

Subject-Object Asymmetry in the Production of Relative Clauses in Cantonese

Jiaying Huang, Caterina Donati
 University of Paris Diderot

Abstract

In the literature about processing of relative clauses (RCs), subject relatives (SRs) are reported to be easier than object relatives (ORs) in a number of languages, but the status of prenominal ORs in languages where the object follows the verb (SVO) is still partly controversial. This study explores the production of RCs in Cantonese in two elicited production experiments and two corpus studies. In the first elicited experiment, an overwhelming preference for SRs was observed. In two corpus studies where the context and the feature of arguments were uncontrolled, the reverse pattern was observed. In order to reconcile the two datasets, we speculate that what counts in object dependencies is the featural endowment of the subject, as in the intervention hypothesis implemented in Friedmann et al. 2009. A second elicited experiment was run to test this hypothesis. The results suggest that production of RCs in Cantonese displays a subject preference in general and that object dispreference is modulated by featural mismatch.

Keywords

relative clauses, Cantonese, adult production, syntax

1 Introduction

In previous research about processing of relative clauses (RCs), it has been well established that subject relatives (SRs) are generally easier to comprehend and to produce and are acquired earlier than object relatives (ORs) across languages (Tavakolian 1981; O'Grady 1997; Vasishth et al. 2013; Jäger et al. 2015, among others). However, studies about RCs in Chinese (both Mandarin and Cantonese), which are special in being prenominal in an SVO language, often exhibit contradictory results (Lin and Bever 2006; Vasishth et al. 2013; Jäger et al. 2015 vs. Hsiao and Gibson 2003, among others). Animacy of arguments has also been reported to have an impact in the RC processing (Wu et al. 2010, Wu et al. 2012). As for Cantonese RCs, most studies (Lau 2006, Lau 2016; Kidd et al. 2015; Yip and Matthews 2007; Matthews and Yip 2017) are about acquisition. The present article focuses on RC production by adults with the purpose of investigating whether this language displays any subject-object asymmetry and what kind of factors influence RC production. In order to address these questions, two elicited production experiments and two corpus studies have been performed.

Before presenting the experiments and the corpus studies, we first introduce some characteristics of Cantonese RCs and review some models of how RC processing works in Section 1. In Section 2, we present the first elicited production experiment, where all the expected items include animate arguments. We then turn to two corpus studies based, respectively, on a TV series and the Hong Kong Cantonese Corpus (Luke and Wong 2015) in Section 3, in order to compare the experimental findings with the spontaneous data. In Section 4, we present the second elicited production experiment, which aims at eliciting items involving four types of arguments. In the last section, we raise some conclusions and propose an appropriate analysis.

1.1 Characteristics of Cantonese RCs

Cantonese, as the other Sinitic languages, is an SVO language with prenominal RCs. This combination yields a mismatch between the verb phrase (VP), which is head-initial, and the noun phrase (NP), which is head-final. This mismatch may give rise to some processing issues, which we will specify in the next subsection.

In addition, Cantonese RCs also exhibit a large amount of syntax variation. There are generally two types of constructions, depending on the presence of an explicit relativizer. 1) With relativizer: the relativizer *ge3* is found at the right edge of the RC, immediately followed by the head. The head can be either a bare noun or a determiner phrase; see an example in (1). 2) Without relativizer: the RC is directly followed by a classifier or a demonstrative with classifier and the head, as in (2). Notice that object RCs without relativizer are thus ambiguous, since they display the same SVO order exhibited by matrix clauses; see (3). Only the context and possibly the prosody might disambiguate them.

With relativizer:

- (1) [_{RC} siu1fong4jyun4 ziu3-zyu6 Ø_i ge3] sei2sau2_i
 fire-fighter light-ASP REL sailor
 ‘(the) sailor that the firefighter is lighting’

Without relativizer:

- (2) [_{RC} siu1fong4jyun4 ziu3-zyu6 Ø_i] go2 go3 sei2sau2_i
 fire-fighter light-ASP Dem Cl.sg sailor
 ‘the sailor (that) the fire-fighter is lighting’

Matrix clause:

- (3) siu1fong4jyun4 ziu3-zyu6 (go2) go3 sei2sau2.
 fire-fighter light-ASP Dem Cl.sg sailor
 ‘The fire-fighter is lighting (that) sailor.’

1.2 Psycholinguistic approaches for subject-object asymmetry

RCs involve typically a long distance dependency, by which the head is related to a gap position internal to the RC. This is illustrated in (4) and (5) in English and Cantonese, respectively.

OR in English:

- (4) *head* *subject* *object gap*
 ↓ ↓ ↓
 The cowboy [_{RC} that the little girl is greeting _____]
 ↑ | ↓

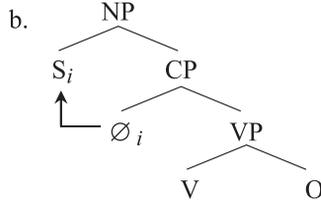
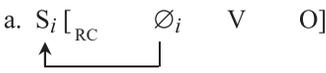
SR in Cantonese:

- (5) *subject gap* *object* *head*
 ↓ ↓ ↓
 [_{RC} _____ man6hau6 ngau4zai2 ge3] sai3lou6nei2
 greet cowboy REL little girl
 | | ↑
- ‘(the) little girl who greets a/the cowboy’

The distance between the head (the filler) and the gap is what can account for the subject preference mentioned in the opening of this article. The exact definition of the notion of “distance” is however still under discussion and is at the basis of the various models explaining processing asymmetries. On one hand, the linear distance-based hypothesis (Dependency Locality Theory: Gibson 1998, 2000) predicts that the shorter the *linear distance* between the filler and the gap, the easier the RC to be processed. On the other hand, the structure distance-based hypothesis (O’Grady 1997) predicts that the shorter the *structural distance* (fewer nodes) between the filler and the gap, the easier the RC to be processed. For SVO languages with postnominal RCs like English, both the two hypotheses predict a subject advantage since the subject gap in these languages is both linearly and structurally closer to its filler in RCs. Linear distance and structural distance in subject and object RCs in English are illustrated in (6) and (7), respectively.

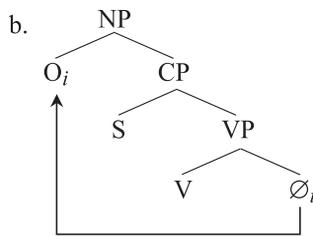
SRs:

(6) the *firefighter*_i [who \emptyset_i is greeting the sailor]



ORs:

(7) the *sailor*_i [that the firefighter is greeting \emptyset_i]



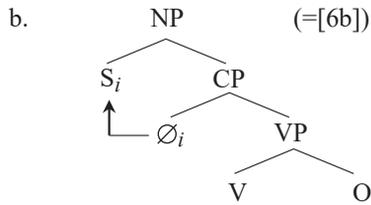
linear distance

structural distance

As for Cantonese, which involves a mismatch between head-initial VP and head-final NP as we mentioned before, the different hypotheses make different predictions: the linear distance-based hypothesis predicts an object advantage; the structural distance-based hypothesis predicts a subject advantage. This makes Cantonese crucial to test them: if Cantonese RCs display a subject advantage, it means that structural distance is more important in processing of RCs; if an object advantage or preference is showed, it means that linear distance is the crucial factor explaining complexity. The mismatch of linear distance and structural distance in RCs in Cantonese is illustrated in the following with an RC with *ge3*. However, the same configurations hold with relatives without relativizer.

SRs in Cantonese:

(8) [_{RC} \emptyset_i ziu3-zyu6 sei2sau2 ge3] siu1fong4jyun4_i
 light-ASP sailor REL firefighter
 '(the) firefighter who is lighting the sailor'

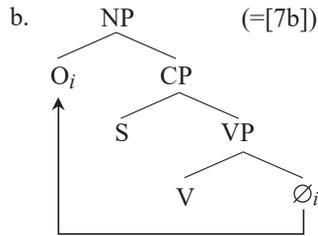


ORs in Cantonese:

(9) [_{RC} siu1fong4jyun4 ziu3-zyu6 \emptyset_i ge3] sei2sau2_i (= [1])
 fire-fighter light-ASP REL sailor
 ‘(the) sailor that the firefighter is lighting’



linear distance



structural distance

Previous studies about RCs in Cantonese report contradictory conclusions: Lau (2006, 2016) reported a subject advantage for the acquisition of RCs by monolingual children (comprehension in 2006, production in 2016); Yip and Matthews (2007) reported an object advantage for the acquisition of RCs by bilingual children. Kidd et al. (2015) reported an overwhelming subject advantage for bilingual children whose dominant language is English and a slight object advantage for monolingual control group. As can be clearly seen, most research about Cantonese RCs concentrated on language acquisition, while processing of RCs by adults remains to be explored. An exception is Pozniak et al. (2017) who reported a slight subject advantage in a visual world eye-tracking experiment. They also reported that the degree of subject advantage is weaker with structure (1) than with (2), with no relativizer. What about production? Does it also reveal a subject-object asymmetry? If so, will it be coherent with what is observed in comprehension? Are there other factors that influence production preferences? In order to answer these questions, we undertook a study about production by adults including two experiments and two corpus studies.

