

No evidence for an input frequency effect for children's learning of participant role markers in Japanese

Ben Ambridge

University of Manchester

ESRC International Centre for Language and Communicative Development (LuCiD)

Motoki Saito

Carl von Ossietzky Universität Oldenburg

Samuel David Jones

Bangor University

Tomoko Tatsumi

University of Liverpool

Kobe University

Kumiko Fukumura

University of Stirling

Ayuno Kawakami

University of Tokyo

Colin Bannard

University of Manchester

Abstract: One of the most crucial tasks facing language learners is learning to indicate 'who did what to whom' ("participant roles") in basic two-participant sentences (cf., *The dog chased the cat; The cat chased the dog*). The present study investigated this phenomenon using an elicited-imitation (sentence-repetition) paradigm with 38 native child learners (ages 4;0-4;11) of Japanese; a language that – at least for active transitive sentences – marks participant roles with the markers *-ga* (e.g., marking the *chaser*) and *-o* (e.g., marking the *chasee*). Constructivist input-based accounts predict that the rate at which children incorrectly produce *-ga*-marked forms in *-o* (OBJECT/PATIENT) contexts will be positively related to the relative input frequency of the *ga*- and *o*- form of the noun in question. In fact, a Bayes Factor analysis (with various supplementary tests) yielded no support for either this prediction or for the null hypothesis of no effect. We conclude with a discussion of the importance of publishing null and inconclusive results, in order to reduce bias in subsequent meta-analyses.

要旨：言語学習者が直面する極めて重要な課題として、基本的な二項文において「誰が誰に何をしたか」（参加者役割）を適切に示せるようになることが挙げられる（例：「犬が猫を追いかけた」と「猫が犬を追いかけた」の区別）。本研究では、日本語を母語とする4歳児（4歳0ヶ月～4歳11ヶ月）38名を対象に、誘発模倣（文の反復）課題を用いてこの現象を調査した。日本語は、少なくとも能動的な他動詞文においては、「～が」（追う側など）や「～を」（追われる側など）といった助詞を用いて参加者の役割を標示する言語である。インプットを基盤とする構築主義の説明では、本来「～を」（目的語／被動作者）とすべき文脈で子供が誤って「～が」を用いた形式を産出する割合は、当該名詞における「～が」形式と「～を」形式の相対的なインプット頻度と正の相関を示すと予測される。しかし実際には、ベイズ因子分析（および各種の補足的検定）の結果、この予測も、また「効果がない」とする帰無仮説も、いずれも支持されなかった。最後に、後続のメタ分析におけるバイアスを低減するため、帰無結果や決定的でない結果を公表することの重要性について論じる。

Keywords: Japanese; word-order; nominative/accusative marking; participant-role marking

Corresponding author: Ben Ambridge, Division of Psychology, Communication and Human Neuroscience, Coupland Building 1, University of Manchester, Manchester, M15 6FH. Email: ben.ambridge@manchester.ac.uk

ORCID IDs: Ben Ambridge: <https://orcid.org/0000-0003-2389-8477>; Motoki Saito: <https://orcid.org/0000-0003-0785-5813>; Samuel David Jones: <https://orcid.org/0000-0002-8870-3223>; Tomoko Tatsumi: <https://orcid.org/0000-0001-5590-016X>; Kumiko Fukumura: <https://orcid.org/0000-0002-1475-9304>; Colin Bannard: <https://orcid.org/0000-0001-5579-5830>

Citation: Ambridge, B., Saito, M., Jones, S.D., Tatsumi, T., Fukumura, K., Kawakami, A., & Bannard, C., (2026). No evidence for an input frequency effect for children's learning of participant role markers in Japanese. *Language Development Research*, 6(1), 408–447. <http://doi.org/10.34842/ldr2026-919>

Introduction

One of the most central questions in developmental science is how children acquire knowledge of their native language. Researchers from a variety of theoretical perspectives (e.g., Gertner et al., 2006; Naigles, 1990; Pinker, 1989; Tomasello, 2003) have long agreed that a particularly key aspect of language acquisition is learning to mark ‘who did what to whom’ in basic transitive sentences (e.g., *The dog chased the cat*; *The cat chased the dog*). Although this question has attracted a great deal of research attention, previous studies have generally focussed on English (which, as we will see shortly, is a rather atypical language in the relevant respects) or investigated how children balance already-acquired cues to meaning in comprehension (e.g., Bates & MacWhinney, 1987; Slobin & Bever, 1982). Outside of English, few studies have investigated how children use and acquire cues to participant roles in production¹.

For learners of languages such as English (i.e., word-order languages), learning to mark ‘who did what to whom’ is almost entirely a matter of learning a particular pattern of word order: [SUBJECT] [VERB] [OBJECT], which – at least for declarative sentences with basic action verbs – corresponds to [AGENT] [ACTION] [PATIENT]² (cf., *The dog chased the cat*; *The cat chased the dog*). The question of how learners of English acquire SVO word order is one of the most heavily researched in language acquisition. For example, Ambridge & Lieven (2011) summarized the findings of 20 experimental studies using tasks as diverse as elicited production (e.g., Tomasello & Brooks, 1998), weird word order (e.g., Akhtar, 1999), syntactic priming (e.g., Savage, et al., 2003), act-out (e.g., Childers & Tomasello, 2001), preferential looking (e.g., Gertner et al., 2006) and forced-choice pointing (Noble et al., 2011).

