Journal of Child Language (2019), **46**, 707–732
doi:10.1017/S0305000919000114

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Testing language acquisition models: null and overt topics in Mandarin

Jingtao ZHU[1,2*] and Anna GAVARRÓ[1]

[1]Departament de Filologia Catalana, Universitat Autònoma de Barcelona, Spain and [2]ClicAsia, Centre d'Estudis Orientals, Barcelona, Spain
*Corresponding author. ClicAsia, Centre d'Estudis Orientals, Carrer de València, 359, 5è 2ª – 6è2ª, 08009, Barcelona, Spain. Tel.: (+34)935143379; E-mail: jtzhu@clicasia.com

(Received 7 August 2017; revised 3 July 2018; accepted 18 February 2019;
first published online 2 April 2019)

Abstract

Parameter setting is either precipitous (Gibson & Wexler, 1994) or it is gradual in response to input frequency (Yang, 2002, 2004). In this study, we compare these models against the empirical domain of subject and (direct) object drop in Mandarin. We conducted a corpus-based study of the speech of 47 Mandarin-speaking children aged 1;2–6;5, and their caregivers, from the CHILDES database. The results show that before age 1;8 all the children used null subjects and null objects in a target-like fashion, which reveals that the parameter that governs null topics is set from very early on, even if the presence of disambiguating evidence for [+Null Topic] patterns is low. Besides, children's *ba* constructions, which require an overt object, reliably included this object from the first occurrence although its frequency was scarce in the input. Our results indicate that the setting of certain parameters occurred early independently of the input.

Introduction

To explain children's rapid mastery of the grammar of any natural language, in the Principles-and-Parameter approach (Chomsky, 1981), acquisition is reduced to lexical learning and parameter setting. Over the past few years, parameter-setting theory has come under criticism in the light of current minimalist thinking. There is controversy about whether the syntax needs to be invariant and parametrization is largely due to some third factors, like frequency (Adger, 2014; Boeckx, 2014) or reduced to variation in the formal features of functional heads (probably only movement-triggering features and unvalued features associated with D, C, and *v*) (Holmberg & Roberts, 2014). In the present study, we compare two classical language acquisition models, namely, the triggering model by Gibson and Wexler (1994) and Yang's (2002, 2004) variational model which relies on input frequency, against the empirical domain of null topic drop as assessed on the basis of null subjects and null objects in Mandarin.

Unlike non-deterministic models, the central claim of the triggering model is that all children start out with either parameter setting, and, based on certain information in

© Cambridge University Press 2019

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

the input data, children can eventually change this initial parameter setting or consistently apply the initial value if it allows them to parse sentences. For the sake of clarity, we limit ourselves to putting forward the algorithmic formulation based on Gibson and Wexler (1994):

(1) Given an initial set of values for n binary-valued parameters P, the learner,
   a. Upon receiving an incoming sentence S, analyzes S with P.
   b. If successful, P remains unchanged; returns to a.
   c. If failure, then
      - changes the value associated with P, obtaining a new parameter setting P_n
      - analyzes S with P_n
      - If successful, adopts P_n
      - otherwise retains P; returns to a.

Therefore, the triggering model is error-driven, in that "the learner does not attempt to change her hypothesis as long as the current input sentence can be syntactically analyzed" (Gibson & Wexler, 1994, p. 410). In addition, learning is online and conservative, in that a grammar G_n(s) is very close to the previous G, differing by only one parameter value. If such a model is on the right track, it follows that:

(2) Learners should have an initial hypothesis about the parameter setting, and they can eventually change the initial value if it does not allow them to parse sentences.

In contrast to the triggering model, the proposal of Yang (2002, 2004) is that learners do not select any values as their initial guess, since both values are available. Yang regards language acquisition as the consequence of a population of competing grammars, whose distribution changes in response to the input presented to the learner. Hence, in this model, the rise of the target grammar is gradual, which means that non-target grammars co-exist for a while before they are eliminated. The procedure can be summarized as follows (from Yang, 2002, pp. 26–7):

(3) For an input sentence S, the child:
   (i) with probability P_i selects a grammar G_i,
   (ii) analyzes S with G_i,
   (iii) • If successful, reward G_i by increasing P_i,
      • otherwise punish G_i by decreasing P_i.

Therefore, in such model, the following predictions are made regarding acquisition:

(4) a. Learners start with two parameter values, and the learning process is a trial-and-error search.
   b. The behavior pattern takes the shape of a gradual curve, that is, abrupt changes in children's linguistic expressions should not be observed. The speed with which the target grammar rises to dominance is determined by the frequency of disambiguating evidence in the input.

With the aim of evaluating these two theoretical models, we compare their predictions against our findings from a corpus-based study of the speech of 47 typically developing,

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

monolingual Mandarin-speaking children and their caregivers from the CHILDES database (MacWhinney, 2000, 2015). We compare the path of acquisition of subjects/objects by children, and rates of these patterns in child-directed speech to evaluate the role of input. All the sentences analyzed included (overt/covert) subjects and objects and appropriateness to context was taken into consideration. Based on that, our goals are, on the one hand, to describe a general developmental pattern of the acquisition of the Chinese Null Topic parameter, and on the other to see how well each model stands up to the empirical findings.

The paper is structured as follows. First, we provide a background description of the Chinese Null Topic phenomenon. Then, we address the respective predictions of the two language models on the acquisition of null subjects and objects. Finally, we present a corpus production study and, on the basis of it, we draw our conclusions.

The Null Topic parameter in Mandarin Chinese

Mandarin Chinese is a discourse-oriented language (Li & Thompson, 1981; Tso, 1977, as cited in Huang, 1984a); both subjects and objects may drop as long as reference can be recovered through discourse. To illustrate, (5) is an example of topic drop.[1]

(5) Speaker A: 你们 都 看 过 泰坦尼克号 了 吗？
Ni-men dou kan guo tai-tan-ni-ke-hao le ma.
2PL all see EXP Titanic PERF SFP
'Have you seen *Titanic*?'
Speaker B: 看 过 了。
Kan guo (Ø) le.
see EXP SFP
'(We) have already seen (it).'

In (5), both the subject 'we' and the direct object *tai-tan-ni-ke-hao* 'Titanic' can be dropped because they can be retrieved from the preceding context. Thus, subject and object drop, as in this case, can be considered instances of topic drop. Roberts (2016) and Roberts and Holmberg (2010) established a hierarchy related to the null-argument system, which is based on whether there are φ-features in the relevant functional heads (in D and T), following a NO > ALL > SOME pattern, where the default option is that NO φ-features are present in D and T; if φ-features are present, the simplest case is that they are present in ALL cases; in the most marked case, they appear in SOME functional projection, but not in all. Thus, they considered Chinese-like topic drop languages to be at the top of the hierarchy, followed by Spanish-like consistent null subjects, which bear φ-features in both D and T, Finnish-like partial null subjects (φ-features are found on T, but not on D), and English-like non-null subject languages (φ-features are found on D, but not on T). Besides, this parameter hierarchy can be paired with a learning path (Biberauer &

[1]The following abbreviations are used in the examples: BA = *ba* construction, CL = classifier, DC = Directional Complement, DE = pre-nominal modification marker or postverbal resultative marker, EXP = experiential aspect, NEG = negation, PERF = perfective aspect, PL = plural, Q = question particle, SFP = sentence-final particle, SG = singular, SHI…DE = focus cleft construction, TOP = topic marker, 3 = third person, Ø = null argument.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Roberts, 2016), namely, learners always assume that the default interpretation is the one that lacks φ-features in all heads and, once φ-features are detected in the primary linguistic data (PLD), learners generalize them to all relevant heads; if then a head which does not bear φ-features is detected, then the learner reconsiders the assumption and postulates that some heads bear φ-features.

However, according to Huang (1989), unlike null objects, not all null subjects in Chinese are null topics. Compare the following interpretations of (6) as (i) and (ii):

(6) 张三 说 很 喜欢 泰坦尼克号。
Zhang-san shuo hen xi-huan tai-tan-ni-ke-hao.
Zhang-san say very like Titanic

(i) 张三$_1$ 说 [Top e_j] 很 喜欢 泰坦尼克号。
'Zhangsan$_i$ said that (he$_j$) liked Titanic.'
(ii) 张三$_1$ 说 [*pro* e_i] 很 喜欢 泰坦尼克号。
'Zhangsan$_i$ said that (he$_i$) liked Titanic.'

Embedded null subjects may refer to a topic, namely some other person in the discourse, as illustrated in (6i) or may be A-bound by the matrix subject *Zhang-san*, as in (6ii). Thus, the null subject in (6ii) is regarded as a null pronominal (*pro*). While null subjects may be analyzed as null *pro*s or null topics, a null object is best analyzed as a *variable* A′-bound by a sentence-initial null topic. The analysis of (5), according to Huang (1984a), is therefore as follows:

(7) Speaker B: [Top e_i], [wo yi-jing kan guo e_i le].
1SG already see EXP SFP
'(It), I have already seen.'

