to use only the largest sum t-value for each experiment in our MAXIMAL sum t-distribution. Finally, we can compute $p$-values by computing the proportion of the maximal sum t-values in the distribution that are greater than the sum t-values for each of the clusters in the observed data. If this proportion is less than .025, then we can conclude that it is significant by a two-tailed test.

Children were tested on all items in the structure type condition to which they were assigned (i.e. either CL or $ge_3$ RCs). However, children are notoriously variable participants and can be affected by individual preferences for stimuli (in our case, toy referents). Therefore, overall looking to the target will vary depending on the participant, the particular sentence being heard, and the sample of toys in the display. This variation works against our goal, which is to understand how the structures that are heard influenced looking behaviour to the target. Hence, we computed the
mean proportions of looks to the target referent at the start of each trial (i.e. beginning of RC) and subtracted this from the looks to the target referent in that trial (we will use **TARGET PROPORTION** to refer to this measure). **Figure 4** shows the mean target proportion averaged across participants and item, with CL RCs shown in the top panel and *ge* RCs shown in the bottom panel. Looks were coded from the onset of the RC (0 ms) for **2400 ms**. The onset and offset of different linguistic units are shown at the top of each figure with solid lines for object RC and dashed lines for subject RCs. The offset of the head noun marks the absolute uniqueness point of each sentence, where the head noun can be unambiguously identified. Note, however, that anticipatory predictive looks are possible because the head-final nature of Cantonese RCs means that the RC comes before the head.

As the first step in the permutation analysis, we applied regressions to each time-window to predict target proportion with subject/object condition (effect coded) and the difference between these *p*-values and .05 are shown in **Figure 4** as bars around −0.1. If the bar extends below −0.1 and is grey, the *p*-value is greater than .05, otherwise, if it is black and above −0.1, it is significant. Adjacent bins were clustered together if they were significant. There was one cluster in the CL study (1716–1815 ms) and four clusters in the *ge3* study (1551–1584 ms, 1617–1650 ms, 1683–2244 ms, and 2310–2442 ms).

In the next step, 1000 experiments were run by permuting condition labels in each significant time bin and applying regression to predict the observed target proportion. Next we produced the sum *t*-distribution by summing the *t*-values produced for each time bin within each cluster (**ALL SUM T**
Fig. 4. Average target proportions for CL (top panel) and ge3 (bottom panel) RCs are shown by solid/dashed lines. Onset/offsets for different units are shown by the size of the rectangles at the top left (solid for subject, dashed for object). Small grey/black bars near 0 are p-values for individual time bins. The large grey bars represent the time-windows identified by the permutation analysis as significant. Curved lines represent 200 ms windows identified as significant by mixed-model post-hoc analysis.

Histograms for CL and ge3 experiments in the top of Figure 5). The dashed lines in Figure 5 show the borders of the band that contains 95% of the sum t-values. Since the ge3 condition has one large cluster and three smaller clusters, there are more datapoints in the centre of the distribution due to the small effects in the smaller clusters. To remove this bias, we select the largest absolute sum t-value for each simulation and place that into the maximal sum t-distribution (bottom two panels in Figure 5). The ge3 maximal sum t-distribution is bimodal, because when we randomly permute and test four clusters, one of the tests will tend to yield a
non-zero sum $t$-value by chance and this value will be the maximal value. The 95% band for the $ge3$ all sum $t$-distribution represents the likelihood of getting a significant effect when four clusters are tested, but we are interested in whether our participants distinguish subject and object RCs, so this one test is better matched by the $ge3$ maximal sum $t$-distribution, which is an exact distribution based on the biggest effect that could occur by chance for our four comparisons.

The maximal sum $t$-distribution for $ge3$ has a bigger band than the distribution for CL because it includes a large cluster of 561 ms, while CL only has one cluster of 99 ms. Since we are interested in whether there is an effect of any length, we will test the observed CL cluster sum $t$-values against the CL distribution, and the observed $ge3$ effect against the $ge3$ distribution. For each of the clusters, $p$-values were the percentage of values in the corresponding distribution that were less than observed sum $t$-values. Unlike traditional linear model approaches (e.g. ANOVA), where theoretical distributions (e.g. normal, $t$) are matched to data by the use of parameters like degrees of freedom, the permutation test is a non-parametric test, because we have computed an exact distribution that takes into account the number of clusters that we are testing, as well as the size and variability of our data.