2 Experiment I: Elicited production with animate full-NP arguments

The first experiment involves elicited production with pictures. In this experiment, we concentrated on whether there is a subject-object asymmetry in the production of RCs in Cantonese by adults. Two factors were mainly taken into account: the grammatical function and the number endowment (singular/plural) of the arguments. In order to prevent the influence of other factors such as (in)animacy of arguments and to be parallel with the eye-tracking experiment mentioned earlier, all the potential arguments of RCs were set to be animate full-NPs. Besides, all the verbs provided were set to be semantically reversible.

2.1 Methods

2.1.1 Participant

A total of 20 native speakers from Hong Kong or Macau aged from 20 to 48 years (mean age 28 years) participated in the experiment, which was conducted at Paris Diderot University. All participants were compensated for their participation.

2.1.2 Materials

The elicited production task used Playmobil characters with various accessories to trigger the production of RCs. Eight cards were set up for the experiment. Each card contained four pictures, and each picture contained two interacting characters with some accessories. The characters were identical in all four pictures except for the accessories. In each picture, characters would perform a different scenario. For each card, a list of actions and relevant accessories was provided.

2.1.3 Procedures

Participants were tested individually. They received eight cards (as in Figures 1 and 2) at the beginning of the experiment session and were told that the researcher had another set of cards in which the accessories indicated in the list were absent and the order of the pictures could be different. The first task for them was to find all the accessories of the list in the cards and to describe what happened in each picture with the verbs provided (in the list of actions). Their second task was to help the researcher identify which character had which accessory by answering questions like *which firefighter is wearing the white helmet?* or *which sailor is taking the blanket?*

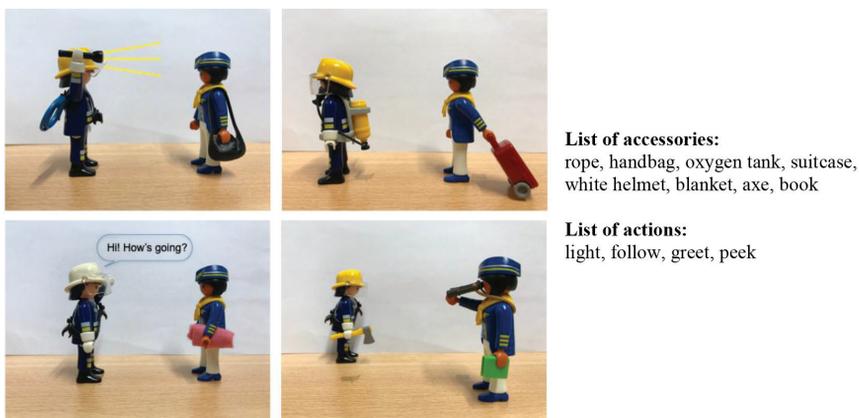


Figure 1 Card for singular condition

These questions were meant to trigger the production of RCs. In the first case, the expected answer was an RC containing a subject gap, something like *the firefighter who is greeting the sailor*, since the firefighter is the agent of the greeting event in Figure 1. This kind of question determines what we call the subject condition. In the second case, since the sailor was the patient of the same greeting event in Figure 1, the expected answer was an RC containing an object gap like *the sailor that the firefighter is greeting*. We call this case the object condition. Notice that in both cases, the entire answer is expected to correspond to an RC, which does not have any grammatical function itself. Some answers did include a presentational copula as an introducer. Fillers were also introduced in the form of questions like

what color is the blanket (that the sailor is holding)? or where is the firefighter in this picture? The ratio of items to fillers was 1:2. A list of probe questions and the corresponding target answers divided by conditions is given in Appendix 1.

In addition to the target production in subject and object conditions, we were also interested in whether the number of the head had an influence. This is why we also set up plural condition cards, in which both the agent and the patient were plural (two characters with the same gender; cf. Figure 2). The experiment had thus a 2×2 mixed design, with four conditions: singular subject condition, singular object condition, plural subject condition, and plural object condition. There were 16 items for each condition, 64 in total. The order of the items during the experiment was pseudo-randomized.



Figure 2 Card for plural condition

2.2 Results

A total of 1226 RCs' productions were collected in this experiment. Among them, 1185 were SRs (cf. [8]), including 567 passivized SRs (cf. [10]), and 41 were ORs (cf. [9]). In all, 97% of produced RCs in the subject condition were subject RCs (Table 1), but only 7% in object condition corresponded to ORs (Table 2). Passivization (89%) was rather the dominant strategy in object condition (Table 2). All in all, subject preference was overwhelming in this task and appeared both in the systematic production of subject RCs in the subject condition and in the massive avoidance of object RCs (replaced by passive subject RCs) in the object condition.

Table 1 Productions in subject condition

Type of answers	Occurrences	%
SRs	618	97
Others	21	3

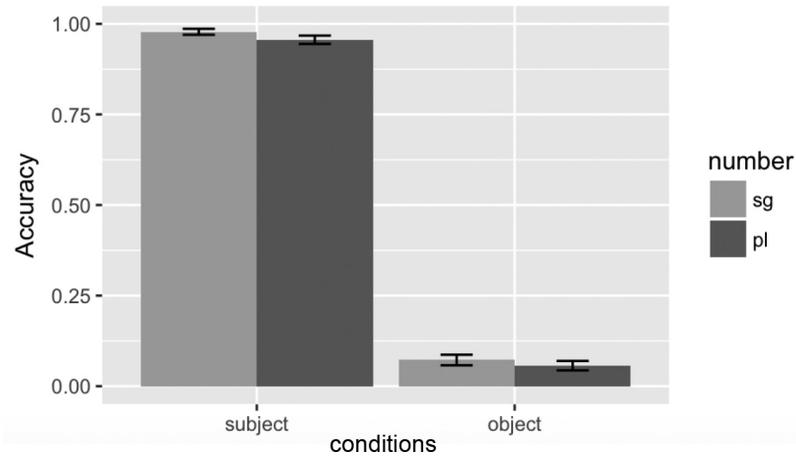
Table 2 Productions in object condition

Type of answers	Occurrences	%
ORs	41	7
Passivized SRs	567	89
Others	27	4

Passivized SR in Cantonese:

- (10) [_{RC} Ø_i bei2 siu1fong4jyun4 ziu3-zyu6 ge3] sei2sau2
 PASS firefighter light-ASP REL sailor
 ‘the sailor who is being lightened by the firefighter’

The grammatical function of the head in the RC had a statistically significant effect in this experiment (subject condition vs. object condition: z -value: -14.486, $\Pr(>|z|)$: $2e-16$; Figure 3). In contrast, the number of the head was not a main factor in this experiment (singular vs. plural: z -value: -0.360, $\Pr(>|z|)$: 0.719; Figure 3). There was no interaction between grammatical function (conditions) and number of the head (z -value: 0.510, $\Pr(>|z|)$: 0.610).¹

**Figure 3** Target production in different conditions

2.3 Interim discussion 1

Comparing our results with those of the eye-tracking experiment we mentioned earlier (Pozniak et al. 2017), we find a much stronger subject preference. Such a difference might be explained as a difference between language comprehension and language production, which is known to involve different

¹ Generalized linear mixed-effect model (GLMER) was applied to analyze the target production, with the conditions (subject vs. object) and the number of the head (singular vs. plural) as fixed effects and with the participants and the items as random effects. Details of the statistic account are given in Appendix 3.

mechanisms (Smolensky 1996; Benedict 1979).² On the other hand, despite this difference, both experiments converge in showing that SRs are indeed easier than ORs in Cantonese both in comprehension and in production. In both experiments, the context was such that it forced a restrictive use of the RC and an interpretation of the head as a definite expression. In both cases, however, the context was very artificial. Will the same asymmetry show up in a spontaneous context?