The direct object (*it*) is bound by an operator that is itself null and appears in the topic sentence-initial position. So, given this analysis, there is no genuine zero pronoun in Chinese object position and what is really missing is the topic, that is, "an object is topicalized first before it is deleted from topic position" (Huang, 1984a, p. 542).

Still, not all objects in a topicalized position can be optionally dropped; in particular, the drop of objects in the *ba* construction is ungrammatical, as in (8). *Ba* was originally a lexical word, with the meaning of 'take hold of' or 'grasp' (Wang, 1957, as cited in Li, 2006); through a process of grammaticalization, *ba* has come to precede an object, although it cannot be used to mark the object in a canonical SVO sentence as in (9); rather it precedes the verb. Thus, the basic pattern of the *ba* construction is illustrated in (10):

(8) *我 把 吃 了。
*Wo ba chi le.
1SG BA eat PERF
'I ate (the cake).'
(9) *我 吃 把 蛋糕 了。
*Wo chi ba dan-gao le.
1SG eat BA cake PERF
'I ate the cake.'

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

(10) 我 把 蛋糕 吃 了。
Wo ba dan-gao chi le.
1SG BA cake eat PERF
'I ate the cake.'

Although there is no consensus on the nature of the *ba* construction (see Huang, Li, & Li, 2009; Liu, 2016), it is agreed in the literature that there are several restrictions concerning this construction (see Chao, 1968; Li, 2006). First, the NP following *ba* – referred to as the post-*ba* NP – is an affectee, that is, in our sentence (10), *dan-gao* 'cake' is directly affected by the predicate *chi* 'eat'. Second, no element can intervene between *ba* and the post-*ba* NP. This requirement can be captured by the adjacency condition on Case assignment if, as suggested by Huang *et al.* (2009), *ba* is a Case marker[2] and the post-*ba* NP receives Case from it. Third, the post-*ba* NP bears a feature of definiteness, so that sentence (11) with an indefinite object is ungrammatical:

(11) *我 把 一 块 蛋糕 吃 了。
*Wo ba yi kuai dan-gao chi le.
1SG BA one CL cake eat PERF
'I ate a cake.'

The ungrammaticality of (11), together with the fact that the post-*ba* NP can be followed by Topic markers like *ya*, *a*, or *ne* (Li & Thompson, 1981), as in (12), gives us good reason to think that the post-*ba* NP is in a low Topic headed by an overt or phonetically null operator. The existence of such a Topic is parameterized.

(12) 我 把 这 个 包 呢 放 桌子 上 了。
Wo ba zhe ge ba ne fang zhuo-zi shang le.
1SG BA this CL bag TOP put table up PERF
'I have put the bag on the table.'

The proposal that, like Italian, Chinese displays Topic and Focus projections within IP is not new. Belletti's (2001, 2004) proposal that the architecture of the domain below IP and above *v*P parallels that of the clause-external left periphery has been shown to be applicable to Mandarin Chinese (Badan, 2008). As for the *ba* construction, Tsao (1987) has argued that the post-*ba* NP is a kind of secondary (inner) topic. The preposed object in the *ba* construction (i.e., post-*ba* NP) is not a Focus since, as shown in (13), it cannot be in a cleft configuration by means of a *shi… (de)* pattern, as would be expected if it were really a Focus (Lee, 2005; Teng, 1979).

(13) *张三 是 把 那 本 书 都 买 了 的。
*Zhang-san shi ba na ben shu dou mai le de.
zhang-san SHI BA that CL book all buy PERF DE
'It is this book that Zhangsan has bought.'

[2]There is another proposal according to which *ba* is a preposition (Chao, 1968; Lü, 1980). However, if one assumes *ba* is a preposition, then it is expected that the post-*ba* NP form a constituent with *ba*, namely, the post-*ba* NP alone (NP without *ba*) and the VP should not form a unit. To the contrary, *ba* can take [NP VP] as its complement in Chinese, as shown by the coordination test (see Huang *et al.*, 2009).

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Moreover, since the post-*ba* NP cannot be omitted, and it can be followed by contrastive topic markers like *ne* (see (12)), we assume that it can be analyzed as a Contrastive Topic, and not a typical 'Aboutness Topic'. Compare the following sentences:

(14) a. 蛋糕, 我 已经 吃 了, 张三 也 吃 了。
Dan-gao, wo yi-jing chi le, Zhang-san ye chi le.
cake, 1SG already eat PERF, Zhang-san also eat PERF
'As for cake, I already ate it and Zhang-san also ate it.'

b. #我 已经 把 蛋糕 吃 了, 张三 也 吃 了。
wo yi-jing ba dan-gao chi le, Zhang-san ye chi le.
1SG already BA cake eat PERF, Zhang-san also eat PERF
'As for cake, I already ate it and Zhang-san also ate it.'

c. 我 已经 把 我的 蛋糕 吃 了, 把 张三的 也 吃 了。
wo yi-jing ba wo-de dan-gao chi le, ba Zhang-san de ye chi le.
1SG already BA my cake eat PERF, BA Zhang-san DE also eat PERF
'I already ate my cake and I also ate Zhangsan's.'

An 'Aboutness Topic' interpretation as in (14b) is not plausible, but the contrastive meaning (14c) is felicitous. This provides evidence to the effect that the object in S*ba*OV is a Topic with a contrastive reading. Thus, it could be the case that the main topic sets up the background and the BA topic specifies a contrastive point within that background. These properties of *ba* would constitute a parameter of Chinese that the child would have to set by exposure to positive evidence; the parameter would be dependent on the acquisition of the lexical functional item *ba*.[3]

Since the *ba* construction also implies a topic structure, Chinese null subjects/objects and the *ba* construction are somehow linked in the acquisition of the Null Topic parameter in Chinese, since null topics are not licensed in every position – in particular, they cannot occur with the *ba* construction. In the next section, we take a closer look at the predictions that the triggering model and the variational model make for the setting of the Null Topic parameter.

The triggering model and the variational model in the Chinese Null Topic parameter

The triggering model predicts that the child consistently applies one value for each parameter unless parameter resetting is required. On the other hand, in the variational model, the speed with which a parameter value rises to dominance correlates with the proportions of unambiguous forms ('signatures', in Yang's (2002, p. 39) terms) in the input received by the child. The signature for a parameter refers to "sentences that are analyzable only if that parameter takes on the correct value of the target language" (Yang, 2011, p. 142). With respect to the Null Topic parameter, since subjects and objects may be either null or overt in Chinese (except for the direct object in the *ba* construction, which can only be overt), null subjects in finite sentences like (5), embedded null subjects of type (6i),

[3]Since the proposal of Borer (1984), later known as the Borer–Chomsky conjecture, parameters have been located in the functional elements in the lexicon. *Ba* would be just one such case. See Baker (2008), Roberts and Holmberg (2010), and Boeckx (2014) for discussion of parameters in current minimalist theory.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

and all null objects could be signatures for the Null Topic parameter (Yang, 2002, 2012). Therefore, by calculating the proportion of grammatical [+Null] subjects/objects and [–Null] subject/objects in child-directed speech, following the variational model we can predict the time needed to acquire a Chinese-like Null Topic grammar. If there are abundant instances of those signatures in the child input, we predict that topic-drop will be learned very fast under Yang's model. To the extent that the input contains more overt forms than null forms, then we expect that the child acquiring the Null Topic parameter setting shows a certain variability for a relatively long time. Yang (2002), based on the data of Wang, Lillo-Martin, Best, and Levitt (1992), asserted that the amount of null object utterances (for obligatorily transitive verbs) in the input is sufficiently high (11.6%) for children to acquire the Null Topic parameter very early.

As for the *ba* construction, owing to the non-optional character of its direct object, clauses with the *ba* construction allow us to make some interesting predictions. Under the triggering model, since the presence of the *ba* construction implies a contrastive topic that would be pragmatically or grammatically motivated in all languages (Lacerda, 2017; Lee & Cho, 2009), we expect the children to correctly set the *ba* construction, namely, produce only overt (contrastive) objects at the early phase of development. By contrast, with respect to the variational model, as the frequency of the *ba* construction is scarce (Zhuo, 2005), then the construction should appear late and be preceded by a period of variability. Under this condition:

(15) It is not possible that children just show one type of object (i.e., overt or null) in the *ba* construction at least at an early age.

Method

The present study is based on the speech of Mandarin-speaking children from the corpus of Zhou (2001) and, since there were no subjects of age six and above in that corpus, five additional children from Chang's corpus (1998) were also included. Zhou's corpus includes children recruited in Nanjing, while Chang's children were recruited in Taiwan. They were all from Mandarin-speaking families as parents speak Mandarin Chinese to their children in everyday life. All transcripts can be found in the Chinese subset of the Child Language Data Exchange System (CHILDES) database (MacWhinney, 2000).