The question of how best to account for these findings has attracted a good deal of research interest. Constructivist accounts (e.g., Ambridge & Lieven, 2015; Tomasello,

¹ Elicited production studies of Polish (Dabrowska, 2004; Dabrowska & Szczerbinski, 2006; Krajewski et al., 2011) and Lithuanian (Saviciute et al., 2018) have investigated the acquisition of case marking in general, but – at least for some gender/declension classes – nominative and accusative marking is identical).

² In the child language acquisition literature (especially for English), acquisition of semantic [AGENT] [ACTION] [PATIENT] and syntactic [SUBJECT] [VERB] [OBJECT] word order are often conflated. For example, children’s success at correctly comprehending [AGENT] [ACTION] [PATIENT] sentences (e.g., *The duck is glorping the bunny*) is sometimes taken as evidence that they have acquired (at least the rudiments of) [SUBJECT] [VERB] [OBJECT] word order. The latter is broader because many SUBJECTS are not AGENTS, many VERBS are not ACTIONS and many OBJECTS are not patients (e.g., [SUBJECT *The book*] [VERB *cost*] [OBJECT \$5]). In the present study, all SUBJECTs are AGENTS, all VERBS ACTIONS, and all OBJECTs PATIENTs. Thus, while the present article discusses acquisition of Japanese “SUBJECT” and “OBJECT” marking (and of SOV/OSV word order), we adopt this familiar conflation purely for terminological convenience: The present data bear only on the acquisition of AGENT and PATIENT marking, not – strictly speaking – of wider SUBJECT and OBJECT marking.

2003) assume that children start out with rote-learned fixed or “frozen” phrases (e.g., *He’s pushing it*; *He’s eating it*), which are always produced in exactly the same form, and lexically-specific slot-and-frame patterns (e.g., *He’s [ACTION]ing it*), which allow for only limited productivity (since the “frame” – here *He’s...ing it* – is invariant across different uses).

Only later, according to constructivist accounts, do these semi-abstract slot-and-frame patterns develop into a fully abstract SVO construction. As evidence for this view, which assumes a gradual and heavily-input-based learning process, constructivist accounts point to the unevenness of children’s performance: In the syntactic priming, act-out, weird-word order and elicited-production studies above, children showed better performance for sentences with pronouns (e.g., *He’s pushing it*) than full noun phrases (e.g., *The dog’s pushing the car*). The claim is that the high frequency of input sentences of the form *He’s [ACTION]ing it* has enabled them to form an abstract slot-and-frame pattern.

Although children’s acquisition of participant-role marking in comprehension has been studied intensively (see Kolak et al., 2025, for a 21-language meta-analysis), production studies have largely been limited to English and other “word-order” languages to which it is typologically related, such as French (e.g., Matthews et al., 2007). This focus is unfortunate, given that, from a typological perspective, languages that convey ‘who did what to whom’ almost exclusively by means of word order are outnumbered (e.g., Comrie, 2013: Chapter 98) by languages that mark participant roles using noun case-marking (e.g., *Dog-NOM chased Cat-ACC*, where NOM and ACC indicate nominative and accusative markers, indicating the SUBJECT/AGENT and OBJECT/PATIENT respectively)³.

The aim of the present study is to begin to remedy this state of affairs by conducting an investigation of children’s acquisition, in production, of a case marker that denotes a participant role; an important topic, since such markers constitute the means to mark ‘who did what to whom’ in basic sentences for a large proportion of the languages of the world. Specifically, we investigate children’s acquisition of the Japanese ACCusative marker *-o* (sometimes written *-wo*), used to mark the direct OBJECT of a transitive verb; the PATIENT of an ACTION (e.g., Minashima, 2001; Miyagawa, 2003):

³ Although such languages almost always have a preferred, conventionalized word order, it is case marking – not word order – that unequivocally marks participant roles. For example, although SVO word order is conventional in Polish, *Dog-ACC chased Cat-NOM* means ‘the cat chased the dog’, not vice versa.

- (1) Neko(-ga) inu-**o** oshiteru
 Cat(-FOCUS) dog-ACC pushing
 'The cat is pushing the dog'

As we will see in more detail later, the precise function of the focus particle *-ga* is a topic of much debate amongst Japanese linguists. For our purposes, the important point is simply that *-o* unambiguously marks the OBJECT/PATIENT of the VERB/ACTION, even for sentences that depart from canonical Japanese SUBJECT OBJECT VERB word order:

- (2) Inu-o neko(-ga) oshiteru
 Dog-ACC cat(-FOCUS) pushing
 'The cat is pushing the dog'

Japanese is a particularly suitable language in which to investigate OBJECT case marking for three reasons. First, the *-o* marker is both *agglutinative* and *overt*; i.e., added to the end of the nouns as an individual syllable (or mora) that is readily observable. This contrasts with, for example, the Russian singular paradigm, in which accusative case marking is usually either absent altogether (for masculine nouns; e.g., *stol-NOM* vs *stol-ACC*; 'table') or is marked with a vowel change (for feminine nouns; e.g., *ruka-NOM* vs *ruku-ACC*, 'hand').