3 Corpus studies

Considering the artificial contexts and other potential limitations in the experiments, we decided to turn to more natural data by assessing two corpora. The first corpus study is based on manual recording data from a TV series. The second is based on the Hong Kong Cantonese Corpus (Luke and Wong 2015),³ which contains natural spontaneous conversations and several broadcast talk shows.

3.1 Corpus study 1: TV data

The first corpus study is based on manual recording data from a TV series named *House of spirits* – 一屋老友記 produced by Hong Kong Television Broadcasts Limited. There are 31 episodes. Each episode lasts 45 minutes.

A total of 252 RCs were collected: 93 were SRs (one with passive) and 159 were ORs. Statistically, ORs occurred significantly more than SRs (p -value = 3.216⁴). Unlike in the elicited production experiment, where the arguments of RCs were all animate full-NPs, the features of the arguments in the RCs collected in TV series exhibited a wide variation, as can be seen in the following tables.

As far as animacy is concerned, 65% of SRs had an animate head (as in [11] and [12]), while 81% of ORs had an inanimate head (as in [16] and [17]). These figures are summarized in Table 3.

2 It has been argued that ambiguity with main clauses might be a factor explaining preferences and dispreferences of RCs in Cantonese. It is a fact that ORs are word by word ambiguous with a canonical main clause up to the relativizer *ge3* and completely ambiguous with an RC without relativizer. This temporal ambiguity has been advocated as a facilitating factor in the acquisition of RCs (Lau 2016). An anonymous reviewer suggests that the avoidance of object RCs that we observe in our experiment might be explained as an avoidance of this ambiguity. While this is certainly an interesting possibility, we fail to see how the same factor could facilitate acquisition and impede production.

3 We are grateful to an anonymous reviewer who suggested that we should also look at the Hong Kong Cantonese Corpus, which brings valuable naturalistic data and a lot of fun. We are also indebted to the editors of the corpus, for making it available to the research community.

4 Chi-square test was applied to test whether the difference between the proportions of SRs and ORs was significant.

Table 3 Distribution of (in)animacy of heads

Animacy of head	SRs	ORs
Animate	60 (65%)	30 (19%)
Inanimate	32 (35%)	129 (81%)

Up to 80% of ORs displayed an animacy mismatch (as in [16] and [18]), while only 18% of SRs did⁵ (illustrative examples are given in [12] and [14]). These figures are summarized in Table 4. The following examples correspond to the various combinations of animacy in the RCs observed in the TV series.

Table 4 Distribution of (in)animacy mismatch

Animacy between arguments	SRs	ORs
Mismatch	17 (18%)	127 (80%)
No mismatch	75 (82%)	32 (20%)

SRs:

With animate subject and animate object:

- (11) nei1 go3 [_{RC} Ø_i joek6saat3 siu2dung6maat6 ge3] jan4zaa1_i
 Dem Cl.sg sadistically kill small animal REL scum
 ‘this scum who sadistically kills small creatures’

With animate subject and inanimate object:

- (12) [_{RC} Ø_i sik1dak1 wai4sau1 nei1 di1 gau6lau4 ge3] si1fu2_i
 know repair Dem Cl.pl old building REL master
 ‘masters who knows (how to) repair these types of old buildings’

With inanimate subject and inanimate object:

- (13) nei1 go3 [_{RC} Ø_i hak1-zo2 bou2zi6 ge3] geng2hyun1_i
 Dem Cl.sg engrave-ASP BOU character REL collars
 ‘the collar that has been engraved a character *BOU*’

With inanimate subject and animate object:

- (14) [_{RC} Ø_i gwu2lai6 ngo5 ge3] syut3waa6_i
 encourage pron.1sg REL words
 ‘words that encourage me’

5 In our data, all ORs had at least two arguments, but the majority of SRs (63%) had only one argument (subject). An anonymous reviewer pointed out that in such a case, number of arguments (1 vs. 2) should be a different variable in the production. Therefore, we counted them as “no mismatch” instead of “same animacy” in order not to confuse “one animacy” with “same animacy”. After the reanalysis, among the 75 SRs that displayed “no mismatch”, 17 (23%) had two (or more) arguments and 58 (77%) had only one argument.

ORs:

With animate subject and animate object:

- (15) [_{RC} aa3baa4 zeoi3 sek3 Ø_i] go2 go3 (jan4)_i
 Dad superlative spoil Dem Cl.sg people
 ‘the one Dad loved the most’

With animate subject and inanimate object:

- (16) [_{RC} maa4maa1 sung3 Ø_i ge3] lai5mat6_i
 mother give REL present
 ‘a present that (your) mother gives (to you)’

With inanimate subject and inanimate object:

- (17) [_{RC} wai4caan2sing4baan6cyu5 gei3-Ø_i-lei4] go2 fung1 seon3_i
 Probate Registry send-RES.come Dem Cl letter
 ‘the letter Probate Registry sent (to us)’

With inanimate subject and animate object:

- (18) [_{RC} ngo5dei2-gung1si1 san1 ceng2-Ø_i-faan1lei4 ge3] contractor_i
 pron.1pl-company newly hire-RES.return REL contractor
 ‘the contractor that our company recently hired’

Besides, the type of subject appearing in ORs also exhibited a wide and interesting variation: 75% of ORs had a pronominal subject (either *pro* or pronoun), 21% of ORs had a proper noun or a referential noun like “Daddy” or “Big brother” as a subject, and only 6% of ORs had generic or indefinite NPs as a subject. These figures are summarized in Table 5. Some examples are given in the following.

Table 5 Distribution of subject types in ORs

Types of subject	Occurrences	Total
<i>Pro</i>	27	120 (75%)
Pronoun	93	
Proper noun	11	33 (21%)
Referential	22	
Other definite (dem-cl-N)	0	
Generic	4	6 (4%)
Indefinite	2	

ORs:

With null subject:

- (19) [_{RC} *pro* m4 jing1goi1 gong2 Ø_i] go2 geoi3 syut3waa6_i
 Neg should say Dem Cl words
 ‘the words (you) should not say’

With pronominal subject:

- (20) [_{RC} ngo5 zeoi3 sam1ngoi3 Ø_i ge3] neoi5jan2_i
 pron.1sg superlative adore REL woman
 ‘the woman that I love the most’

With proper noun subject:

- (21) [_{RC} saam1bou2 sung3 Ø_i bei2 lei5] go3 sau2doi2_i
 SaamBou give prep.to pron.2sg Cl.sg handbag
 ‘the handbag SaamBou gave to you’

With referential subject:

- (22) [_{RC} aa3baa4 zeoi3 sek3 Ø_i] go2 go3 (jan4)_i (= [15])
 Dad superlative spoil Dem Cl.sg person
 ‘the one Dad loved the most’

With generic subject:

- (23) [_{RC} neoi5jan2 jing1 jau5 Ø_i ge3] ngaang6gin2_i
 woman should have REL hardware
 ‘hardwares that women should have’ (it means: symbols of womanhood)

With indefinite subject:

- (24) [_{RC} kei4taa1jan4 gong2 Ø_i] go2 di1 je5_i
 other people say Dem Cl.pl thing
 ‘those things other people said’

3.2 Corpus study 2: Hong Kong Cantonese Corpus (Luke and Wong 2015)

The second corpus study is based on the Hong Kong Cantonese Corpus produced by Luke and Wong (2015). This corpus contains 30 hours of recordings, up to 180,000 words, extracted from 52 spontaneous conversations and 42 radio programs.

A total of 283 RCs were collected in this corpus: among them, 153 were SRs and 130 were ORs. Although SRs occurred more frequently than ORs, the difference was not statistically significant⁶ (p -value = 0.1716). As in the TV data, the features of the arguments including the RCs collected in this corpus also exhibited a variation, as represented in Table 6.

Table 6 Distribution of (in)animacy of heads in Hong Kong Cantonese Corpus

Animacy of head	SRs	ORs
Animate	96 (63%)	12 (9%)
inanimate	57 (37%)	118 (91%)

6 Chi-square test was applied to test whether the difference between the proportions of SRs and ORs was significant.

Comparing now the two corpora, similar patterns emerged: as for animacy, 63% of SRs had an animate head vs. 91% of ORs had an inanimate head (Table 6). Moreover, 74% of ORs vs. only 22% of SRs displayed an animacy mismatch (Table 7). Of the remaining 120 SRs that did not display any animacy mismatch, only 29 (24%) included two arguments sharing the same animacy, while 91 (76%) had only one argument.