All children in this combined corpus were between the ages of 1;2 and 6;5 and were typically developing children. We selected our samples randomly. The corpus includes a total of 4624 child utterances and 8810 utterances from their adult interlocutors.[4]

For the purpose of this study, children were divided into nine age-groups: the 1;2 age-group, the 1;8 age-group, the 2;2 age-group, the 2;8 age-group, the 3;0 age-group, the 3;6 age-group, the 4;0 age-group, the 5;0 age-group and the 5;5–6;5 age-group (with five speakers for each age-group, except for the last group ranging from 5;5 to 6;5, which included seven speakers) as shown in Table 1.

The analysis of the child and adult production was carried out manually unless otherwise indicated. With respect to subject production, we calculated the proportion

[4]Our sample size of 6,005 input sentences with [±Topic] is bigger than Legate and Yang's (2007) sample of 2,226 sentences for Spanish and 2,231 for French, and smaller than their 70,047 sample for English.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 1. Child subjects with age, mean MLU, and Standard Deviation

Age group	N° of Subjects	Mean MLU	Standard Deviation
1;2	5	1.28	0.15
1;8	5	2.01	1.05
2;2	5	2.78	0.81
2;8	5	3.19	0.42
3;0	5	2.40	0.36
3;6	5	2.55	0.17
4;0	5	3.08	0.73
5;0	5	3.10	0.65
5;5-6;5	7	5.44	1.63

of [±Null] subjects in topicalized structures; we found only two child-directed embedded sentences, with only one potential instance of a null *pro* subject like that exemplified in (6ii) – which was excluded from the recounts as involving another parameter. The proportion of [±Null] objects was calculated using a similar method. However, since all null objects are null topics, following Huang (1984a), we considered all the [±Null] objects, including the *ba* construction. In addition, when calculating null objects, following Huang's Phrase Structure Condition (1984b) and the criteria of Packard (2000), we excluded VO compound words as V + O structures. The constructions considered excluded an analysis in terms of VP ellipsis, since they did not involve the copular verb *shi* or any of the modal verbs found in VP ellipsis in Chinese (see Cheung, 2008, and Zhou, 2014, for VP ellipsis in Chinese and its acquisition); in Chinese, unlike what happens in other languages (e.g., Portuguese; Santos, 2009), sentences with null objects and VP ellipsis are not syntactic homophones.

Furthermore, the transitive sentences were picked based not only on the main verbs used, but also the context in which the sentence in question was actually produced, because a verb can be optionally transitive. Thus, grammaticality and acceptability were judged with respect to the given discourse context. We excluded any utterance that appeared to be an immediate imitation from earlier in the same transcript (Demuth, 1996) or 'chunks' that frequently occur in the input (Tomasello, 2003). Such utterances were excluded on the grounds that they do not reliably indicate the child's own grammatical knowledge.

Since efficient automated analysis tools are now readily available, we used CLAN tools (MacWhinney, 2015), in particular the command ¨kwal + t*CHI + s"把" @¨ and ¨kwal + t*MOT + s"把" @¨ to cull all of the utterances that contained the *ba* construction in the children's and adults' spontaneous speech.

Statistical methods

First, descriptive statistics are reported. Then, a Poisson Regression model for the null subject and null object was set up. An offset term and age group as covariate were included. The role of the offset term in the Poisson Regression model is used to

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

model rates rather than counts; this is necessary for instance when individuals are not followed for the same amount of time. In this paper the offset term for the null subject model was total subjects = null subjects + overt subjects and total objects = null objects + overt objects for the null object model. Results from the model are reported below with the estimated proportion and its confidence interval. All results were obtained using R software (version 3.5.0). R packages used in the analysis were stats, emmeans, and RcmdrMisc.

Results

Results for subject production and omission

First we focus on the overall percentage of subject omission in the corpus; we turn to the omission of canonical objects and the *ba* construction in the last two subsections. Out of all sentences, the mean percentage of sentences with null subjects produced by the Chinese children was 48.98 (Standard Deviation, $SD = 24.86$; Standard Error mean, $SE = 3.62$), while it was 49.83 ($SD = 10.82$; $SE = 1.58$) for the Chinese adults. We analyzed whether there were differences in the proportion of subject omission as a function of age and found that initially (in the period starting at age 1;2) all five children studied started out by omitting all subjects ($M = 100$, $SE = 0$). At 1;8, the production of subjects by children increased, and in addition children showed an early sensitivity to language-specific properties of the input data with a subject being produced when required, as in (16a), and a null subject produced when the context favored it (16b).

(16) a. CHI: 我 要 喝 茶。 (Jiang Weiying, 1;8)
Wo yao he cha.
1SG want drink tea.
'I want to drink tea.'

b. MOT: 要 笔 吗 ?
Yao bi ma?
want pen Q
'Do you want a pen?'

CHI: 要 (Jiang Weiying, 1;8)
(Ø) yao
want
'(I) want (it).'

The percentage and number of null subjects found in child- and child-directed speech (CDS) appear in Table 2. We excluded the 1;2 age group from the developmental curve of null subjects because children in this age group had not yet reached the two-word stage (their MLU was below 1.75), so it is by definition impossible for them to produce the arguments of an overt verb, or at least to do so in any systematic way. The developmental curve of subject omission for the remaining eight age groups is shown in Figure 1.

The results indicate that the omission of subjects by children stays around 50% from 1;8 to 6;5 (the oldest age considered). Moreover, a Poisson regression model with an offset term was established to predict the number of null subjects produced by children based on an age effect. From this model, we report the estimated proportion of null subjects and its 95% confidence intervals (CIs) for each age-group, as shown in Table 3.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 2. Percentage (%) and number (N) of null subjects found in child and child-directed speech

	Null subject (Child)		Null subject (CDS)	
Age group	%	N	%	N
1;2	100	18/18	55.66	177/318
1;8	58.43	52/89	55.52	307/553
2;2	51.050	97/19	37.31	175/469
2;8	61.67	74/120	56.67	187/330
3;0	46.15	54/117	52.67	256/486
3;6	42.22	57/135	53.05	270/509
4;0	58.90	129/219	44.23	230/520
5;0	35.10	119/339	47.41	238/502
5;5–6;5	49.49	97/196	47.27	52/110
Total	48.98		49.83	

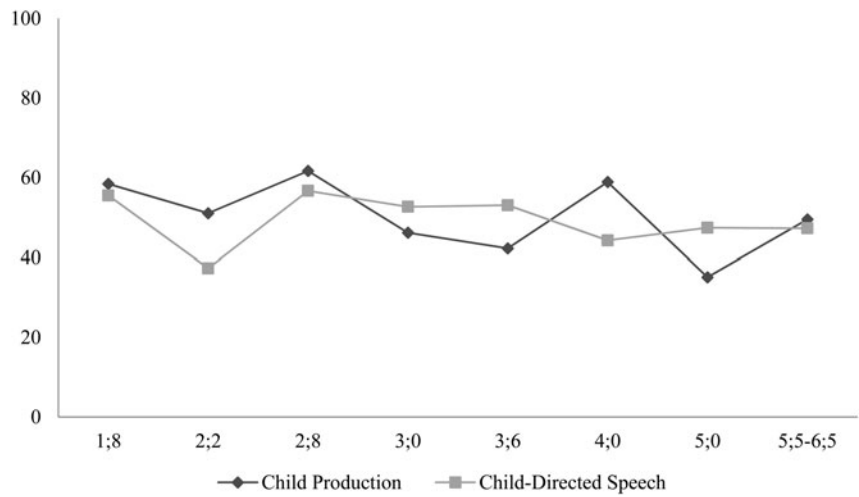


Figure 1. Developmental curve of subject omission in child and child-directed speech.

To examine if there is a significant difference between ages 1;8 ($M = 58.43\%$, $SE = 12.60$) and 2;2 ($M = 51.05\%$, $SE = 0.79$) with respect to the percentage of null subjects produced by children, we ran a post-hoc test. Though the descriptive statistics showed a decline, these differences were not statistically significant ($t(8) = 1.84$, $p = .104$), which means there is no abrupt change in children's performance.

It is possible that the children omitted subjects because they are exposed to instances of null subjects. We calculated the percentage of null subjects in all sentences heard by the children (including fragments, imperatives, etc.), which constitutes 21.5% of the total number of sentences in the corpus (N = 8810). We also examined the incidence of null subjects in adult use at each point in the child's development represented

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 3. Estimated mean of null subjects and its 95% confidence intervals

Age group	Mean (%)	Lower confidence limit	Upper confidence limit
1;8	58.43	44.52	76.67
2;2	51.05	41.84	62.29
2;8	61.67	49.10	77.45
3;0	46.15	35.35	60.26
3;6	42.22	32.57	54.74
4;0	58.90	49.57	70.00
5;0	35.10	29.33	42.01
5;5–6;5	49.49	40.56	60.39
Adult	49.83	47.63	52.13

in Table 2 and Figure 1: the post-hoc test shows that there is no significant difference between children and adult performance by the period starting at age 1;8 ($t(4.83) = -2.04$, $p = .099$). Moreover, a Poisson regression model shows that from 1;8 (the youngest stage considered), age has no statistically significant effect on the production of null subjects by the children. Moreover, from 1;8, children's performance on null subjects is not significantly different from the adult's production, which confirms again our conclusion that the children's production of null subjects has become adult-like by 1;8, that is, quantitatively speaking, by age 1;8 the child production of null subjects has become adult-like. Although the children's subject-drop rate does not approach that of the adults before age 1;8, at that stage the parameter value that allows null subjects seems to be set.