Second, Japanese children have been observed to make errors of commission, at least for sentences that use noncanonical OBJECT-*o* SUBJECT-*ga* VERB (PATIENT AGENT ACTION) word order (e.g., Hakuta, 1982). For example, when attempting to describe a scene in which a cat pushes a dog, instead of the correct target sentence (2), a child might produce

- (3) Inu(-ga) neko-o oshiteru
 dog(-FOCUS) cat-ACC pushing
 'The dog is pushing the cat'

failing to mark the OBJECT/PATIENT dog with the *-o* marker, and instead erroneously marking it with the focus particle *-ga*, and hence reversing her intended meaning. These errors appear to stem from the fact that AGENT-PATIENT *ga...o* sentences are considerably more frequent than PATIENT-AGENT *o...ga* sentences (Kuno, 1973, puts the ratio at 17:1) and/or to an AGENT-first bias that may have its roots in information structure (e.g., Abbot-Smith et al., 2017; Huang & Arnold, 2016; Huang et al., 2013; MacDonald, 2013). Furthermore, while *-o* is straightforwardly an OBJECT/PATIENT marker, *-ga* marks a wide variety of roles, including both SUBJECT (e.g., 1-3) and – for stative, unrealized or otherwise low-transitivity events – OBJECT, as in the following examples from Takano (2003):

(4) John-wa eigo-ga wakaru.
 John-TOPIC English-FOCUS understands
 'John understands English'

(5) John-wa dansu-ga deki-ru.
 John-TOPIC dancing-FOCUS is capable of
 'John is capable of dancing; John can dance'

Indeed, a corpus count of the nouns used in the present study found that while all *-o* forms were OBJECTS, around 13% of *-ga* forms were OBJECTS (4-5) rather than SUBJECTS (1-3). It makes sense, therefore, that children struggle to learn both the functions of *-ga* and when to mark OBJECTS with *-o* (i.e., when they are objects of highly transitive actions), as opposed to *-ga* (i.e., when they are not).

The third reason why Japanese is a particularly suitable language in which to investigate object case marking is that Japanese nouns vary considerably as to the frequency with which they occur with each case marker. As illustrated by the corpus analysis we conducted for the present study (discussed in more detail below), some occur much more frequently with *-ga* than *-o*, others reverse this pattern, and others are equi-biased.

Together, these properties of Japanese allow us to design a study that tests a central claim of constructivist accounts of language acquisition: unevenness of development. In the present domain, the constructivist prediction is as follows: Because children's knowledge of *-o* is initially not abstract, but tied to particular case-marked noun forms stored in vocabulary, young children's ability to produce the *-o* rather than *-ga* form in contexts in which the former is required will be positively related to the relative input frequency of the *-o* versus *-ga* form of that noun (e.g., *inu-o* vs *inu-ga*), hence referred to as "input bias". The flip side of this coin is that nouns that are heavily biased away from *-o* and towards *-ga* are predicted to cause errors in which children produce *-ga*-marked nouns in contexts that require *-o* (see Ambridge et al., 2015, for a review of the effect of competing erroneous forms in constructivist accounts). A complication is that while this prediction clearly holds for children in the earliest stages of acquisition (around 2-3 years), it may not necessarily hold for older children, who – at least under some versions of a constructivist account – are assumed to have built productive representations that abstract away from the original exemplars. We return to this issue in the Discussion.

The question of whether rival, non-constructivist accounts predict these types of lexical input effects is also unclear. Early-abstraction accounts, which assume "predispositions to link semantic and structural abstractions" (Gertner et al., 2006: 690), would seem to predict that the participants of the present study, who are

relatively old in terms of learning participant-role marking, will be well beyond the stage characterized mainly by lexical item-based knowledge, and will have formed abstractions that allow them to correctly mark the role of any participant. For example, in Gertner et al.'s (2006) study of English, children aged 1;9 were able to interpret cues to participant-role marking with completely novel (i.e., zero-frequency) verbs; a finding that the authors interpreted as evidence against a lexical (i.e., constructivist) account (see p.689 "We consider this unlikely...").