Table 7 Distribution of (in)animacy mismatch in Hong Kong Cantonese Corpus

Animacy between arguments	SRs	ORs
Mismatch	33 (22%)	96 (74%)
No mismatch	120 (78%)	34 (26%)

As for the distribution of subject types in ORs, we adopted a similar classification as in the TV data and observed that 82% of ORs had a pronominal subject (either *pro* or pronoun), 12% of ORs had either a proper noun or a referential noun as a subject, 4% of ORs had a demonstrative-classifier-noun sequence as a subject, and 2% of ORs had a generic bare noun as a subject. These figures are summarized in Table 8.

Table 8 Distribution of subject types in ORs in Hong Kong Cantonese Corpus

Type of subject in ORs	Occurrences	Total
Pronoun	70	107 (82%)
<i>Pro</i>	37	
Proper noun	12	15 (12%)
Referential noun	3	
Other definite (dem-cl-N)	5	5 (4%)
Generic	3	3 (2%)
Indefinite	0	

3.3 Interim discussion 2

In general, we observed converging patterns in the distribution of animacy of the head, of animacy mismatches and of subject types in ORs in the two corpus studies. A difference emerged in the distribution of SRs and ORs: while a significant OR preference was observed in TV data, Hong Kong Cantonese Corpus data exhibited a slight (but not significant) SR preference. This difference may be due to the small sample size of our datasets (252 in TV data and 283 in the Hong Kong Cantonese Corpus) or other unknown factors.

It should also be considered that SRs are more difficult to identify in Cantonese than ORs. This is because in Cantonese, the copula can be omitted with adjectival predicates. An example is given in (25).

- (25) di1 coeng4 fan2 hou2 hou2mei6.
 Cl.pl rice noodle roll very delicious
 ‘Rice noodle rolls are very delicious.’

As a result, a structure like (26), including an adjectival phrase preceding a nominal phrase, can be either interpreted as a simple NP modified by an adjective or as an RC construction including a null copula (see Simpson 2001).

- (26) [_{RC} \emptyset_i hou2 hou2mei6] ge3 coeng4 fan2_i
 very delicious GE3 rice noodle roll
 ‘delicious rice noodle rolls’
 (example extracted from the TV series)

For sake of simplicity, these ambiguous structures were not counted in our analysis. This might have led to underestimate the frequency of SRs in the corpora.

Another factor that makes SRs difficult to identify in Cantonese is *pro*-drop. This entails that a string like (27) can be analyzed as either an SR or an OR. For this kind of ambiguous structures, their classification depends on whether the agent of the event is explicit in the context. If the agent is explicit, we counted them as an OR, otherwise an SR.

- (27) go2 di1 [_{RC} tip3 hai2dou2 ge3] je5_i
 Dem Cl.pl paste here GE3 stuff
 a. SR: go2 di1 [_{RC} \emptyset_i tip3 hai2dou2 ge3] je5_i
 b. OR: go2 di1 [_{RC} pro tip3 \emptyset_i hai2dou2 ge3] je5_i
 Dem Cl.pl paste here GE3 stuff
 ‘those stuff that are pasted here’
 ‘those stuff that someone pastes here’
 (example extracted from Hong Kong Cantonese Corpus)

All in all, the results of the corpus studies appear to be in a striking contrast with respect to the results of the elicited production: while an overwhelming subject preference was observed in the experiment, this predominance of SRs disappears in spontaneous environments. However, if we compare the elicited production experiment and the corpus data in more detail, other differences emerge. First of all, the context in elicited production systematically favored a definite interpretation of the head, while the context in the corpora exhibited a wide variety. Second, all the arguments in the elicited experiment were animate, while the arguments in spontaneous data were diverse. Third, all the arguments were full-NPs in the experiment, while there were many null and pronominal subjects in the corpus.

With this in mind, we can reconcile the results of the two datasets in the following way: SRs are easier to produce in a definite context with animate arguments. ORs are facilitated on the other hand when the arguments of the RC display a featural mismatch: either in (in)animacy (e.g., animate subject vs. inanimate object) or at the lexical level (e.g., pronominal subject vs. full-NP

object – *She vs. president*⁷). This observation is compatible with the intervention hypothesis (Friedmann et al. 2009).

The intervention hypothesis predicts that the difficulty associated with object extraction can be modulated by the featural endowment of the object and the subject: if the subject that *intervenes* structurally between the object gap position and the head of the RC shares the same features with the object, it creates an interference effect in the dependency called an *intervention effect*. If on the other hand, the subject and object display a mismatch in features, the intervening subject will trigger no intervention effect on object extraction. Therefore, ORs should be (almost) as easy as SRs. What we observed earlier might be explained along these lines. Remember that most ORs found in spontaneous data displayed a featural mismatch between subject and object, while all the arguments in the elicited production experiment shared the same features (number, animacy, gender, etc.).

Other differences between the experimental environment and the spontaneous environment might also influence the production in a general way. In order to test the intervention hypothesis in a controlled context, we decided to run a new elicited experiment based on the observations in the elicited experiment and in the corpus studies.

4 Experiment II: Elicited production with four types of arguments

In the new experiment, the design remained similar to that of the first experiment. The only difference was that four conditions were set in accordance with the previous observations: animate full-NP condition (as in experiment I), inanimate object condition, pronoun subject condition and proper noun subject condition. These conditions correspond to the factors that appear to favor the production of ORs in the corpus studies. All conditions were object conditions, i.e., conditions where an OR was expected. What varied were the features of one of the arguments. An example for each condition is given in the following.

OR with animate full-NPs in Cantonese:

- (28) [_{RC} siu1fong4jyun4 ziu3-zyu6 Ø_i ge3] sei2sau2_i (= [1])
 fire-fighter light-ASP REL sailor
 ‘(the) sailor that the firefighter is lighting’

OR with inanimate object:

- (29) [_{RC} go3 nei5si6 cing3-gan2 Ø_i] go2 di1 coi3_i
 Cl.sg lady weigh-ASP Dem Cl.pl cabbage
 ‘the cabbages the lady is weighing’

7 This facilitating effect of a pronominal subject in the production of object relative has been observed in many languages, such as English (Fox and Thompson 1990; Reali and Christiansen 2007), Hebrew (Arnon 2010), and German (Brandt et al. 2009).

OR with pronominal subject:

- (30) [_{RC} keoi5 gan1-zyu6 Ø_i] go2 go3 siu1 fong4jyun4_i
 pron.3sg follow-ASP Dem Cl.sg firefighter
 ‘the firefighter she is following’

OR with proper noun subject:

- (31) [_{RC} Lea gan1-zyu6 Ø_i] go2 go3 siu1 fong4jyun4_i
 Lea follow-ASP Dem Cl.sg firefighter
 ‘the firefighter Lea is following’

A similar test was also run in French as a sort of a control, since featural mismatch has widely been reported to modulate object dispreference in this language (Durrleman 2017). We were also interested in directly comparing the impact of featural mismatch on object relativization when the subject intervenes only structurally as in Cantonese – see (8) above – as opposed to when the subject intervenes both structurally and linearly, as in French – see (6) above. An example for each condition in French is given in the following.

OR with animate full-NPs in French:

- (32) le marin_i [_{RC} que le pompier éclaire Ø_i]
 the sailor that the firefighter light
 ‘the sailor that the firefighter lights’

OR with inanimate object:

- (33) les choux_i [_{RC} que la dame pèse Ø_i]
 the cabbages that the lady weigh
 ‘the cabbages that the lady weighs’

OR with pronominal subject:

- (34) le pompier_i [_{RC} qu’ elle suit Ø_i]
 the firefighter that pron.fem.3sg follow
 ‘the firefighter that she follows’

OR with proper noun subject:

- (35) le pompier_i [_{RC} que Léa suit Ø_i]
 the firefighter that Léa follow
 ‘the firefighter that Lea follows’

4.1 Methods

4.1.1 Participants

A total of 17 native speakers of Cantonese from Hong Kong or Macau aged from 20 to 58 years (mean age 30 years) and 17 native speakers of French aged from 19 to 41 years (mean age 25 years) participated in the experiment, which was

conducted at Paris Diderot University. All participants were compensated for their participation.

4.1.2 Materials

The materials were similar to the ones used in experiment I. Eight cards were set up: each card contained four pictures and each picture contained either two interacting characters as in experiment I or one character and a particular object in the inanimate object condition. The characters were identical in all four pictures except for the accessories. In each picture, a different scenario was displayed. In the cards where the patients of the event were inanimate, the objects were the same in all four pictures except for the colors or patterns. There were a list of actions and a list of accessories for each card. For the cards for the inanimate object condition, there was also a list of objects.