Results across MLU in subject omission

In order to determine whether there is any relationship between the null subject phenomenon and the child's linguistic development, the percentage of null subjects was re-calculated on the basis of the children's MLU. Based on Brown's stages of language development (Brown, 1973), we divided child production into five MLU groups, as in Table 4. A strong positive correlation was found between MLU and the chronological age of our Chinese children ($r = .745$, $p < .001$).

We excluded those children whose MLU ranged below 1.75 (Brown's MLU stage I) from the curve (see Figure 2), since they had not fully reached the two-word stage; because subject production requires the presence of a verb as well, subject omission before MLU 2 may simply result from processing factors – we take those to be responsible for a low MLU in early production. The curve confirms what we found when considering age. There is no statistically significant difference from MLU 1.75–2.25 (Brown's Stage II) to 2.25–2.75 (Stage III), and there is no significant difference between children and adult performance at MLU Stage II regarding subjects drop, which means that by MLU 2.25–2.75 children's use of null subjects has already become adult-like.

In sum, whether divided by age or by MLU, children set the correct parameter value at a very early age (at age 1.8 or before MLU 2.25). This is also consistent with

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 4. Percentage of null subject sentences on the basis of MLU, standard deviation (*SD*), and standard error mean (*SE*) at each stage

MLU stage	MLU range	Null subject	*SD*	*SE*
I	1.0–1.75	93.18%	11.11	3.70
II	1.75–2.25	60.42%	21.07	14.90
III	2.25–2.75	49.27%	23.92	6.63
IV	2.75–3.5	49.21%	17.33	5.23
V	3.5–4.0	36.57%	20.57	8.40
VI	4.0+	73.56%	8.09	3.30

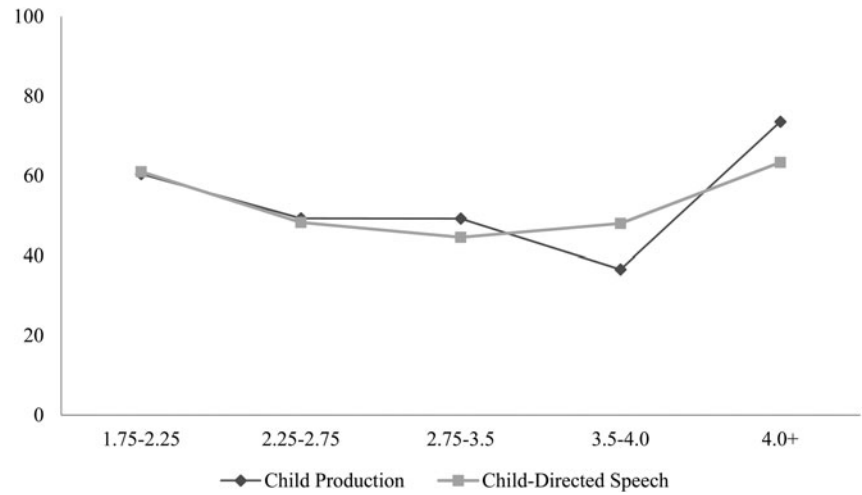


Figure 2. Developmental curve of null subjects by the Chinese children by MLU and child-directed speech.

individual data; that is, no variability among the children was found. All children correctly set the parameter value for null subjects by age 1;8.

Results for object production and omission

The mean percentage of sentences with null objects produced by the Chinese children was 33.19 (*SD* = 26.67; *SE* = 3.96), while it was 34.42 (*SD* = 16.13; *SE* = 2.24) for the Chinese adults.

The results for child and child-directed speech appear in Table 5 and Figure 3. For statistical validity, as in the case of subjects, we also excluded children in the age group of 1;2 from our calculation, since at this age children omitted objects 100% of the time.

The results in Figure 3 indicate that null object production by children stays around 35% from 1;8 to 6;5. We also report the estimation of null objects and its 95% CIs in Table 6. Again, a post-hoc test was run to verify if there is a significant difference

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 5. Percentage (%) and number (N) of null objects found in child and child-directed speech

	Child production		Child-directed speech	
Age group	%	N	%	N
1;2	100	14/14	43.05	65/151
1;8	42.86	15/35	38.67	87/225
2;2	34.57	28/81	26.73	50/187
2;8	41.77	33/79	32.93	54/164
3;0	30.59	26/85	31.87	109/342
3;6	31.31	31/99	28.21	110/390
4;0	28.66	45/157	40.36	134/332
5;0	31.67	57/180	37.35	121/324
5;5–6;5	29.80	59/198	32.26	30/93
Mean	33.19		34.42	

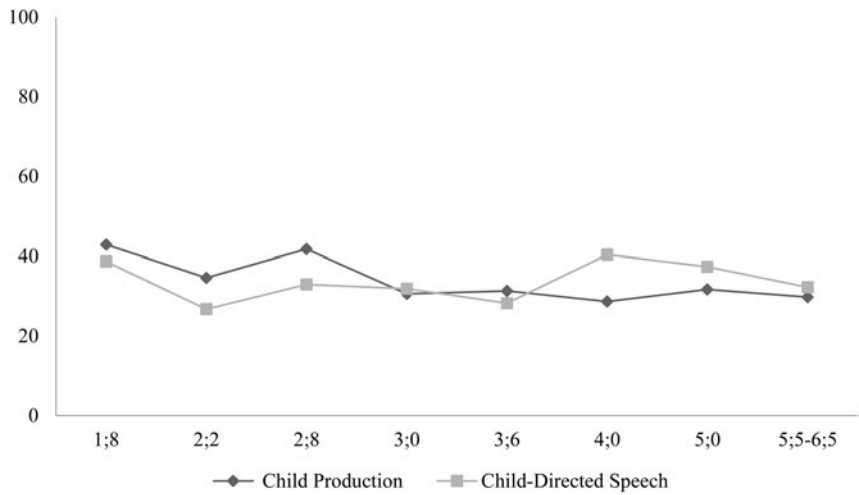


Figure 3. Developmental curve of object omission in child and child-directed speech.

between ages 1;8 ($M = 42.86\%$, $SE = 13.56$) and 2;2 ($M = 41.76\%$, $SE = 10.15$), but none was found ($t(8) = 0.58$, $p = .578$).

The percentage of null objects constitutes 8.6% over the total number of sentences in the corpus (including fragments). Following the same methodology used with null subjects, we also examined the incidence of null objects in adult use at each point in child development: first, there is no systematic decrease in the rate of null objects in the adults' speech that parallels the decrease in the children's use of null objects, as might be expected if children merely imitated the input. Second, there is no significant difference between child and adult performance by the period starting at

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 6. Estimated mean of null objects and its 95% confidence intervals

Age group	Mean (%)	Lower confidence limit	Upper confidence limit
1;8	42.86	25.84	71.09
2;2	34.57	23.87	50.06
2;8	41.77	29.70	58.76
3;0	30.59	20.83	44.93
3;6	31.31	22.02	44.53
4;0	28.66	21.40	38.39
5;0	31.67	24.43	41.05
5;5–6;5	29.80	23.09	38.46
Adult	34.42	32.06	36.96

age 1;8 ($t(8) = -1.1$, $p = .304$), and the Poisson model also confirms our finding; that is, quantitatively, as early as 1;8 the child production of null objects is very close to that of Chinese adults.

In the qualitative study, we counted all verbal utterances in the period starting at 1;8 for which the object is obligatory based on adults' judgments in the specific discourse context. They are illustrated in (17).

(17) (a) MOT: 告诉 阿姨 小 白兔 吃 什么？
Gao-su a-yi xiao bai-tu chi shen-me?
tell auntie little rabbit eat what
'Tell the auntie what rabbits eat.'
CHI: 吃 萝卜 (Xu'er, 1;8)
Chi luo-bo.
eat radish (*Ø)
'(They) eat radish.'

(b) MOT: 你 要 做 哈？
Ni yao zuo ha?
2SG want do q
'What do you want to do?'
CHI: 我 要 拿 这个。 (Jiang Weiying, 1;8)
Wo yao na zhe-ge.
1SG want take this
'I want to take this.'

The use of null objects in that period is also target-consistent. All children produced null objects just like adults, as exemplified in (18).

(18) (a) %act: MOT is teaching how to pull the pen cup out.
CHI: 奶奶 拔。 (Xue'er, 1;8)
Nai-nai ba.
grandma pull out
'Grandma pull (it) out.'