Similarly, words-and-rules accounts (e.g., Pinker, 1999) assume that once a regular morphological marker has been learned (a "eureka moment"; p.202), children are able to apply it across the board, and so would not seem to predict lexical effects. The situation is also similar for accounts based on minimalist syntax, such as Wexler (1998)⁴, which assume "Very Early Knowledge of Inflection", that "the two-word stage [is] the point at which parameters and inflectional properties are known", at which "NOM/ACC features are known". To be sure, these accounts do not explicitly *rule out* lexical frequency effects (or mention them at all). But the very point of an abstraction (Gertner et al., 2006), of a formal rule (Pinker, 1999) or of knowledge of parameters and inflectional properties would seem to be to explain how children can correctly inflect low-frequency (or novel) items. In sum, while lexical frequency effects are a core prediction of constructivist accounts, these various types of early-abstraction accounts (see also Clahsen et al., 1992; Fodor, 1998; Hoekstra & Hyams, 1998; Rispoli et al., 2009; Schuler et al., 2016) could explain such effects only as an add-on to the core learning mechanism (i.e., abstract knowledge of case marking plus faster and/or more accurate processing for frequent forms). This issue of processing and performance limitations is again one to which we will return in the Discussion.

Although, to our knowledge, the present study is the first to investigate input-frequency effects in the acquisition of Japanese case-marking, a number of previous studies of Japanese have investigated either frequency effects or case-marking separately.

Considering, first, comprehension studies looking at the acquisition of Japanese case-marking, Matsuo et al., (2012) found that children aged 2;4 were able to use syntactic frames to infer the meaning of novel verbs; for example that, when given (in translation) *The duck is glorpung the bunny* and *The duck and the bunny are kradding*, children were able infer that *glorpung* has a transitive meaning (e.g., similar to

⁴ For example, Wexler (1998: 51) formulates the rule for SUBJECT marking as: "Both AGRS and TNS have a D feature which must be eliminated by checking against the D-feature of a DP which raises up for checking". The grammatical details are not important for our purposes; but what is crucial is that the rule is formulated in such a way that it applies to any DETERMINER PHRASE (DP), and indeed makes no reference to the lexical content of the DP (e.g., *The man*, *The boy*, *John* etc.)

squashing) and *kradding* an intransitive meaning (e.g., similar to *dancing*). Crucially, however, they were able to do so only when the arguments (e.g., *the duck* and *the bunny*) were marked with *-ga* and *-o*. In a modified replication, Suzuki showed that children succeeded with these types of sentences (i.e., with *-o* marking the PATIENT/OBJECT), even when the AGENT/SUBJECT was omitted altogether). Together these studies show that even by age 2;4 – much younger than those who took part in the present study – children are adept with *-ga* and *-o* marking, at least in comprehension, and with canonical (S)OV sentences.

When it comes to noncanonical OSV sentences, even older children struggle. Kolak et al.'s (2026) 21-language meta-analysis of comprehension studies found that children (ranging from 3;4 to 10;0) showed only 66% correct performance for noncanonical sentences (mainly OVS or OSV) as compared to 87% for canonical sentences (mainly SVO or SOV). Indeed, even adults showed a small but statistically significant disadvantage for noncanonical sentences (92% vs 95%). The findings for the three Japanese studies included in this meta-analysis (Hakuta, 1982, for children; Shigenaga, 2014 and Yano & Koizumi, 2018, for adults) were in line with this general pattern. In production, too, Japanese learning children struggle at first to mark case in noncanonical OSV sentences, with error rates of 43% and 30% observed for 2-6 and 9-year-olds respectively (Hakuta, 1982; Murao et al., 2017)

In summary, then, the previous empirical literature suggests that mastery of Japanese AGENT/SUBJECT *-ga* and PATIENT/OBJECT *-o* marking is a protracted process. On the one hand, even 2-year-olds know a good deal about the functions of these case markers. On the other, even adults do not show flawless performance with case-marked OSV sentences in comprehension studies. This makes it very difficult to say whether the 4-year-old participants of the present study would be expected to have acquired productive use of these markers; but, certainly, we would not expect these children to have mastered them fully.

With regard to lexical frequency effects, we are aware of only two papers that have investigated these effects in the context of Japanese morphological acquisition. In a series of four studies with children aged 2;7-5;8, Tatsumi et al., (2017; 2018) investigated the acquisition of (a) simple versus stative verb past-tense forms (*-ta* vs *-teta*) and (b) past versus non-past forms (*-ta* vs *-ru*). In both cases, an input-bias measure reflecting the relative frequency of each verb in (a) *-ta* versus *-teta* and (b) *-ta* versus *-ru* form significantly predicted the rate at which children erroneously produced (a) *-ta* in contexts that required *-teta*, (b) *-ta* in contexts that required *-ru*, and vice-versa. Although the studies of Tatsumi et al. (2017; 2018) investigated the acquisition of verb rather than noun morphology, they are otherwise broadly similar in most respects to the study reported here, including with regard to the calculation of the input-bias predictor and the age of the children studied. Thus, there is every

reason to expect the input-frequency effect observed in these previous studies to also be observed here.