Fig. 5. Distributions created by permutation distribution for CL and $ge3$ studies with either all sum $t$-values or only maximal sum $t$-values. Dashed lines represent range for 95% of the $t$-values.
The permutation analysis revealed one significant cluster for each structure type. For CL RCs, the children look significantly more at the head referent of object RCs between 1716 ms and 1815 ms post RC onset (total window time = 99 ms, sum $t = 6$, $p < .003$), suggesting a significant object advantage. For ge3 RCs, the children looked significantly more at the head referent of subject RCs between 1683 ms and 2244 ms post RC onset (total window time = 561 ms, sum $t = 43$, $p < .001$), suggesting a significant subject advantage. These significant clusters are denoted in Figure 4 by the long grey shading.

Comparison with linear mixed-effects analysis over prespecified time-windows

Since permutation analyses have not been used extensively for studying child language data, it is worthwhile to compare our permutation analysis with a traditional mixed-model analysis. To do this, we averaged the target proportion for each 200 ms window for each subject in each condition (CL and ge3) for each RC extraction type (subject/object). We then applied a mixed model to the proportion target looks with window, condition (CL and ge3), and RC type (all centred). Subject and items were random effects and the maximal model had random slopes for window and extraction type for both subjects and items. In this analysis, there was a main effect of window ($\beta = 0.052$, $SE = 0.0024$, $\chi^2(1) = 432.3$, $p < .001$), an interaction of window with RC type ($\beta = -0.01$, $SE = 0.0048$, $\chi^2(1) = 3.97$, $p = .046$), and a three-way interaction of window, RC type, and extraction ($\beta = 0.036$, $SE = 0.0097$, $\chi^2(1) = 13.92$, $p < .001$). To explore this three-way interaction, post-hoc comparisons were performed comparing subject and object conditions in each window in both RC types ($p$-values were adjusted for the 24 multiple comparisons; Bretz, Hothorn & Westfall, 2011). The only significant differences between subject and object was in the 1800–2000 ms window in the GE study ($\beta = 0.33$, $SE = 0.14$, $t(54) = 2.3$, $p = .025$).

The significant region in the mixed-model analysis is shown as a curved line in Figure 4. Although the mixed model identifies a fairly strong three-way interaction, only one ge3 region is identified in the post-hoc analysis. The fact that this region is smaller than the region identified by the permutation analysis is due in part to the fact that the cluster is divided across multiple windows, and these windows are treated as independent events. However, this assumption does not hold: there is a correlation of 0.94 between the target preferences in the 1800–2000 and the 2000–2200 ms ge3 windows. The post-hoc analysis does not identify the significant CL cluster that was found in the permutation analysis, and this is because the window is larger than the cluster and hence it potentially includes more noise than the cluster used by the permutation analysis.
analysis. Furthermore, the post-hoc analysis uses a $p$-value threshold that is adjusted for multiple comparisons, and that could help to explain why fewer regions are significant. Thus, the mixed model is a weaker analysis than the permutation test, because it assumes that independent processing components take place in 200 ms windows, and multiple comparisons with higher thresholds for significance are needed to find the window where the effect of RC extraction can be found.

**DISCUSSION**

In the current study, we investigated four-year-old monolingual Cantonese-speaking children’s online processing of two types of RC. Chinese RCs are important in the context of current competing theories of syntactic processing and development because they separate two cues to interpretation that are confounded in more well-studied languages like English. Namely, because Chinese RCs have the typologically rare combination of SVO canonical word order and head-final RCs (Dryer, 2005), object RCs follow the canonical SVO word order, whereas subject RC have non-canonical VOS word order. This therefore pits one cue to interpretation—word order (Diessel, 2007; MacDonald & Christiansen, 2002)—against another cue—subject prominence (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2009; Friedmann et al., 2009; Vasishth et al., 2013; Yun et al., 2015), and allows an explicit investigation of how children may weigh these cues. Our results intriguingly suggest that RC processing in Cantonese depends significantly on the type of RC used and their relationship to other structures in the language. Specifically, for CL RCs, we observed no difference in offline accuracy across subject and object RCs, and a brief but significant time-window which suggested an online processing advantage in favour of object RCs. In contrast, for $ge3$ RCs, we see the more familiar subject advantage, although this was only statistically reliable in the online looking behaviour. We discuss each result in turn.