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

(b) %act: CHI is taking the pen out.
CHI: 妈妈 拿 好。 (Jiang Weiying, 1;8)
Ma-ma na hao.
mom take well
'Mom, take (it) well.'

Utterances (18a) and (18b) are fully acceptable with the reference of the null objects determined by the discourse topic. In this regard, the children's null object use at age 1;8 has already converged to the adult grammar. Besides, drawing on examples such as (19), we can assert that children show an early sensitivity to related language-specific properties of the input data. Unlike in English, verbal answers (which recover a form of the verb in the question) are one of the possible answer patterns to *yes/no* questions in Chinese.[5] As noted earlier, verbal answers are not cases of VP ellipsis in Chinese, and so they illustrate null objects.

(19) MOT: 要 喝 点 水 啊？
yao he dian shui a ?
want drink little water Q
'Do you want to drink some water?'
CHI: 喝 (Haohao, 1;2)
drink (Ø)
'I do.'

Results across MLU for object production and omission

The results in Table 7 show that all the children start out by omitting objects very frequently. This result is hardly surprising, given that children at Stage 1 have a very low MLU, and an overt object (just like an overt subject) requires a two-word utterance at least.

Thus, if we exclude the first stage from the calculation, the curve by MLU in Figure 4 confirms what we found when considering age. There was no statistically significant difference between MLU 1.75–2.25 (Stage II) to MLU 2.25–2.75 (stage III) in children's production with respect to object drop and, from MLU 1.75–2.25, the children's performance regarding null objects has become target-like.

In sum, whether divided by age or by MLU, children set the correct parameter value for the parameter determining null objects at a very early age (age 1;8 or before MLU 2.25). It seems unlikely that children omit objects because they mimic the input. If this were the case, we should have found a systematic change in the input (adult child-directed speech) parallel to the change in the output (child speech). However, this was not the case. Thus, different performance among the nine age groups cannot be attributed to the different rate of object drop in the input available to the child.

Results for object production and omission in the ba *construction*

Now we turn to object production in the *ba* construction. Both in young children and adult speech its frequency is relatively low according to our data. In young child

[5]See Jones (1999) on the availability of verbal answers in several languages, including Mandarin and Cantonese.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 7. Percentage of sentences with a null object based on MLU

Stage	MLU range	Null object	*SD*	*SE*
I	1.0–1.75	85.19%	22.74	7.58
II	1.75–2.25	34.33%	9.68	5.39
III	2.25–2.75	31.21%	10.43	2.82
IV	2.75–3.5	33.42%	12.60	3.52
V	3.5–4.0	26.93%	13.52	8.56
VI	4.0+	37.56%	34.43	14.06

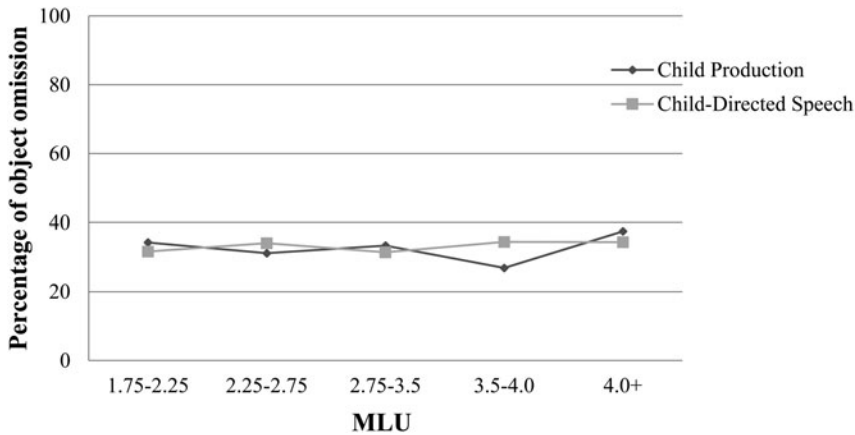


Figure 4. Developmental curve of object drop by Chinese children by MLU and child-directed speech.

production, the total number of sentences involving the *ba* construction in our corpus is 63 and in adult speech it is 298; the overall percentage of the *ba* construction against transitive sentences is 4.9% ($SE = 1.14$) in child production, while in adult speech it is 12.9% ($SE = 1.42$). We also calculated the frequency of adult usage of the *ba* construction in the input (including fragments), which constitutes 3.1% of all utterances. Table 8 reports the results of this analysis.

As shown in the developmental curve in Figure 5, the *ba* construction is first attested at 1;8 and reaches adult levels of production by 4;0. There was a significant change between 3;6 and 4;0 ($t(8) = -2.56$, $p = .034$) and it represented a large-sized effect ($d = 1.62$). However, the input received did not present a significant change ($t(8) = 1.07$, $p = .315$), suggesting that the change could not be the result of mimicking the input.

The percentage of the *ba* construction against transitive sentences based on the children's MLU was also calculated, as seen in Table 9; the result is the developmental curve in Figure 6 (again we excluded the first MLU stage). As can be observed, although the Mandarin-speaking children did not produce the *ba* construction during the first two stages studied (MLU 1.0–2.25), after Stage II, as their MLU increased, the mean percentage of sentences with the *ba* construction increased. By MLU 3.5 their use of the *ba* construction is approaching that of

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 8. Percentage and number of *ba* constructions against transitive sentences produced by children and adults

	Child production		Child-directed speech	
Age group	%	N	%	N
1;2	0.00	0/14	17.88	27/151
1;8	2.86	1/35	22.22	50/225
2;2	2.47	2/81	22.99	43/187
2;8	2.53	2/79	9.76	16/164
3;0	3.53	3/85	12.87	44/342
3;6	2.02	2/99	12.56	49/390
4;0	10.19	16/157	11.14	37/332
5;0	9.44	17/180	6.79	22/324
5;5–6;5	10.10	20/198	10.75	10/93
Mean	4.90		12.90	

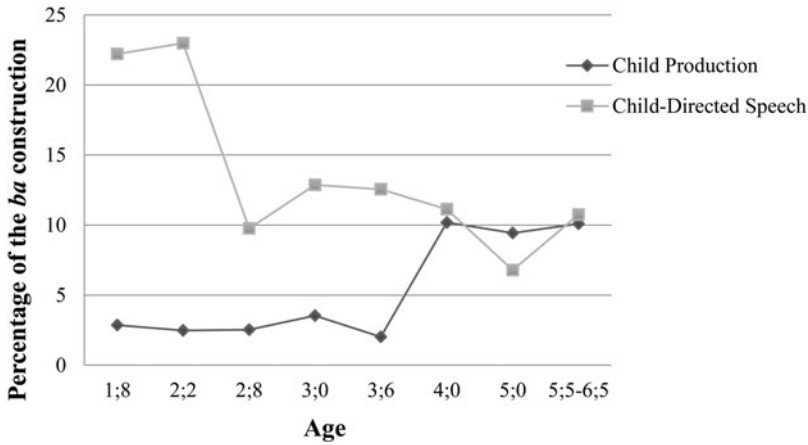


Figure 5. Developmental curve of the *ba* construction by Chinese children by age and by adult speakers.

Chinese adults ($t(10) = 0.15$, $p = .883$). Most importantly, all the children whose productions we analyzed produced an overt object (i.e., post-*ba* NP) in the *ba* construction from the outset (63 sentences with *ba* + NP and no sentences with *ba* + object omission). Examples are provided in (20).

(20) (a) CHI: 把 大灰狼 赶 走 了。 (Haohao, 1;8)
Ba da-hui-lang gan zou le.
BA wolf drive away PERF
'I drove the wolf away.'

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 9. Percentage of the *ba* construction against transitive sentences on the basis of MLU

Stage	MLU range	*ba* construction	*SD*	*SE*
I	1.0–1.75	0.00%	0.00	0.00
II	1.75–2.25	0.00%	0.00	0.00
III	2.25–2.75	3.00%	4.64	1.29
IV	2.75–3.5	6.03%	6.96	1.71
V	3.5–4.0	10.91%	6.68	6.83
VI	4.0+	14.22%	13.98	5.71

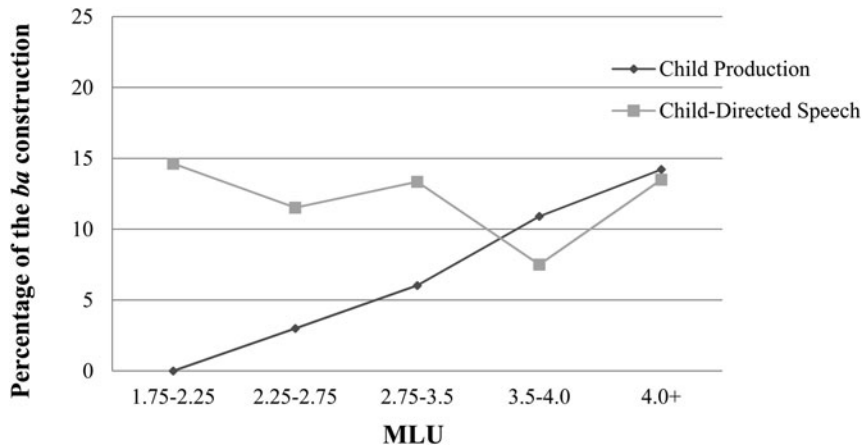


Figure 6. Developmental curve of the *ba* construction against transitive sentences by MLU and child-directed speech.