Method

Preregistration

We consider the present study to be partially pre-registered, as follows. An earlier version of this study was accepted in principle as a Registered Report at *Developmental Science* and publicly uploaded to the Open Science Framework (<https://osf.io/cq4zv>). This earlier version specified a target population of “monolingual Japanese-speaking children aged 2;6-3;6”, with the sample size to be determined using a Sequential Bayes Factor (SBF) analysis (Schönbrodt & Wagenmakers, 2017). However, following restrictions due to the Covid 19 pandemic, the study was converted to remote administration via Zoom (with a live experimenter). Because we felt that an online study would be unworkable with children aged 2;6-3;6, we changed the age range to 4;0-5;0. Due to uncertainties around the feasibility of recruiting children for the online study, particularly given the pandemic, we also abandoned the Sequential Bayes Factor approach, and instead simply recruited as many participants as possible in the time available (final $N=38$, with an additional 12 children tested but subsequently excluded). Given that these children were considerably older, we also changed the distractor element of the task (described below) from a silent 1-second waiting period to a 5-second task in which children were asked to repeat three digits shown onscreen and spoken live by the experimenter. In all other respects – including the methods, materials and preregistered R syntax for data analysis – the study remains identical to the Registered Report version. While these changes are too major to allow us to proceed with the Registered Report, we consider this previous publicly shared version to constitute a relatively robust preregistration document. Accordingly, the remainder of the present Method section is taken verbatim from the Registered Report, except for (a) the inclusion of changes necessitated by the differences described above (b) rewriting of the methods from future into past-tense (purely as a convenience to the reader) and (c) adding changes requested by two anonymous *Language Development Research* reviewers (all clearly indicated as such).

Overview of Method

The aim of this study is to investigate whether young Japanese-speaking children’s knowledge of OBJECT-*o* case marking constitutes knowledge of individual case-marked nouns. In order to investigate this question, we asked children to repeat OSV sentences which – because they reverse typical Japanese SOV order – unambiguously require *wo...ga* marking in order to convey the correct meaning. (When the intended meaning is clear without them – e.g., from the use of typical SOV word order – these markers are often omitted in informal speech). Importantly, because all sentences

use highly actional transitive verbs, correct OBJECT/PATIENT marking unambiguously requires *-o*, rather than *-ga* (which, for such sentences, unambiguously marks the SUBJECT/AGENT). We also included SOV sentences, essentially as fillers, since we anticipated close-to-ceiling performance on these sentences. **We investigated children’s acquisition of ACC *-o* marking, by varying the *o*-vs-*ga* bias of the first noun in OSV sentences**, using nontarget (filler) nouns that are equibased on this measure:

TARGET-<i>o</i> NONTARGET- <i>ga</i> VERB	OSV sentence	
NONTARGET- <i>ga</i> TARGET-<i>o</i> VERB	SOV sentence	(control)

The target nouns were chosen to vary with regard to their relative input frequency in *o*-vs-*ga* form but were presented exclusively in *-o* form. The nontarget nouns were chosen to be equi-biased with regard to their relative input frequency in *o*-vs-*ga* form. If, as constructivist accounts predict, young children’s knowledge of case marking is indeed tied to knowledge of individual case-marked forms, then – across nouns – the frequency with which children produce correct OBJECT-*o* marking versus incorrect OBJECT-*ga* marking (with all other responses discarded as missing data⁵) will be related to an input-bias measure that reflects the frequency of each target noun with *o*-vs-*ga* marking in child-directed speech (e.g., the relative input frequency of *banana-o* vs *banana-ga*).

A repetition (elicited imitation) paradigm was chosen to ensure that children were attempting to produce (a) OSV responses rather than the more common SOV, and (b) overt case-marked forms, since *-ga* and *-o* are often omitted in informal speech (though, in Japanese culture, a school-based “test” conducted by an unfamiliar experimenter would normally be considered a relatively formal context). This paradigm is also less demanding than elicited production for young children. Although intuitively it might seem that children are likely to achieve ceiling performance by mere “parroting”, previous studies (e.g., Bannard & Matthews, 2008) have shown this not to be the case (see e.g., Ambridge & Rowland, 2013; Lust et al., 1996, for reviews). Provided the utterance to be repeated is sufficiently long and/or complex – or an intervening distractor task is used – children are not able to store the sentence verbatim and instead recreate it using normal language production mechanisms (indeed, it is often argued that even adults generate sentences anew during recall; e.g., Potter & Lombardi, 1998).

⁵ Though, as explained in more detail below, we also include an exploratory non-preregistered analysis which also includes omission of *-o* – as opposed to, for the main analysis, incorrect replacement of *-o* with *-ga* – as an error.

Ethics

The study was approved by the ethics committee of the University of Liverpool. Since we did not have a Japanese partner university with a suitable ethics committee, an independent local ethics review was generously conducted by a Professor of Ethics at a Japanese University, Shunzo Majima of Hokkaido University (we thank an anonymous reviewer for assistance with clarifying the language around this step).