4.1.3 Procedure

As in the previous experiment, participants were first asked to identify the accessories (and the objects) from the list of the cards and to describe the events in each picture with the verbs provided. They were then asked to help the researcher identify the accessories or the objects by answering questions like *which woman is buying the cabbages?* or *which cabbages are yellow?*, which should elicit the production of object RCs. In the inanimate object condition, a question like *which cabbages are green?* was meant to elicit an answer like *the cabbages that the woman is carrying* (cf. Figure 4).⁸ In the pronominal subject condition, participants were asked to use a *pronoun* to refer to a specific character (agent of the event) and to answer a question like *which firefighter is carrying an oxygen tank?* The expected answer was *the firefighter who she is following*. In the proper noun subject condition, participants were asked to use a proper noun to refer to a specific character (agent of the event) like *Lea* in our example (cf. Figure 5) and use this name to answer a question like *which firefighter is carrying an axe?* The expected answer was then *the firefighter who Lea is peeking*. Fillers in this experiment contained irrelevant questions about the color of the accessories or the position of the characters in the pictures. The context of the experiment was controlled as to favor a definite interpretation. A list of probe questions and the corresponding target answers divided by conditions is given in Appendix 2.

8 Remember that our aim here was to verify whether a mismatch in animacy would facilitate object extraction and hence the production of object RCs. For this reason, the subject was always animate in this “inanimate object condition”. As an anonymous reviewer correctly pointed out, it would have been interesting to check whether the sole factor “inanimate object” would affect the production of object RCs even without mismatching. This would require introducing a further condition including an inanimate subject and an inanimate object. This configuration is indeed attested in the corpora; see for example (13). We leave to further research this important verification.



- List of accessories:**
 elastic, red sleeves, red cap, toque
List of objects (cabbage):
 yellow, white, green, red
List of actions:
 weigh, buy, carry, cook

Figure 4 Card for inanimate object condition



- List of accessories:**
 rope, handbag, oxygen tank, suitcase,
 white helmet, blanket, axe, book
List of actions:
 light, follow, greet, peek

Figure 5 Card for the other conditions

Since in the previous experiment number was not a main factor and that number and grammatical function did not display any interaction, plural conditions were not set up in this experiment. Besides, since in the previous experiment the subject condition displayed a high accuracy and that ORs were crucial for testing the intervention hypothesis, this time we only considered the production in object conditions, while the production in the subject condition was kept as a control. Therefore, there were 32 items and 64 fillers in total, with eight items for each condition. The order of the items during the experiment was pseudo-randomized. The visual materials used for the test in Cantonese and in French were identical.

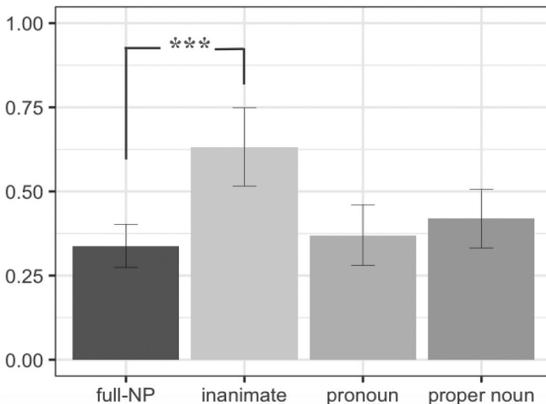
4.2 Results

4.2.1 Results of the test in Cantonese

With 17 native speakers of Cantonese, 537 RCs were collected in the various conditions, that all targeted object RCs. Among them, 236 were ORs and 301 were passivized SRs. Passivized relatives were still produced more often than ORs (55.3% vs. 43.4%; Table 9). ORs were produced most in the inanimate object condition and least in the animate full-NP arguments condition. The production of passivized relatives displayed an opposite pattern: they were found most in the animate full-NP arguments condition and least in the inanimate object condition. A generalized linear mixed model fixed was applied to analyze the target production, with the condition (four-level variable) as fixed effect and the numbers of participants and items as random effects. In comparing the proportions of ORs in the four conditions (Figure 6), only the difference between animate full-NP arguments condition and inanimate object condition resulted as statistically significant (z -value: 5.170, $Pr(>|z|)$: 2.35e-7).

Table 9 Distribution of RCs in four conditions in Cantonese

Conditions	ORs	Passivized SRs	Others
Animate full-NPs	46 (33.8%)	90 (66.2%)	
Inanimate object	84 (61.8%)	47 (34.6%)	5 (3.6%)
Pronominal subject	49 (57%)	85 (62.5%)	2 (1.5%)
Proper noun subject	57 (42%)	79 (58%)	
Total	236 (43.4%)	301 (55.3%)	7 (1.3%)



z -value of inanimate condition: 5.170; $Pr(>|z|)$: 2.35e-07***

Figure 6 Proportions of ORs in four conditions in Cantonese

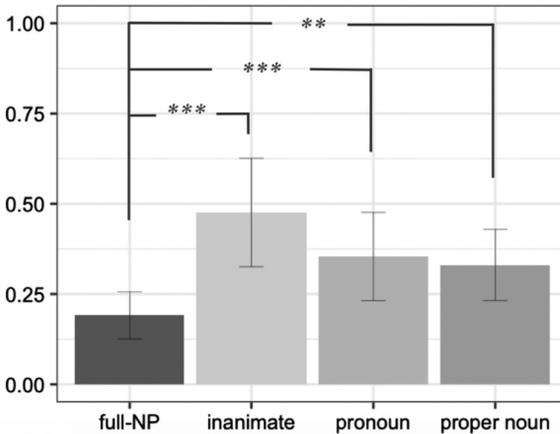
4.2.2 Results of the test in French

With 17 native speakers of French, 536 RCs were collected. Among them, 183 were ORs and 353 were passivized SRs. Passivized SRs were produced much

more often than ORs (64.9% vs. 33.6%; Table 10). ORs were produced most in the inanimate object condition, least in the animate full-NP arguments condition. Conversely, passivized SRs were produced most in the animate full-NP arguments condition and least in the inanimate object condition. As for the comparison of the proportions of ORs in the four conditions, a statistic calculation similar to the one applied for the Cantonese test was conducted. The difference between the animate full-NP condition and the three other conditions ended up being all statistically significant, particularly the difference between the animate full-NP condition and the inanimate object condition (see Figure 7).

Table 10 Distribution of RCs in four conditions in French

Conditions	ORs	Passivized SRs	Others
Animate full-NPs	26 (19.1%)	105 (77.2%)	5 (3.7%)
Inanimate object	64 (47%)	70 (51.5%)	2 (1.5%)
Pronominal subject	48 (35.5%)	87 (64%)	1 (0.7%)
Proper noun subject	45 (33%)	91 (67%)	
Total	183 (33.6%)	353 (64.9%)	8 (1.5%)



z-value of full-NPs condition: 4.045; $\Pr(>|z|)$: 5.24e-05***
z-value of inanimate condition: -5.777; $\Pr(>|z|)$: 7.60e-09***
z-value of pronoun condition: -3.549; $\Pr(>|z|)$: 0.000387***
z-value of proper noun condition: -3.060; $\Pr(>|z|)$: 0.02212**

Figure 7 Proportions of ORs in four conditions in French

4.3 Interim discussion 3

If we compare the results of the test in Cantonese and the test in French, a similar effect concerning the distribution of RCs under the four conditions is observed: the production of ORs is largest in the inanimate object condition and least in the animate full-NP arguments condition. This effect is stronger in Cantonese.

Passivized SRs show the opposite pattern in both tests: they are most produced in the animate full-NP arguments condition and least produced in the inanimate object condition. A greater tendency toward passivization is observed in French. As for nontarget productions, namely, when participants replied with an utterance that was not an RC at all, there was a lot of variation across speakers, ranging from completive clauses like *(in) the picture that the firefighter lights the sailor* or coordinated clauses like *there is a firefighter lighting the sailor and that sailor is carrying a purse*.⁹ Their distribution was strikingly different in the two languages: in Cantonese, they were mainly given in the inanimate object condition, while they were mostly produced in the animate full-NP arguments condition in French.

As for the greater tendency toward passivization observed in French, it might be due to two factors. One is the fact that French allows for more syntactic options for passive. There are three ways to form a passive in French: the copula followed by past participle, the pseudo-causative *se fait*, and a reduced relative. There is only one way in Cantonese, with *bei* (see Tang 2000 and Tang 2001 for a discussion).