The significant object advantage for CL RCs is consistent with past comprehension research with four-year-old monolingual children (Chan et al., 2011; cf. Lau, 2016), extending the result to online data. The result is consistent with the suggestion that children gain added value from the isomorphism between simple SVO transitives and classifier object RCs (Chan et al., 2011; Diessel, 2007). Note that this is not to say that the children were processing the sentences as simple canonical sentences. Had they have done this, they would have selected the RC subject rather than the head noun. Instead, the children were processing both sentence types as noun modifiers, as evidenced by their above-chance accuracy in offline responding. As such, we suggest that two properties of Cantonese CL
object RCs may have contributed to the object advantage. First, object RCs follow canonical word order, which may facilitate thematic role assignment (Chen & Shirai, 2015; Diessel, 2007). Second, the possibility that classifier object RCs can be analyzed as internally headed RCs may facilitate processing further (Chan et al., 2011).

Further evidence in favour of the internally headed analysis and against a purely word order based explanation comes from the results of the ge3 RCs. Although CL and ge3 RCs share word order similarities, with object RCs following canonical word order, the children showed a distinct subject advantage when processing ge3 RCs. This result is consistent with developmental studies that have reported offline experimental comprehension data in Mandarin (Hu, Gavarró, Vernice & Guasti, 2016), and with adult studies that report a consistent subject preference (e.g. Jäger, Chen, Li, Lin & Vasishth, 2015; Vasishth et al., 2013). The data from the ge3 RCs therefore support theoretical approaches predicting a subject preference for Chinese RCs (e.g. Bornkessel-Schlesewsky & Schlesewsky, 2009; Friedmann et al., 2009; Vasishth et al., 2013; Yun et al., 2015).

An important question to address from these results is why two largely similar sets of sentences, which can be used for the same function of noun modification, yield different results. One possible explanation for the difference derives from linguistic analyses of the two. Cantonese classifier RCs and ge3 RCs have been argued to have different syntactic structures, reflected by their distinct patterns of syntactic behaviours such as ellipsis and topicalization. Specifically, Cheung and Li (2015) argue that ge3 RCs involve a complementation relationship between RC and the head noun, while classifier RCs involve an adjunction relationship between RC and the head noun (cf. Cheng & Sybesma, 2009). Therefore, one could crucially derive from Cheung and Li’s (2015) analysis that extraction or filler–gap dependency is possible for complementation structures only (i.e. ge3 RCs), but not for adjunction structures with classifier RCs. If Cantonese CL vs. ge3 RCs differ in whether filler–gap syntactic dependency is involved, the subject advantage observed for ge3 RCs may reflect similar structural constraints in the processing of filler–gap dependency structures hypothesized to contribute to processing complexity in European languages (such as structural distance, favouring subject RCs; Friedmann et al., 2009). However, it must be acknowledged that there is considerable debate amongst linguists regarding the analysis of Chinese RCs. The ‘non-uniform’ approach argues that not all types of Cantonese/Mandarin RCs involve filler–gap dependencies (e.g. Cheng & Sybesma, 2009; Cheung & Li, 2015), whereas the ‘uniform’ approach, based largely on Mandarin RCs, argues that all Chinese RCs involve filler–gap dependencies (Aoun & Li, 2003; Simpson, 2002).
While syntactic analyses may provide an independent and linguistically motivated explanation for the difference across CL and ge3 RCs, there are likely to be other sources of information that contribute to or perhaps even explain the effect. One possibility is that structural frequency in the input significantly influences the processing of these structures (Ambridge et al., 2015; MacDonald, 2013). To examine this possibility, we extracted all of the morphologically labelled Cantonese utterances from CHILDES and searched for utterances using go2 CL or ge3 (25,635 utterances from the HKU corpus (Fletcher, Leung, Stokes & Weizman, 2000); 196,931 utterances from the Lee et al. (1994) corpus). No utterances containing ge3 RCs were found, attesting to its low frequency in adult child-directed speech. We extracted 589 utterances containing a verb followed by the demonstrative go2 followed by a classifier (each element separated by any number of words). Since our goal was to understand the biases for subject and object RCs, we further restricted this set to those items with either a noun / pronoun / proper noun before or after the verb preceding go2 (186 potential RCs in adult child-directed speech). Out of this set, only 44 utterances contained CL RCs, with 27 object RCs (i.e, 61.4%) and 17 subject RCs (38.6%). There were an additional 54 simple SVO transitives that were similar in form to the object CL RCs. Overall, we found that CL RCs are more often object- than subject-extracted, and that simple transitives, which share surface identity with object CL RCs, are even more frequent. Consequently, the naturalistic data are consistent with the suggestion that structural frequency affects online processing.