Participants

Participants were 50 monolingual Japanese-speaking children who were reported by their parents as not displaying any apparent language difficulties. Children ranged in age from 48 months (4;0) to 59 months (4;11), with a mean of 53 months ($SD=3.66$). Children participated via Zoom with their parents present, and parents gave informed written consent. Children gave verbal assent. Of the original sample of 50 children, 12 were excluded for not providing any scorable responses (as described below). We retained from the Registered Report the criterion to exclude and replace any child who did not produce a scorable response (defined below) for at least 50% of the test trials.

Design and Materials

Nouns were selected from the input portion of a composite corpus created from all of the Japanese corpora available on CHILDES (MacWhinney, 2014): the Hamasaki, Ishii, MiiPro, Miyata, Noji, Ogawa, Okayama, Ota, Paido Japanese, Stanford Japanese, and Yokoyama corpora.

For each noun in the composite input corpus ($N = 2022$) we obtained the frequency of the relevant noun in *-o* and *-ga* form and calculated a measure of *o*-vs-*ga* bias: the direction-corrected log-transformed p value of the Fisher-Yates exact test (e.g., Stefanowitsch & Gries, 2003). This statistic reflects the extent to which, relative to all other nouns in the corpus, the noun in question is biased towards *-o* versus *-ga* (greater bias = larger negative value) or *-ga* versus *-o* (greater bias = larger positive value). Table 1 illustrates the calculation of this *input-bias* predictor for *Kippu*, ‘ticket’, the noun with the largest *-o* versus *-ga* bias of those included in the study.

Table 1. Calculation of the input-bias predictor (p value of Fisher-Yates exact test) for the *o*-biased noun, *kippu*, ‘ticket’

Items	+ <i>ga</i>	+ <i>o</i>
kippu, ‘ticket’	(a) 9	(b) 26
all other nouns	(c) 20227	(d) 6084

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}} = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{a! b! c! d! n!}$$

For *kippu*, ‘ticket’, the very small p value (2.22E-10) and corresponding large (direction-corrected) negative log-transformed value (-22.25), reflect the fact that the ratio of *-o* to *-ga* forms for this noun (roughly 3:1) is substantially larger than for the other nouns in the corpus (roughly 1:3).

We selected as target nouns the 10 nouns with the largest *o*-vs-*ga* bias, and the smallest *o*-vs-*ga* bias (i.e., the largest *ga*-vs-*o* bias), excluding nouns that refer to (pseudo-) animates (e.g., people, animals, and vehicles, which are considerably more plausible as SUBJECTs than OBJECTs), to abstract concepts that are difficult to illustrate in animations (e.g., *care*, *health*), to concepts that are unfamiliar to young children (e.g., *seal*), and generics (e.g., *thing*). We also excluded *te*, ‘hand’, since it constituted an outlier with regard to its degree of *o*-bias (direction-correct log p value of -152.468; cf.. values in Table 2). The use of an extreme-groups approach (e.g., Preacher et al., 2005) is justified by the need to restrict the study to 40 sentence repetition trials (20 target nouns, each in an OSV and SOV sentence); the upper limit of what we felt would be tolerable for young children (e.g., Bannard & Matthews, 2008, used 32 sentence repetition). The selected target nouns, together with their score on the input-bias measure, are shown in Table 2.

We selected as nontarget nouns 20 equi-biased nouns, nouns with Fisher-Yates p values of >0.999, following the same constraints as for the target nouns, and focussing on nouns with relatively high overall frequency. These nontarget nouns are equi-biased not in the sense that they are *equally* frequent in *o*-vs-*ga* form, but in the sense that their bias for *o*-vs-*ga* forms is virtually identical to that observed in Japanese child-directed speech (or, at least, our sample thereof) as a whole. It is important to acknowledge at this point that the operationalization of an “equi-biased” noun is not uncontroversial. An anonymous reviewer suggested that it may be more appropriate to define an equi-biased noun simply as one that appears in *-ga* and *-o* form with roughly equal frequency in the corpus. Our position is that because *-o* forms are considerably less frequent than *-ga* forms in the language as a whole, nouns with approximately equal raw frequency in *-o* and *-ga* form in fact show a considerable bias in favour of *-o* form, relative to Japanese in general, and thus are not “equi-biased” in the relevant sense. But we acknowledge that the issue is controversial, and we would welcome future research along the lines of the present study that instead employs raw-frequency-equi-based nouns as nontarget nouns.

Table 2. Target nouns biased towards *o*-vs-*ga* form (negative values) and *ga*-vs-*o* form (positive values), and equibiased nontarget nouns