Another factor for the weaker avoidance of ORs in Cantonese might be more directly related to intervention and to the mismatch between linear distance and structural distance. Since in French, the subject intervenes both structurally and linearly in object extraction, the processing burden associated with such an extraction might be stronger. In Cantonese, where the subject intervenes only hierarchically but not linearly, object extraction might be less taxing.

In general, these largely converging results of the two tests are in line with the intervention hypothesis: when the arguments of the RC share the same features, object extraction is more difficult. This explains why ORs were found the least in the animate full-NP arguments condition. When the arguments display a featural mismatch, object dispreference can be modulated. In both our tests, ORs were indeed produced more often in conditions involving a featural mismatch and in particular in those displaying an animacy mismatch (inanimate object condition).

5 Conclusion and general discussions

In our study, two elicited production experiments and two corpus studies have been performed. In the first elicited production experiment, which elicited only animate full-NP arguments within a definite context, an overwhelming subject preference was observed. In the corpus studies, where contexts were different and the features of arguments exhibited a wide variation, the frequency of OR was massive instead. These contradictory results were tentatively interpreted in terms of feature-based intervention (Friedmann et al. 2009): subject extraction is easier due to the shorter structural distance between the head and the gap; object

9 In the French test, we also observed other types of production of non-relative clauses, such as temporal adverbial clauses (introduced by *lorsque*).

extraction is more difficult but only insofar as the intervening subject shares the same features as the head. If the subject and the object mismatch in features, the difficulty associated with object extraction decreases. The strong OR avoidance observed in the first experiment is due to a strong intervention effect since both subject and object were definite animate full-NPs.

In order to test this hypothesis, a new elicited production experiment was conducted, where the featural composition of the arguments was manipulated in four conditions: 1) animate full-NP arguments as in the first experiment, predicted to disfavor ORs; 2) inanimate object condition; 3) pronoun subject condition; and 4) proper noun subject condition. The last three conditions, where the subject does not match with the object for animacy, pronominal status, and proper-noun-hood, were expected to enhance object RC production. On account of the previous observation concerning featural mismatches of RCs in French, a parallel test in French was performed. A subject preference was observed in both languages, but feature mismatches did favor the production of ORs. This result supports the structural intervention hypothesis for the two languages: in both cases, varying the featural composition of the subject with respect to the object increases the production of ORs. In the two languages, although, all feature mismatches do not have the same impact. Mismatch in animacy seems to be the most effective in favoring object extraction in both languages. As for the pronominal status of the subject, however, we did not find the facilitation effect predicted on the basis of the corpus studies.

We are well aware that the impact of animacy on object extraction is a well-documented phenomenon (see a.o. Mak et al. 2002, Mak et al. 2006; Traxler et al. 2002, Traxler et al. 2005; Diessel 2007) that has received many explanations in the literature. While a simple facilitation effect associated with an inanimate object would be compatible with many different analyses, what we observed in the corpus studies and experiment II is a more specific effect that speaks in favor of an intervention-like hypothesis: it is not the features of the object *per se* that favor the production of ORs but rather a featural mismatch with respect to the subject.

In all our experiments, we observed a great tendency toward avoiding object relativization by producing passive subject RCs. This tendency is well known to the literature and has been reported in many languages both in adults and in children (for Italian and French, see Utzeri 2007; Guasti and Cardinaletti 2003; Labelle 1990; Contemori and Belletti 2014; for English, see McDaniel et al. 1998; and Novogrodsky and Friedmann 2006 for Hebrew, a.o.). Passivization is interpreted by many authors as a strategy that speakers develop in order to avoid the intervention effect that typically occurs with object extraction: adopting in particular Collins's (2005) smuggling analysis, the preference for passivization over object relativization can be explained along the following terms: passivization involves the movement of the entire VP at the periphery of the clause (in Force), "smuggling" the object across the subject, hence avoiding the intervention problems

or burdens for computation that typically arise when the object is directly moved over the subject, as in object RCs.

In the present article, the distribution of the different constructions for RCs, which are an important characteristic of Cantonese, is not discussed. In the two elicited experiments, however, a clear preference for the construction with demonstrative and classifier was observed, even more so in the object condition. Remember that in both experiments, the context always favored a definite interpretation for the head. The overwhelming preference for this construction is consistent with the hypothesis of Cheng and Sybesma (1999), who claim that the *ge3* construction is only compatible with an indefinite interpretation of the head. Still, a certain amount of *ge3* constructions were indeed produced in the elicited contexts, suggesting that more research is needed in order to establish which exact factors determine the distribution of the various constructions in Cantonese RCs (see also Matthews and Yip 2001, Matthews and Yip 2013; Yu 2006 for other hypotheses). This is left to further research.

Acknowledgments

This research was funded by CNRS Labex EFL, REL working package. We would like to thank Anne Abeillé, Barbara Hemforth, and Marie-Claude Paris for their suggestions for the design of the second experiment; Ewan Dunbar, Dorian Gras, Yair Haendler, and Barbara Hemforth for their advices for the statistic analyses; Céline Pozniak for her contribution to the test in French; Patty Garet and Yuan Zhongjun for their participation in the pilot tests in French and in Cantonese; and Victor Junnan Pan for his encouragement and precious help. Special thanks to all the participants; some of them stayed temporary in Paris, some just began their new life here and some have already spent half of their lives in the city. They did not only come to take the tests but also shared their stories and listened to ours. We had some great times together, and we thank them for this.

References

- Arnon, Inbal. 2010. Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language* 37(1). 27–57.
- Benedict, Helen. 1979. Early lexical development: Comprehension and production. *Journal of Child Language* 6(2). 183–200.
- Brandt, Silke, Evan Kidd, Elena Lieven & Michael Tomasello. 2009. The discourse bases of relativization: An investigation of young German and English-speaking children's comprehension of relative clauses. *Cognitive Linguistics* 20(3). 539–570.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30(4). 509–542.
- Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8(2). 81–120.

- Contemori, Carla & Adriana Belletti. 2014. Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension. *Applied Psycholinguistics* 35(6). 1021–1053.
- Diessel, Holger. 2007. A construction-based analysis of the acquisition of East Asian relative clauses. *Studies in Second Language Acquisition* 29(2). 311–320.
- Durrleman, Stephanie. 2017. Featural mismatches and the comprehension of relative clauses in French: Comparing gender and number. Paper presented at the 13th Generative Approaches to Language Acquisition Conference (GALA 13), Universitat de les Illes Balears, 7–9 September.
- Fox, Barbara A., & Sandra A. Thompson. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66(2). 297–316.
- Friedmann, Naama, Adriana Belletti & Luigi Rizzi. 2009. Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua* 119(1). 67–88.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O’Neil (eds.), *Image, language, brain: Papers from the First Mind Articulation Project Symposium*, 95–126.
- Guasti, Maria T. & Anna Cardinaletti. 2003. Relative clause formation in Romance child’s production. *Probus* 15(1). 47–89.
- Hsiao, Franny & Edward Gibson. 2003. Processing relative clauses in Chinese. *Cognition* 90(1). 3–27.
- Jäger, Lena, Zhong Chen, Qiang Li, Chien-Jer Charles Lin & Shravan Vasishth. 2015. The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language* 79–80. 97–120.
- Kidd, Evan, Angel Chan & Joie Chiu. 2015. Cross-linguistic influence in simultaneous Cantonese–English bilingual children’s comprehension of relative clauses. *Bilingualism: Language and Cognition* 18(3). 438–452.
- Labelle, Marie. 1990. Predication, wh-movement, and the development of relative clauses. *Language Acquisition* 1(1). 95–119.
- Lau, Elaine. 2006. *The acquisition of relative clauses by Cantonese children: An experimental approach*. Hong Kong: The University of Hong Kong M.phil thesis.
- Lau, Elaine. 2016. *Acquisition of relative clauses in Cantonese: A multi-factorial analysis*. Manoa, HI: University of Hawai’i at Manoa dissertation.
- Lin, Chien-Jer Charles & Thomas G. Bever. 2006. Subject preference in the processing of relative clauses in Chinese. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 254–260. Somerville, MA: Cascadilla Proceedings Project.
- Luke, Kang Kwong & May L. Y. Wong. 2015. The Hong Kong Cantonese corpus: Design and uses. In Benjamin K. Tsou & Oi Yee Kwong (eds.), *Linguistic Corpus and Corpus Linguistics in the Chinese Context*. [Monograph series number 25]. *Journal of Chinese linguistics*, 309–330.
- Mak, Willem M., Wietske Vonk & Herbert Schriefers. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language* 47(1). 50–68.