(b) CHI: 帮 我 把 它 拿 出来。 (Liuzonghao, 2;2)
Bang wo ba ta na chu-lai.
help 1SG BA it take out.
'Help me to take it out.'

(c) CHI: 他 把 苹果 扔 掉 了。 (Marui, 2;8)
Ta ba ping-guo reng diao le.
3SG BA apple throw away PERF
'He threw away the apple.'

(d) CHI: 你把 这个 拿 走, 把 那个 拿 出来。 (Lianlian, 3;0)
Ni ba zhe-ge na zou, ba na-ge na chu-lai.
2SG BA this take away, BA that take out
'You take this away, and take that out.'

The observation that children's *ba* productions reliably included an overt object is also supported by a paired *t*-test, which indicates that the frequency of production of an overt object in the presence of *ba* is significantly greater than its frequency when *ba* is absent ($t(41) = -10.42$, $p < .001$, $d = 3.25$). Although in the corpus we did not find any use of topic markers after the post-*ba* NP by children or by adults, the contexts in which children used

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

post-*ba* NPs are compatible with a contrastive topic interpretation (see 20d). In addition, the fact that all children produced post-*ba* NPs just like adults shows, if we adopt the standard criterion (e.g., Brown, 1973; Thornton & Tesan, 2007) that 90% adult-like usage in obligatory contexts indicates that a grammar structure has been acquired, that the Chinese *ba* construction is an early acquisition (attained by 1;8), although the frequency of occurrence of this structure is low in our corpus (only 3.1%).

Most importantly, the group results are corroborated by individual data. Over the course of language acquisition, the individual use of null objects became adult-like by age 1;8. Regarding the *ba* construction, target-inconsistent use was unattested for the entire dataset.

Results for additional ages between 1;2 and 1;8

Since there is a relatively abrupt change between 1;2 and 1;8, it is possible that more evidence of gradual change might have been found if additional ages between 1;2 and 1;8 were examined. Thus, we conducted a follow-up study using the corpora of Xu/Min/Chen (seven transcripts from two participants; Chen, 2008), Tong (one transcript from one participant; Deng & Yip, 2018), and TCCM (seven transcripts from four participants; Cheung, Chang, Ko, & Tsay, 2011) from CHILDES. The results show an abrupt change in null object production between 1;2 and 1;3; however, for the reasons adduced above, we attribute that not to parameter setting but to the fact that early processing limitations at age 1;2 (and before that age) impinge on the production of multiple-word sequences. The following example (21a) shows that null object production is set correctly even at age 1;3.

(21) (a) FAT: 没 叫 过 他 吗？
Mei jiao guo ta ma?
NEG call PERF 3SG SFP
'Did (you) call him?'
CHI: 没 叫 过 。 (Ma, 1;3.15)
Mei jiao guo.
NEG call PERF (Ø)
'(I) didn't call (him).'

(b) MOT: 你 要 什么？
Ni yao shen-me?
2SG want what
'What do you want?'
CHI: 要 布丁。 (Yang, 1;5)
Yao bu-ding.
want pudding
'(I) want pudding.'

The examples in (21) support a parameter-setting model over an input-driven learning model, with a null topic being produced when the context allows for it (21a), and an overt object produced when required (21b).

Discussion

If verbal answers correlate with the default interpretation of a *yes/no* question, which is a topic-centered interpretation (see Reinhart, 1982; Santos, 2004), the fact that verbal

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

answers are available from the earliest stages in Chinese implies that children can correctly identify the topic in a question. This is consistent with the recent finding that 22-month-old children are able to interpret sentences with Topics in child French (Lassotta, Omaki, & Franck, 2015) and that Korean children are learning to understand and produce topic markers, including contrastive topic markers, before age 2;0 (Lee & Cho, 2009). Furthermore, it has been shown cross-linguistically that children produce full DP subjects and objects in an adult-like manner from early on whether in a pro-drop language (Lorusso, Caprin, & Guasti, 2005, for Italian; Gavarró & Cabré, 2009, for Catalan) or topic-drop language (Sugisaki, 2009, for Japanese). Our study provides evidence that children acquiring Chinese are sensitive to topicalization at around 20 months of age, something that has been argued for also in comprehension studies on other languages (see, for example, De Cat, 2000, 2008, on French; Abalada, 2013, on Portuguese).

The fact that there is no abrupt change in the children's distribution of null subjects and null objects (if we exclude the 1;2 age group, which we take to be limited for processing reasons) can be explained by both the variational and the triggering model. For the variational model, language acquisition is the consequence of competing grammars determined by the input frequency; thus, as predicted in (4), the behavior pattern should take the shape of a gradual curve; as for the triggering model, since children start out with either parameter, a sharp change is only observed in cases where parameter resetting is required (i.e., the selected grammar cannot parse the sentence) and when this process occurs after production starts, and so is visible.

On the other hand, the input frequency of null objects (8.6%) is very low compared to that of null subjects (21.5%), which might imply a longer variational stage with null objects than with null subjects. According to Pierce (1989), French-speaking children learn that French is a verb raising language by 1;8. Yang (2002), based on the CHILDES database, estimated that signature sentences for verb raising, which is expressed by the form V_{FIN} Neg/Adv, constitute 7% of all the sentences that children acquiring French hear. Thus, he concluded that the frequency of the signature for a parameter to be set early must be at least 7% of the input data. Yang (2002) also asserted that the amount of evidence for assessing null object utterances in the input is sufficiently high for children to acquire the Null Topic parameter very early, based on the data in Wang *et al.* (1992), and at first sight our data seem to support the quantitative analysis of Yang. Null subject sentences constitute 21.5% of the input data, and null object sentences constitute 8.2% of the input data, both percentages above the 7% threshold, thus guaranteeing early acquisition. The difference in the relative frequencies of null subjects and null objects can be considered to have no import for acquisition even in a model based on frequency given the fact that all null topics fall under the [±Null Topic] parameter.

On the other hand, recall that there is not just one but two parameters controlling the use of null subjects: the Null Topic parameter (regulating null variables) and the *pro* null subject parameter affecting embedded sentences. There were no embedded sentences in the child production data examined in this paper, allowing for comparison between null topics and null *pro* subjects. We leave for future research the important issue of how children set the *pro* subject parameter in a language like Chinese, where evidence for *pro* subjects like those exemplified in (6ii) is scarce, unlike in the Spanish-like consistent null subject languages.[6]

[6]Likewise, the more general question of what amount of triggering data is required for a parameter to be set is not addressed in this paper.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

However, the *ba* construction, which requires the obligatoriness of an overt object, which in turn is constrained by another parameter, determining the existence of a low Topic with contrastive force, comprises just 3.1% of the input. According to Yang's (2002) calculation, this structure should be acquired late, because its frequency is less than the threshold for an early acquisition (i.e., 7%). Yet, contrary to this prediction, children are able to correctly set the post-*ba* NP from the outset.

If the child initially has probabilistic access to both the positive and the negative value of any parameter,[7] and this increases and decreases in weight as in (3) or (4), there is a probability at least greater than zero that the Null Topic parameter value can be used as evidence for the child to produce a null Topic in the *ba* construction. In other words, the relatively high frequency of null objects in Chinese and the low frequency of the *ba* construction in the input should lead children to drop objects (i.e., post-*ba* NP) in this construction at least at the early stage of *ba* production. However, this prediction is not fulfilled: even at age 1;8 children produce post-*ba* NPs in the *ba* construction consistently (see (20)). Again, there is no variational stage such as the one expected in the variational model.[8]

Even if there is no evidence for a variational stage in child production, it is possible that input factors are responsible for the different rates and patterns of production as proposed by Yang (2002, 2004). Under the variational approach of probabilistic learning (Legate & Yang, 2007; Yang, 2016), if one grammar has a higher number of ambiguous forms, the child will take much longer to acquire that grammar. Thus, by calculating the proportion of [±Null Topic] subjects and objects in the input, we can predict the time to acquire a Mandarin-like [±Null Topic] grammar.

Based on that, we compared the overall rate of [±Null Topic] subjects and objects in the adults' production. The results of our counts are summarized in Table 10. We found 2652/6005 (44.2%) recorded sentences that unambiguously implicate a [+Null Topic] grammar. These are countered by 3353/6005 (55.8%) sentences recorded that, by virtue of being consistent with the [–Null Topic] grammar, may impede the acquisition of the target Chinese [+Null Topic] grammar. This means that the numerical disadvantage of the target null Topic grammar is 11.6%.