Noun	Romaji	English	Noun_GA	Noun_O	Other_GA	Other_O	Input_Bias	Noun_ID
切符	kippu	ticket	9	26	20217	6074	-22.247	o_bias_01
紙	kami	paper	33	42	20193	6058	-20.855	o_bias_02
荷物	nimotsu	baggage	13	28	20213	6072	-20.670	o_bias_03
ご飯	gohan	rice	48	43	20178	6057	-14.437	o_bias_04
靴	kutsu	shoes	17	23	20209	6077	-12.690	o_bias_05
絵	e	picture	27	29	20199	6071	-12.585	o_bias_06
ドレス	doresu	dress	1	10	20225	6090	-12.467	o_bias_07
本	hon	book	54	43	20172	6057	-12.241	o_bias_08
パン	pan	bread	20	23	20206	6077	-10.493	o_bias_09
服	fuku	clothes	13	17	20213	6083	-9.458	o_bias_10
雨	ame	rain	106	4	20120	6096	17.665	ga_bias_01
花	hana	flower	36	4	20190	6096	2.838	ga_bias_02
耳	mimi	ear	33	4	20193	6096	2.521	ga_bias_03
骨	hone	bone	17	1	20209	6099	2.369	ga_bias_04
電池	denchi	battery	38	5	20188	6095	2.293	ga_bias_05
レモン	remon	lemon	10	0	20216	6100	2.039	ga_bias_06
風船	fūsen	balloon	18	2	20208	6098	1.636	ga_bias_07
機械	kikai	machine	19	2	20207	6098	1.634	ga_bias_08
鼻	hana	nose	60	12	20166	6088	1.559	ga_bias_09
コーヒー	kōhī	coffee	8	0	20218	6100	1.554	ga_bias_10
牛乳	gyūnyū	milk	10	3	20216	6097	0	equi-bias_01
バナナ	banana	banana	13	3	20213	6097	0	equi-bias_02
チーズ	chīzu	cheese	8	2	20218	6098	0	equi-bias_03

月	tsuki	moon	3	0	20223	6100	0	equi-bias_04
蜜柑	mikan	mandarin	17	5	20209	6095	0	equi-bias_05
ニンジン	ninjin	carrot	7	2	20219	6098	0	equi-bias_06
おにぎり	onigiri	riceball	4	1	20222	6099	0	equi-bias_07
時計	tokei	clock	22	6	20204	6094	0	equi-bias_08
オレンジ	orenji	orange	4	1	20222	6099	0	equi-bias_09
机	tsukue	desk	4	1	20222	6099	0	equi-bias_10
スプーン	supūn	spoon	4	1	20222	6099	0	equi-bias_11
大根	daikon	radish	3	0	20223	6100	0	equi-bias_12
傘	kasa	umbrella	9	2	20217	6098	0	equi-bias_13
シュー	shū	cream puff	0	0	20226	6100	0	equi-bias_14
チューリップ	chūrippu	tulip	4	1	20222	6099	0	equi-bias_15
パイナップル	painappuru	pineapple	4	1	20222	6099	0	equi-bias_16
扇風機	senpūki	fan	0	0	20226	6100	0	equi-bias_17
ティッシュ	tisshu	tissue	2	0	20224	6100	0	equi-bias_18
ミキサー	mikisā	mixer	0	0	20226	6100	0	equi-bias_19
扉	tobira	gate	5	1	20221	6099	0	equi-bias_20

Finally, we selected five high-frequency, highly-actional transitive verbs, chosen to be familiar to young children and easy to depict in a simple animation: 持ち上げる, (mochiageru, ‘lift’), 引っ張る (hipparu, ‘pull’), 押す (osu, ‘push’), 転がす (korogasu, ‘roll’) and ひっくり返す (hikkurikaesu, ‘turn over’). All verbs were always presented in progressive nonpast form (‘is [VERB] ing’) form (持ち上げている, 引っ張っている, 押している, 転がしている, ひっくり返している) in the experiment.

Inspection of the participant roles of instances of these 40 nouns in the input corpus revealed that while all *-o* forms were indeed clear PATIENT/OBJECTs, around 13% of *-ga* forms were not clear AGENT/SUBJECTs, but OBJECTs of stative, unrealized or otherwise low-transitivity events (e.g., 4-5). In the context of the present study, this is not a bug but a feature, as it increases the likelihood of the crucial errors in which children substitute *-ga* for *-o* in OBJECT contexts, particularly for nouns biased away from *-o* and towards *-ga*. However, at the suggestion of two anonymous reviewers, we later added supplementary analyses that exclude non-SUBJECT uses of *-ga* to address this issue.

The 20 target nouns, the 20 nontarget nouns, and the 5 verbs (in present progressive *-teru* form) were used to construct – for each child individually – 20 OSV sentences ([**TARGET**]-*o* [NONTARGET]-*ga* VERB). These 20 OSV sentences were constructed randomly for each child, such that each noun – target and nontarget – appeared exactly once, and each verb exactly four times. A further 20 SOV sentences ([NONTARGET]-*ga* [**TARGET**]-*o* VERB) were constructed in the same manner. Note that, because SOV sentences follow canonical Japanese *ga...o* word order we anticipate close to ceiling performance, and exclude these sentences from the primary statistical analysis. However, they play an important role as fillers: If children were presented only with OSV sentences, they could adopt a task-based strategy of always marking the first noun with *-o*, even if they had little or no understanding of case marking.