- Mak, Willem M., Wietske Vonk & Herbert Schriefers. 2006. Animacy in processing relative clauses: The hikers that rocks crush. *Journal of memory and language* 54(4). 466–490.
- Matthews, Stephen & Virginia Yip. 2001. Aspects of contemporary Cantonese grammar: The structure and stratification of relative clauses. In Hilary Chappell (ed.), *Sinitic grammar: Synchronic and diachronic perspectives*. Oxford & New York: Oxford University Press.
- Matthews, Stephen & Virginia Yip. 2013. *Cantonese: A comprehensive grammar*. London: Routledge.
- Matthews, Stephen & Virginia Yip. 2017. Noun-modifying clauses in Cantonese. In Yoshiko Matsumoto, Bernard Comrie & Peter Sells (eds.), *Noun-modifying clause constructions in languages of Eurasia: Rethinking theoretical and geographical boundaries*, 105–120. Amsterdam & Philadelphia: John Benjamins.
- McDaniel, Dana, Cecile McKee & Judy B. Bernstein. 1998. How children's relatives solve a problem for minimalism. *Language* 74(2). 308–334.
- Novogrodsky, Rama & Naama Friedmann. 2006. The production of relative clauses in syntactic SLI: A window to the nature of the impairment. *Advances in Speech Language Pathology* 8(4). 364–375.
- O'Grady, William. 1997. *Syntactic development*. Chicago: University of Chicago Press.
- Pozniak, Céline, Jiaying Huang & Barbara Hemforth. 2017. Relative clause processing, structural and linear distance matter – Evidence from Mandarin, Cantonese and English Visual World experiments. Paper presented at the 30th Annual CUNY Conference on Human Sentence Processing, MIT, 30 March–1 April.
- Reali, Florencia & Morten H. Christiansen. 2007. Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology* 60(2). 161–170.
- Simpson, Andrew. 2001. Definiteness agreement and the Chinese DP. *Language and Linguistics*, 2(1). 125–156.
- Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27(4). 720–731.
- Tang, Sze-Wing. 2000. Yueyu beidongju shishizhe de shengluo he yuanze-yu-canshu yufa [Omission of the agent argument in Cantonese passives and the principles-and-parameters framework], *Zhongwen Xuekan* [The Chinese academic journal] 2. 243–260.
- Tang, Sze-Wing. 2001. A complementation approach to Chinese passives and its consequences. *Linguistics* 39(2). 257–295.
- Tavakolian, Susan. 1981. The conjoined-clause analysis of relative clauses. In Susan Tavakolian (ed.), *Language acquisition and linguistic theory*, 167–187. Cambridge, MA: MIT Press.
- Traxler, Matthew J., Robin K. Morris & Rachel E. Seely. 2002. Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47(1). 69–90.
- Traxler, Matthew J., Rihana S. Williams, Shelley A. Blozis & Robin K. Morris. 2005. Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language* 53(2). 204–224.

- Utzeri, Irene. 2007. The production and the acquisition of subject and object relative clauses in Italian: A comparative experimental study. In Kensuke Takita & Chisato Fuji (eds.), *Papers from the consortium workshops on linguistic theory, 2006–2007*, vol. 1. [Special issue 3]. *Nanzan Linguistics*. 283–313.
- Vasishth, Shravan, Zhong Chen, Qiang Li & Gueilan Guo. 2013. Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS One* 8(10), e77006.
- Wu, Fuyun, Elsi Kaiser & Elaine Andersen. 2010. Subject preference, head animacy and lexical cues: A corpus study of relative clauses in Chinese. In Hiroko Yamashita, Yuki Hirose & Jerome Packard (eds.), *Processing and producing head-final structures*, 173–193. Dordrecht: Springer.
- Wu, Fuyun, Elsi Kaiser & Elaine Andersen. 2012. Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes* 27(10). 1489–1524.
- Yip, Virginia & Stephen Matthews. 2007. Relative clauses in Cantonese-English bilingual children: Typological challenges and processing motivations. *Studies in Second Language Acquisition* 29(2), 277–300.
- Yu, Dominic. 2006. Relative clauses and nominal modifiers in Cantonese. Unpublished manuscript, University of California, Berkeley. Available on <http://linguistics.berkeley.edu/~dom/cantonese-rc.pdf>.

Appendix 1: Elicitation of RCs in experiment I

Elicitation of an SR:

Probe questions:

- 1) 請問邊個人拖住個籠？水手定消防員？

Please answer that who is carrying a suitcase? The sailor or the firefighter?

- 2) 咁（即係）四個水手入面，邊一個（水手）拖住籠呢？

Then among the four sailors, (it is) which one is carrying a suitcase?

Target answer:

（即係）跟住消防員嗰個水手。

It is the sailor who is following the firefighter.

Filler question:

個籠係咩顏色㗎？

Which color is the suitcase?

Answer:

（係）紅色嘅。

(It is) red.

Elicitation of an OR:

Probe questions:

3) 請問邊個人預住個氧氣筒？水手定消防員？

Please answer that who is carrying an oxygen tank? The sailor or the firefighter?

4) 咁（即係）四個消防員入面，邊一個（消防員）預住個氧氣筒？

Then among the four firefighters, (it is) which one is carrying an oxygen tank?

Target answer:

（即係）水手跟住嗰個消防員。

It is the firefighter that the sailor is following.

Filler question:

咁呢個消防員企喺圖嘅咩位置？

Then where is this firefighter in this picture?

Answer:

（佢企喺）圖嘅左手邊

(He is) on the left side.



List of accessories:

rope, handbag, oxygen tank, suitcase, white helmet, blanket, axe, book

List of actions:

light, follow, greet, peek

Figure 1 Card for singular condition

Appendix 2: Elicitation of RCs in experiment II

Elicitation of an SR (control):

See 1) and 2) in Appendix I

Elicitation of an OR with animate full-NP arguments:

See 3) and 4) in Appendix I

Elicitation of an OR with inanimate object:



List of accessories:
elastic, red sleeves, red cap, toque
List of objects (cabbage):
yellow, white, green, red
List of actions:
weigh, buy, carry, cook

Figure 4 Card for inanimate object condition

Probe question:

請問邊齣菜係綠色嘅？

Please answer: Which cabbages are green?

Target answer:

個女仔搬緊嗰齣（菜）。

The one that the girl is carrying.

Filler questions:

1) 咁喺呢幅圖入面，你仲見到乜嘢呢？

In this picture, what else do you see?

Answer:

蛋、蘿蔔、薯仔，仲有啲蒜頭。

Eggs, turnips, potatoes, and some heads of garlic.

2) 圖中嘅女仔有乜嘢特徵呢？

What is the characteristic of the girl in this picture?

Answer:

佢戴住一頂紅色嘅 cap 帽。

She is wearing a red cap.

Elicitation of an OR with a pronoun subject:



List of accessories:

rope, handbag, oxygen tank, suitcase, white helmet, blanket, axe, book

List of actions:

light, follow, greet, peek

Figure 5 Card for the other conditions

Probe questions:

1) 請問邊個人預住一條藍色繩索？水手定消防員？

Please answer: Who is carrying a blue rope? The sailor or the firefighter?

2) 請你描述一下呢個消防員嘅特徵？

Please describe the characteristics of this firefighter.

3) 請喺以下嘅問題裡面用“佢”來代替呢個消防員。

Please use “he” to denote this firefighter in the following question.

4) 請問邊一個水手預住一個黑色手袋？

Please answer: Which sailor is carrying a black handbag?

Target answer:

佢照住嗰個水手。

The sailor that he is lighting.

Filler questions:

1) 請問呢個水手企喺圖嘅咩位置？

Please answer: Where is this sailor in this picture?

Answer:

(佢企喺) 圖嘅右手邊。

(She is) on the right side.

2) 佢邊隻手拎住個手袋？

With which hand is she taking the handbag?

Answer:

右手。

Right hand.

Elicitation of an OR with proper noun subject:



List of accessories:
rope, handbag, oxygen tank, suitcase,
white helmet, blanket, axe, book

List of actions:
light, follow, greet, peek

Figure 5 Card for the other conditions

Probe questions:

1) 請問邊個人預住一條藍色繩索？水手定消防員？

Please answer: Who is carrying a blue rope? The sailor or the firefighter?

2) 請你描述一下呢個消防員嘅特徵？

Please describe the characteristics of this firefighter.

3) 可唔可以幫佢改個名啊？(e.g., Patrick)

Could you name him?

4) 請喺以下嘅問題裡面用“Patrick”來代替呢個消防員。

Please use “Patrick” to substitute this firefighter in the following question.