In the case of the *ba* construction, since an overt object is always required, there is no [+Null Topic] option and the child only has evidence to consider her grammar as non-null object. If we exclude this structure from the setting of the [+Null Topic] grammar due to its non-optionality, the numerical advantage of the target [+Null Topic] grammar is still low (i.e., lower than 7%).

One can find more or less the same numerical percentage in Legate and Yang's (2007) work in the context of the acquisition of English [+Tense] grammar – see Table 11. The numerical advantage of our [+Null Topic] grammar is much lower than the advantage of the [+Tense] grammar of English (5.8%). Based on that, we expect, if we adopt Yang's model, that Chinese children will converge to a target

[7]Here, we put the question of the brain's computational capability aside (see Wexler 2011).

[8]Alternatively, the fact that children always produce an overt object at the early phase could be explained if we assume that children are guided by a universal constraint on contrastive elements in a salient position, which should be phonologically overt. This would not be parameterised, but follow from a universal (third factor) constraint. In this case, the results so far would be amenable to a variational account simply by arguing that some of the properties of *ba* fall outside the range of variation. However, the fact that children always produce the *ba* construction with the correct word order (i.e., non-canonical S*ba*OV, rather than SV*ba*O) can hardly be captured without taking syntactic properties into account.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 10. Numerical advantage of [–Null Topic] grammar in the adults' production

Rewards [+Null Topic] subject	1892/3797 (49.83%)
Rewards [–Null Topic] subject	1905/3797 (50.17%)
Rewards [+Null Topic] object	760/2208 (34.42%)
Rewards [–Null Topic] object	1448/2208 (65.58%)
([–Null Topic] subject + [–Null Topic] object] – [+Null Topic] object + [+Null Topic] subject)%	55.8% – 44.2% = 11.6%

parameter setting (that is, a Null Topic grammar) even much later than English-speaking children acquire a [+Tense] grammar (i.e., more than 3;5). However, both quantitative and qualitative data show that Chinese children are performing at ceiling before 3;5, which reveals that the cross-linguistic implications of Yang's model are not fulfilled.

Moreover, our findings show that the *ba* construction is found at a rate of 3.1% of all the sentences in the input; therefore, as we discussed above, following the variational model, this structure should be acquired late. However, this is in disagreement with the fact that all the children in our study were able to produce the post-*ba* NP in a target-like fashion from the outset at age 1;8, which reveals that the *ba* construction is an early acquisition.

We think that our results on the acquisition of null topics could be explained by the parameter hierarchy of Roberts (2016) and Roberts and Holmberg (2010) presented at the beginning of the paper, which may lead to a learning path as assumed by Biberauer and Roberts (2016). As we move successively down in the hierarchy, systems become more marked, and more computationally complex for learning; children only adopt such settings when forced to by PLD incompatible with the less marked setting. Thus, Mandarin-speaking children set the [+Null Topic] parameter very quickly because it is at the top of the hierarchy, which is the option preferred by the learner and is assumed to represent their initial hypothesis.

The acquisition of the *ba* construction depends on the acquisition of the features of the lexically marked element *ba*; here an overt object is expressed in 100% of all relevant utterances, which leads children to produce correctly the *ba* construction from early on (for an alternative view on the acquisition of micro-parameters, see Westergaard's, 2009, micro-cue approach).

Conclusion

Our results show that Mandarin-speaking children have a grammar that allows for null topics at an early age in their naturalistic speech, that is, roughly at the age of 1;8, or before MLU 2.25; the production of both null subjects and null objects by Chinese children is indicative of a target-like grammar. Moreover, a numerical advantage of the [–Null Topic] grammar was found, which means, if we adopt Yang's variational model, that Mandarin-speaking children should struggle for a long time to figure out the setting, since there is little disambiguating evidence available in the input.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Table 11. Quantitative evidence in favor of the [+Tense] grammar and the reported duration of the RI stage in three languages (from Legate & Yang, 2007, p. 336)

Language	% [+T] – % [–T]	Duration
Spanish	60.2%	~2;0
French	39.6%	~2;8
English	5.8%	>3;5

In the second place, the *ba* construction is scarce in the input, thus it should appear late and be subject to a period of variability if the variational model is adopted. However, our analysis of the *ba* construction shows that, first, it is an early acquisition (as early as 1;8), and second, there is no such variational stage because all the children observed produced overt objects in the *ba* construction, like adults, from the first occurrence.

These observations provide counter-evidence to the variational proposal. Admittedly, the conclusions we can draw on the basis of naturalistic data depend on the number of datapoints considered; the corpus examined here is of considerable size, and the results are very consistent across ages. Although we believe that our study of the null topic parameter in child Chinese favors some form of the triggering model, this does not mean that factors like frequency play no role at all in language acquisition. Many studies point out that children are powerful statistical learners (e.g., Kuhl, 2010; Rische & Komarova, 2015; Yang, 2013), and the role of frequency has long been recognized in lexical and phonological processes (e.g., Labov, 1981; Phillips, 1984). The evidence we have presented shows, instead, that the setting of certain parameters can be very quick, despite how much (or little) unambiguous evidence there is in the input. The setting of the *ba* construction also suggests that children are not simply setting parameters based on what appears most frequently in their input, which means that parameter setting cannot be explained simply by input frequency, at least not in a superficial sense, as observed by many others (Friedmann & Costa, 2011; Gavarró, Lewandowski, & Markova, 2010; Guasti, Gavarró, De Lange, & Caprin, 2008; Roeper, 2007; Thornton & Tesan, 2007), and our study urges us to reconsider the role of the classical triggering theory in language acquisition as suggested by Sakas (2016).

Acknowledgments. The authors would like to acknowledge the comments and suggestions by two anonymous reviewers, as well as the help of Anna Espinal and Ester Boixadera at the Servei d'Estadística Aplicada of the Universitat Autònoma de Barcelona. This work has been supported by projects FFI2017-87699-P, FFI2014-56968-C4-1-P, and 2017 SGR 634.

References

Abalada, S. (2013). Acquisition of the left and right peripheries in European Portuguese. In S. Stavrakaki, P. Konstatinopoulou, & M. Lalioti (Eds.), *Advances in language acquisition: Proceedings of GALA 2011* (pp. 4–13). Newcastle: Cambridge Scholars Publishing.

Adger, D. (2014). Variability and grammatical architecture. In M. C. Picallo (Ed.), *Linguistic variation in the minimalist framework* (pp. 179–96). Oxford University Press.

Badan, L. (2008). Preposed object and low periphery in Mandarin Chinese. *Working Papers on Language and Cognition* (StiL CISCL) *2*, 18–41.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Baker, M. (2008). The macroparameter in a microparametric world. In T. Biberauer (Ed.) *The limits of variation* (pp. 351–374). Amsterdam/Philadelphia: John Benjamins.

Belletti, A. (2001). 'Inversion' as Focalization. In A. Hulk & J. Y. Pollock (Eds.), *Inversion in Romance and the theory of Universal Grammar* (pp. 60–90). Oxford University Press.

Belletti, A. (2004). Aspects of the low IP area. In L. Rizzi (Ed.), *The structure of IP and CP: the cartography of syntactic structures, Vol. 2* (pp. 16–51). Oxford University Press.

Biberauer, T., & Roberts, I. (2016). *Emergent parameters and syntactic complexity: new perspectives*. Paper presented at the 49th Annual Meeting of the Societas Linguistica Europaea, Università degli Studi di Napoli Federico II, Naples.

Boeckx, C. (2014). What principles and parameters got wrong. In M. C. Picallo (Ed.), *Linguistic variation in the minimalist framework* (pp. 155–78). Oxford University Press.

Borer, H. (1984). *Parametric syntax: case studies in Semitic and Romance languages*. Dordrecht: Foris.

Brown, R. (1973). *A first language: the early stages*. Cambridge, MA: Harvard University Press.

Chang, C. (1998). The development of autonomy in preschool Mandarin Chinese-speaking children's play narratives. *Narrative Inquiry*, *8*(1), 77–111.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, J. (2008). *The acquisition of verb compounding in Mandarin*. Doctoral dissertation. Nijmegen & Amsterdam: Max-Planck Institute for Psycholinguistics & Free University of Amsterdam.

Cheung, H., Chang, C.-J., Ko, H.-W., & Tsay, J. (2011). Taiwan Corpus of Child Mandarin (TCCM). (NSC 96-2420-H-002-030).

Cheung, L. Y.-L. (2008). First language acquisition of elliptical structures in Cantonese. In H. Chan, E. Kapia, & H. Jacob (Eds.), *BUCLD 32 Online Proceedings Supplement*. Retrieved from <http://www.bu.edu/bucld/proceedings/supplement/vol32/>.

Chomsky, N. (1981). *Lectures on government and binding: the Pisa lectures*. Dordrecht: Foris.

De Cat, C. (2000). Structure building and the acquisition of dislocations in child French. In C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Language acquisition and development: Proceedings of BUCLD 24* (pp. 242–52). Somerville, MA: Cascadilla Press.