Although we did not find ge3 RCs, there are several studies which report input frequencies for the comparable Mandarin de structures, which we compare to our results. Using the 5 million word Sinica corpus,3 Vasishth et al. (2013) found that sentences following a subject RC pattern (V-N-de-N) are more frequent than structures following an object RC pattern (N-V-de-N) by a ratio of 5.5:1. The pattern holds when specifically looking at subject and object RCs with two animate nouns (4.3:1). Similarly, Yun et al. (2015) found that subject RCs are more frequent than object RCs in the Chinese Treebank 7 (Xue, Xia, Chiou, & Palmer, 2005). In contrast to Cantonese CL RCs, these corpus data suggest a subject preference, which is consistent with the subject preference we observed for ge3 RCs. It must be acknowledged, however, that analyses of child language corpora have yielded different results. Chen and Shirai (2015) found that object RCs were three times more likely than the subject RCs in the input to children (see also Liu, 2015). Importantly, Chen and Shirai (2015) did not control for animacy in their analysis, which is likely to have played a significant role in object RC use. For

instance, work across several languages has shown that children have less difficulty processing object RCs that have inanimate heads (e.g. Kidd et al., 2007; Kirjavainen et al., 2017), which are more frequent in the input.

Predictions based on linguistic analyses and/or input-based accounts mainly predict that one structure will be better than another overall, rather than making strong predictions about the timecourse of processing. In our data, both CL conditions and the ge3 subject RC condition rise to a target proportion of 0.6, which suggests that by 2.5 seconds the children converge on the correct target to a similar degree relative to the start of the trial. However, the ge3 object condition only reaches a target proportion of 0.4, which suggests that children experienced considerable difficulty processing the structures, even after head noun offset. One explanation of this difference can be found in expectation-based accounts of parsing (Hale, 2006; Levy, 2008). Yun et al. (2015) found a subject RC bias in entropy reduction in a minimalist model trained on Mandarin Chinese input (where there was a subject RC bias in the input). Critically, the largest proportional change in entropy was at the head noun, with entropy reduction twice the magnitude in object RC condition compared to the subject RC condition. The effect of extraction for ge3 RCs begins to emerge during the head noun, suggesting that it may in fact begin at ge3. One possibility is that, in the object ge3 condition, children anticipate a simple transitive interpretation following the N-V segment, but must renalalyze the parse at the end of the RC. Renalalysis is costly, especially for children (Kidd, Stewart & Serratrice, 2011; Trueswell, Sekerine, Hill & Logrip, 1999), and may be especially so for ge3 object RCs because of their rarity in the input. The subject ge3 RCs are also a low-frequency structure, but since they are clearly non-canonical, children may identify them as RCs at an earlier point in the sentence.

On the other hand, the object bias in CL RCs only appears for a short period (99 ms), and then the subject RC target proportion catches up and, by the end of the trial, both are near 0.6. One interpretation of this result, which is consistent with the ge3 results and our corpus analysis, is that children are parsing the object RC as a transitive and, given the high frequency of the structural pattern (which is not the case for ge3 object RCs), they show a benefit over the subject RC, which is not similar to canonical SVO transitives. However, once the sentence is experienced, and the participant must select the appropriate toy, they must renalalyze this transitive into a structure which helps them to identify the appropriate referent. This post-structural processing is more difficult for the object RC than the subject RC, because the subject RC is constructed as a RC from the beginning. The short bias for the object CL RCs is then due to the ease of structural analysis and the difficulty of referential processing, while subject CL RCs are the opposite. This referential account depends on
having a discourse context where there is an expectation for noun modification, as in our test materials. When this expectation is not set up, as in Lau (2016), children are often garden-pathed, and do interpret the sentences as SVO transitives (see also Kidd et al., 2015, who showed this occurs in Cantonese–English bilinguals).