5) 請問邊一個水手預住一個黑色手袋？

Please answer: Which sailor is carrying a black handbag?

Target answer:

Patrick 照住嗰個水手。

The sailor that Patrick is lighting.

Filler questions:

1) 請問呢個水手嘅頭髮係咩顏色㗎？

Please answer: What color is the hair of this sailor?

Answer:

(係) 黑色嘅。

(It is) black.

2) 咁佢對鞋呢？

Then what about her shoes?

Answer:

(係) 藍色嘅。

(They are) blue.

Appendix 3: Statistic calculations

Experiment I: Elicitation production

a. Target as function of subject/object condition:

```
Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: target ~ conditions + (1 | subject) + (1 | code)
Data: sub_rc
Control: glmerControl(optimizer = "bobyqa")
```

AIC	BIC	logLik	deviance	df.resid
447.9	468.5	-219.9	439.9	1270

Scaled residuals:

Min	1Q	Median	3Q	Max
-19.7795	-0.1569	0.0314	0.1324	10.7174

Random effects:

Groups	Name	Variance	Std.Dev.
code	(Intercept)	0.6468	0.8042
subject	(Intercept)	1.2697	1.1268

Number of obs: 1274, groups: code, 64; subject, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.1161	0.4231	9.729	<2e-16 ***
conditionsobject	-7.6015	0.5248	-14.486	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

b. Target as function of number of arguments (singular vs. plural):

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]
 Family: binomial (logit)
 Formula: target ~ number + (1 | subject) + (1 | code)
 Data: sub_rc
 Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	deviance	df.resid
588.5	609.1	-290.3	580.5	1270

Scaled residuals:

Min	1Q	Median	3Q	Max
-10.8599	-0.1794	0.0242	0.1154	6.9203

Random effects:

Groups	Name	Variance	Std.Dev.
code	(Intercept)	28.889	5.375
subject	(Intercept)	1.229	1.109

Number of obs: 1274, groups: code, 64; subject, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.8431	1.0127	0.832	0.405
number2	-0.5002	1.3889	-0.360	0.719

c. Interaction between fixed effects (condition*number):

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [glmerMod]
 Family: binomial (logit)
 Formula: target ~ conditions * number + (1 | subject) + (1 | code)
 Data: sub_rc
 Control: glmerControl(optimizer = "bobyqa")

AIC	BIC	logLik	deviance	df.resid
449.9	480.8	-219.0	437.9	1268

Scaled residuals:

Min	1Q	Median	3Q	Max
-18.1633	-0.1672	0.0348	0.1444	11.3205

Random effects:

Groups	Name	Variance	Std.Dev.
code	(Intercept)	0.5635	0.7507
subject	(Intercept)	1.2602	1.1226

Number of obs: 1274, groups: code, 64; subject, 20

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.4683	0.5334	8.377	<2e-16 ***
conditionsobject	-7.7745	0.6411	-12.127	<2e-16 ***
number2	-0.6868	0.5589	-1.229	0.219
conditionsobject:number2	0.3675	0.7205	0.510	0.610

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

d. Correlation between fixed effects (number+condition):

```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: target ~ number + conditions + (1 | subject) + (1 | code)
Data: sub_rc
Control: glmerControl(optimizer = "bobyqa")

      AIC      BIC   logLik deviance df.resid
 448.2   473.9  -219.1   438.2   1269

Scaled residuals:
   Min      1Q   Median      3Q      Max
-18.7189 -0.1700  0.0349  0.1407  11.6181

Random effects:
Groups Name      Variance Std.Dev.
code   (Intercept) 0.582    0.7629
subject (Intercept) 1.266    1.1253
Number of obs: 1274, groups: code, 64; subject, 20

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.3469    0.4678   9.292 <2e-16 ***
number2        -0.4672    0.3531  -1.323  0.186
conditionsbjet -7.5927    0.5188 -14.636 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr) numbr2
number2      -0.447
condtnsbjct -0.644  0.104
    
```

Experiment II: Elicitation of production with four types of arguments

a. Test in Cantonese: target as function of lexical condition (four types of arguments)

```

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial ( logit )
Formula: Obj ~ lexical.condition + (1 | participant) + (1 | items)
Data: sub_can
Control: glmerControl(optimizer = "bobyqa")

      AIC      BIC   logLik deviance df.resid
 455.0   480.7  -221.5   443.0   531

Scaled residuals:
   Min      1Q   Median      3Q      Max
-4.3709 -0.3867 -0.0790  0.3087  3.2449
    
```

```

Random effects:
Groups      Name          Variance Std.Dev.
items      (Intercept) 0.375   0.6124
participant (Intercept) 8.168   2.8580
Number of obs: 537, groups: items, 32; participant, 17

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.3066    0.7843  -1.666  0.0957 .
lexical.conditioninanimate  2.5912    0.5012  5.170 2.35e-07 ***
lexical.conditionpronoun     0.3122    0.4805  0.650  0.5159
lexical.conditionproper noun  0.7637    0.4778  1.598  0.1100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

b. Test in French: target as function of lexical condition (four types of arguments)

```

Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) [glmerMod]
Family: binomial ( logit )
Formula: Obj ~ lexical_cond + (1 | participant) + (1 | items)
Data: sub_fr2
Control: glmerControl(optimizer = "bobyqa")

```

```

      AIC      BIC logLik deviance df.resid
516.9    542.7  -252.5   504.9     530

```

```

Scaled residuals:
  Min      1Q  Median      3Q      Max
-2.8008 -0.5275 -0.2185  0.5485  3.1943

```

```

Random effects:
Groups      Name          Variance Std.Dev.
items      (Intercept) 0.001264 0.03555
participant (Intercept) 3.945076 1.98622
Number of obs: 536, groups: items, 32; participant, 17

Fixed effects:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.3025    0.5693  -4.045 5.24e-05 ***
lexical_cond2  2.0699    0.3583  5.777 7.60e-09 ***
lexical_cond3  1.2333    0.3475  3.549 0.000387 ***
lexical_cond4  1.0706    0.3499  3.060 0.002212 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Corpus study 1: TV data

Difference between the proportions of SRs and ORs

Hypothesis 0: SRs and ORs are equally distributed

Chi-squared test for given probabilities

data: table

X-squared = 17.286, df = 1, p-value = 3.216e-05

Corpus study 2: the Hong Kong Cantonese Corpus

Difference between the proportions of SRs and ORs

Hypothesis 0: SRs and ORs are equally distributed

Chi-squared test for given probabilities

data: table

X-squared = 1.8693, df = 1, p-value = 0.1716

Jiaying Huang

Mailing address: Laboratoire de Linguistique Formelle (LLF), Bâtiment Olympe de Gouges,
Université Paris Diderot, Place Paul-Ricœur, 75013 Paris, France

Email: jiaying.huang@etu.univ-paris-diderot.fr

Caterina Donati

Mailing address: Laboratoire de Linguistique Formelle (LLF), Bâtiment Olympe de Gouges,
Université Paris Diderot, Place Paul-Ricœur, 75013 Paris, France

Email: caterina.donati@linguist.univ-paris-diderot.fr

Received: October 2, 2018

Accepted: June 24, 2019

對粵語主賓關係從句產出的不對稱性研究

黃嘉穎、Caterina Donati

巴黎狄德羅大學（巴黎七大）

提要

過往關於關係從句加工的研究指出，主語關係從句在很多語言中較賓語關係從句更易處理。然而，對於關係從句出現在名詞之前的 VO 語言是否存在主語優勢爭議較大。本研究通過兩個語言產出誘導實驗以及兩個會話語料研究來調查粵語關係從句的產出優勢。實驗一的結果顯示主語關係從句的產出優勢明顯：在論元特徵相同的情況下，成年母語者傾向使用主語關係從句；而兩個語料研究的結果則呈現相反趨勢。考慮到實驗的背景設置與自然會話不同，兩項研究中論元的特徵也不盡相同，我們認為實驗結果與語料研究的差異可以採用 Friedmann et al. (2009) 提出的針對關係從句的阻隔效應理論來解釋。賓語關係從句中，中心語與移出點所構成的依存關係是否會受到主語阻隔，取決於主語的相關特徵：主語特徵與賓語特徵越相似，阻隔效應越強。為驗證這一假設，我們設計了第二個誘導實驗。該實驗的結果顯示粵語關係從句的產出主要呈現主語優勢，而賓語從句的劣勢則的確能通過“錯開”中心語與主語的特徵得到改善。

關鍵詞

關係從句，粵語，語言產出，句法