De Cat, C. (2008). Experimental evidence for preschoolers' mastery of topics. In A. Gavarró, & M. J. Freitas (Eds.), *Proceedings of GALA 2007* (pp. 155–65). Newcastle: Cambridge Scholars Publishing.

Demuth, K. (1996). Collecting spontaneous production data. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's syntax* (pp. 3–22). Cambridge, MA: MIT Press.

Deng, X., & Yip, V. (2018). A multimedia corpus of child Mandarin: the Tong corpus. *Journal of Chinese Linguistics*, *46*(1), 69–92.

Friedmann, N., & Costa, J. (2011). Acquisition of SV and VS order in Hebrew, European Portuguese, Palestinian Arabic, and Spanish. *Language Acquisition*, *18*(1), 1–38.

Gavarró, A., & Cabré, Y. (2009). Subjects, verb classes and word order in child Catalan. In J. Grinstead (Ed.), *Perspectives on typical and atypical Hispanic language development* (pp. 175–94). Amsterdam: John Benjamins.

Gavarró, A., Lewandowski, W., & Markova, A. (2010). An approach to multiple interrogatives in child Bulgarian and Polish. In J. Costa, A. Castro, M. Lobo, & F. Pratas (Eds.), *Language acquisition and development: Proceedings of Generative Approaches to Language Acquisition 2009* (pp. 170–81). Newcastle: Cambridge Scholars Publishing.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, *25*, 407–54.

Guasti, M. T., Gavarró, A., De Lange, J., & Caprin, C. (2008). Article omission across child languages. *Language Acquisition*, *15*(2), 89–119.

Holmberg, A., & Roberts, I. (2014). Parameters and the three factors of language design. In M. C. Picallo (Ed.), *Linguistic variation in the minimalist framework* (pp. 61–81). Oxford University Press.

Huang, J. (1984a). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, *15*(4), 531–74.

Huang, J. (1984b). Phrase structure, lexical integrity, and Chinese compounds. *Journal of Chinese Language Teachers Association*, *19*, 53–78.

Huang, J. (1989). Pro-drop in Chinese: a generalized control theory. In O. Jaeggli & K. Safir (Eds.), *The null subject parameter* (pp. 185–214). Dordrecht: Kluwer Academic Publishers.

Huang, J., Li, A., & Li, Y. F. (2009). *The syntax of Chinese*. New York: Cambridge University Press.

Jones, B. M. (1999). *The Welsh answering system*. Berlin: Mouton de Gruyter.

Kuhl, P. (2010). Brain mechanisms in early language acquisition. *Neuron Review*, *67*(5), 713–27.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, *57*, 267–308.

Lacerda, R. (2017). *Information structure in child English: contrastive topicalization and the dative alternation*. Paper presented at the Boston University Conference on Language Development 41, Boston.

Lassotta, R., Omaki, A., & Franck, J. (2015). *Abstract knowledge of non-canonical word order by age 2*. Paper presented at the Boston University Conference on Language Development 39, Boston.

Lee, H. C. (2005). *On Chinese focus and cleft constructions*. Doctoral dissertation, Taiwan: National Tsing Hua University.

Lee, C., & Cho, S. W. (2009). Acquisition of the subject and topic nominals and markers in the spontaneous speech of young children in Korean. In C. Lee, G. B. Simpson & Y. Kim (Eds.), *The handbook of East Asian psycholinguistics, vol. III* (pp. 23–33). Cambridge University Press.

Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, *14*(3), 315–44.

Li, C., & Thompson, S. (1981). *Mandarin Chinese: a functional reference grammar*. Berkeley, CA: University of California Press.

Li, Y. H. (2006). Chinese *ba*. In M. Everaert & H. Van Riemsdijk (Eds.), *The Blackwell companion to syntax, vol. I* (pp. 374–468). Malden: Blackwell.

Liu, H. Y. (2016). *Also introduing arguments: the Mandarin* ba *construction*. Paper presented at the Meeting of the Canadian Linguistic Association, Calgary.

Lorusso, P., Caprin, C., & Guasti, M.-T. (2005). Overt subject distribution in early Italian children. In A. Brugos, M. R. Clark-Cotton, & S. Ha (Eds.), *Supplement: Proceedings of the 29th BUCLD*. Retrieved from <http://www.bu.edu/bucld/files/2011/05/29-LorussoBUCLD2004.pdf>.

Lü, S. X. (1980). *Xiandai Hanyu Babai Ci* [800 words of contemporary Chinese]. Beijing: The Commercial Press.

MacWhinney, B. (2000). *The CHILDES project: tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2015). *The CHILDES project: tools for analyzing talk*. Retrieved from <https://talkbank.org>.

Packard, J. (2000). *The morphology of Chinese: a linguistic and cognitive approach*. Cambridge University Press.

Phillips, B. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–42.

Pierce, A. (1989). *On the emergence of syntax: a crosslinguistic study*. Doctoral dissertation, Massachusetts Institute of Technology.

Reinhart, T. (1982). *Pragmatics and linguistics: an analysis of sentence topics*. Bloomington, IN: Indiana University Linguistics Club.

Rische, J., & Komarova, N. (2015). Regularization of languages by adults and children: a mathematical framework. *Cognitive Psychology*, *84*, 1–30.

Roberts, I. (2016). *Null subject parameters and determiner parameters*. Talk presented at the 1st SynCart workshop: From maps to principles: Syntactic cartography and locality. Chiusi.

Roberts, I., & Holmberg, A. (2010). Introduction. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric variation: null subjects in minimalist theory* (pp. 1–56). Cambridge University Press.

Roeper, T. (2007). What frequency can do and what it can't. In I. Gülzow & N. Gagarina (Eds.), *Frequency effects in language acquisition: defining the limits of frequency as an explanatory concept* (pp. 23–50). Berlin: Mouton de Gruyter.

Sakas, W. G. (2016). Computational approaches to parameter setting in generative linguistics. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 696–724). Oxford University Press.

Santos, A. L. (2004). How do children say 'yes' in European Portuguese? In J. Van Kampen & S. Baauw (Eds.), *Proceedings of Generative Approaches for Language Acquisition 2003*, vol. 2. (pp. 433–43). Utrecht: LOT Occasional Series.

Santos, A. L. (2009) *Minimal answers: ellipsis, syntax and discourse in the acquisition of Portuguese*. Amsterdam: John Benjamins.

Sugisaki, K. (2009). The acquisition of argument ellipsis in Japanese: a preliminary study. *Nanzan Linguistics*, *5*, 61–73.

Teng, S. H. (1979). Remarks on cleft sentences in Chinese. *Journal of Chinese Linguistics*, *7*, 101–12.

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press

Thornton, R., & Tesan, G. (2007). Categorical acquisition: parameter-setting in Universal Grammar. *Biolinguistics*, *1*, 49–98.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tsao, F. F. (1987). A topic–comment approach to the *ba* construction. *Journal of Chinese Linguistics*, *15*, 1–54.

Wang, Q., Lillo-Martin, D., Best, C., & Levitt, A. (1992). Null subject versus null object: some evidence from the acquisition of Chinese and English. *Language Acquisition*, *2*(3), 221–54.

Westergaard, M. (2009). *The acquisition of word order: micro-cues, information structure and economy*. Amsterdam: John Benjamins.

Wexler, K. (2011). Grammatical computation in the optional infinitive stage. In J. De Villiers & T. Roeper (Eds.), *Handbook of generative approaches to language acquisition* (pp. 53–118). Berlin: Springer.

Yang, C. (2002). *Knowledge and learning in natural language*. New York: Oxford University Press.

Yang, C. (2004). Universal grammar, statistics, or both? *Trends in Cognitive Science*, *8*(10), 451–6.

Yang, C. (2011). Computational models of language acquisition. In J. De Villiers & T. Roeper (Eds.), *Handbook of generative approaches to language acquisition* (pp. 119–54). Berlin: Springer.

Yang, C. (2012). Computational models of syntactic acquisition. *Cognitive Science*, *3*, 205–13.

Yang, C. (2013). Who's afraid of George Kingsley Zipf? Or: Do children and chimps have language? *Significance: Magazine of the The Royal Statistical Society*, December, 29–34.

Yang, C. (2016). *The price of productivity*. Cambridge, MA: MIT Press.

Zhou, J. (2001). *Pragmatic development of Mandarin speaking young children: from 14 months to 32 months*. Doctoral dissertation, The University of Hong Kong. Available from HKU Thesis Online (HKUTO).

Zhou, P. (2014). Children's knowledge of ellipsis constructions in Mandarin Chinese. *Journal of Psycholinguistic Research*, *43*, 421–45.

Zhuo, J. S. (2005). *Dramatized discourse: the Mandarin Chinese* ba*-construction*. Amsterdam: John Benjamins.

Cite this article: Zhu J, Gavarró A (2019). Testing language acquisition models: null and overt topics in Mandarin. *Journal of Child Language* **46**, 707–732. https://doi.org/10.1017/S0305000919000114

https://doi.org/10.1017/S0305000919000114 Published online by Cambridge University Press