Our results therefore provide a novel insight into the specificity with which word order regularities exert an effect on syntactic processing. MacDonald and Christiansen (2002; see also Wells et al., 2009) suggested that differences in word order frequencies (i.e. in English, NVN for subject RCs and NNV for object RCs) explains the subject–object asymmetry in English, because subject RCs, although not particularly frequent, follow the frequent canonical word order (the \text{frequency} \times \text{regularity} interaction). Our data do not support the most general interpretation of this claim – that the effect derives from the frequency of abstract word orders, which in the case of Cantonese predicts an object advantage for both CL and \text{ge}_3 \text{RCs} because both are NVN. Instead, the data, along with our corpus analyses, suggest that structural frequencies within the language can both help (for CL RCs) and hinder (for \text{ge}_3 \text{RCs}) processing, highlighting the tight link between online processing and the input-based learning of structures in typologically different languages (e.g. Chang, 2009; Fitz et al., 2011).

\textbf{CONCLUSION}

Relative clauses in East Asian languages like Cantonese are important in debates concerning syntactic acquisition and processing, since the typological features of East Asian languages potentially allow for long-standing debates, such as the source of the subject–object asymmetry in RCs, to be addressed in languages with fewer confounds than are found in European languages. Consistent with acquisition work on typologically diverse languages such as Finnish (Kirjavainen et al., 2017) and Quechua (Courtney, 2011), the results from the present study highlight the significance of language-specific influences on syntactic processing. We have, for the first time, identified an asymmetry in the online processing of different RC types within the one language, and have argued that this is likely to derive from differences in structural frequencies that either support or do not support correct syntactic predictions to be made at crucial points in the sentence. The results are important because data from new languages widen the evidential base upon which theories can be tested and developed (Kelly, Kidd & Wigglesworth, 2015). They also highlight the value of online studies in acquisition, which have the potential to reveal differences in processing that may not be obvious in offline data.
REFERENCES


CANTONESE RELATIVE CLAUSES


CANTONESE RELATIVE CLAUSES


**APPENDIX A**

Can you pick up [relative clause] head noun? DEM: demonstrative; CL: classifier

**Subject-extracted CL relative clause**

1. 追獅子嘅狗仔
   
   *zeoi* si*zi*2 go2 *zek*3 gau2zai2
   
   chase lion DEM CL dog
   
   ‘the dog that chased the lion’

2. 踢斑馬嘅熊人
   
   *tek*3 baan1maa5 go2 *zek*3 hung4jan2
   
   kick zebra DEM CL bear
   
   ‘the bear that kicked the zebra’

3. 抹豬仔嘅馬騾
   
   *maat*3 zyu1zai2 go2 *zek*3 maa5lau1
   
   wipe pig DEM CL monkey
   
   ‘the monkey that wiped the pig’

4. zit1 馬騾嘅牛牛
   
   *zit*1 maa5lau1 go2 *zek*3 ngau4ngau2
   
   ‘the monkey that kicked the zebra’
APPENDIX B

Can you pick up [relative clause] head noun?

Subject-extracted ge3 relative clause

1. 舔斑馬嘅獅子
   laai2 baan1maa5 ge3 si3zi2
   ‘the lion that licked the zebra’

2. 撞熊人嘅老虎
   zong6 hung4jan2 ge3 lou5fu2
   ‘the tiger that bumped the bear’

3. 咬牛牛嘅大象
   ngaau5 ngaau4ngau2 ge3 daai6zoeng6
   ‘the elephant that the tiger fed’
‘the elephant that bit the cow’

4. 推 長頸鹿 嘅 老虎
teoⅰ coeng4geng2luk5 ge3 lou5fu2
the tiger that push the giraffe’

Object-extracted ge3 relative clause

1. 熊貓 瓦 嘅 獅子
hung4maau1 laai2 ge3 si1zi2
the lion that the panda licked’

2. 大象 追 嘅 老虎
daai6zoeng6 zeoⅰ ge3 lou5fu2
the tiger that the elephant chased’

3. 豬仔 踢 嘅 牛仔
zyu1zai2 tek3 ge3 ngau4zai2
the cow that the pig kicked’

4. 大象 撞 嘅 長頸鹿
daai6zoeng6 zong6 ge3 coeng4geng2luk5
the giraffe that the elephant bumped’