Additional non-preregistered exploratory predictors

Recall that the main *ga-vs-o* bias predictor, as set out above, was the (scaled and centred) direction-corrected log-transformed *p* value of the Fisher-Yates exact test (see the Input_Bias column of Table 2). Recall also that this *p* value was calculated from the raw frequency with which each noun appears in *-ga-* and *-o* form in the input corpus. However, as noted by two anonymous reviewers, basing this calculation on the raw frequency of the relevant *ga-* and *o-* forms, *regardless of the syntactic role and position of these nouns*, may result in an inappropriate operationalization of input-bias for two reasons.

First, in the present study, which featured only high-transitivity two-participant events, *-ga* always marked AGENT SUBJECTs. In real life, however, *-ga* marks a variety

of roles, including even – for low-transitivity events – OBJECTs (see examples 3-5 in the Introduction). We therefore obtained a new set of *-ga* counts (using the same corpora and coders) including only clear AGENT-SUBJECT uses (i.e., those for which both coders agreed).

Second, in the present study – at least, for the crucial OSV sentences – all nouns marked with *-o* (which unambiguously marks OBJECTs) occurred at the start of the clause. In real life, however, because default Japanese word order is SOV, *-o* forms usually appear later in the clause. We therefore obtained a new set of *-o* counts (using the same corpora and coders) including only those that clearly occurred at the start of the clause (again, requiring agreement by both coders).

These new predictors were then used in additional exploratory analyses. Because we have both old and new versions of both the *-ga* and *-o* counts, this brings the total number of input-bias predictors, and – consequently – analyses, to four:

1. **Main preregistered analysis:** Original (raw number) *-ga* counts; original (raw number) *-o* counts
2. New (AGENT-SUBJECT-only) *-ga* counts; original (raw number) *-o* counts
3. Original (raw number) *-ga* counts; New (clause-initial-OBJECT-only) *-o* counts
4. New (AGENT-SUBJECT-only) *-ga* counts; New (clause-initial-OBJECT-only) *-o* counts

In each case, an input-bias predictor was calculated from the *p* value of the Fisher-Yates exact test in the same way as already described above for the main, preregistered analysis.

Procedure

All trials were presented using a custom-written script for the Processing software package (www.processing.org), which automatically generates animations and target sentences using supplied still pictures and audio files of the relevant nouns in *-o* and *-ga* form. The script and all associated materials can be downloaded from the project OSF site.

Before the main test session, children completed a noun training session to ensure that they labelled each of the items pictured in the target sentences with the relevant nouns. The experimenter showed each child a sheet containing all target and nontarget nouns and elicited from the child a label for each in turn. If a given child did not produce the target noun for any object, the experimenter supplied the target noun, asked the child to repeat it, and then restarted the training session. This procedure was repeated until the child produced the entire set with no errors or

corrections. Any child who did not meet this criterion was to be excluded and replaced, though, in fact, all 50 children who were tested did meet this criterion.

Next, children completed a training session designed to ensure that they understood the “parrot game” of repeating exactly the sentence spoken by an onscreen robot, with a short intervening distractor task. Children completed four training trials, all using intransitive sentences in which the ACTOR is marked with the focus particle *-wa* (to avoid providing children with specific practice on *o-* or *-ga*): *Otokonoko-wa waratteiru/ utatteiru/ mawatteiru/ taoreteiru* ‘The boy is laughing / singing / turning / falling down). The program generated the audio sentences on the fly by combining pre-recorded audio files of each noun in *-o* and *-ga* form. This procedure was used to avoid a “clever Hans” effect whereby an experimenter who recorded entire sentences (or spoke them live) could unwittingly adopt a more natural intonation for utterances which include more plausible, higher-frequency *-o...-ga* (or *-ga...-o*) combinations. This resulted in somewhat stilted audio recordings (by design, since the aim is to neutralize the naturalness of the speech contour), which was attributed to the fact that the sentences are spoken by a robot (shown onscreen).

For each training trial (see Figure 1), the program displayed an animation depicting the scene and then played the target sentence in audio form, alongside an animation of a talking robot. Next, the child completed a 5-second distractor task which involved repeating three digits shown on an otherwise-blank black screen and spoken live (via Zoom) by the experimenter. Given the relatively advanced age of the children, a relatively long and complex distractor task (cf., the 1-second silent wait envisaged for children aged 2;6-3;6) was required in order to force children to attempt to reconstruct (rather than simply “parrot”) the target sentence. At the end of the 5-second distractor task, a beep sound was played, indicating that the child should begin the repetition attempt (with the animation shown again, in order to aid memory). For the first training trial, the experimenter elicited this repetition explicitly (“Now you have to say what the robot just said”) and continued to do so on the second and third training trial where necessary. On the final training trial, the experimenter elicited the sentence repetition simply by, on the bleep, turning her head towards and looking at the child; the procedure to be followed for the subsequent test sentences. Any child who did not spontaneously repeat the target sentence on the final training trial was to be excluded before beginning the test phase, though in fact no children were excluded on this basis. However, following the preregistration set out in the Registered Report, 12/50 children were excluded for failing to provide scorable responses (defined below) for at least 50% of the test trials.

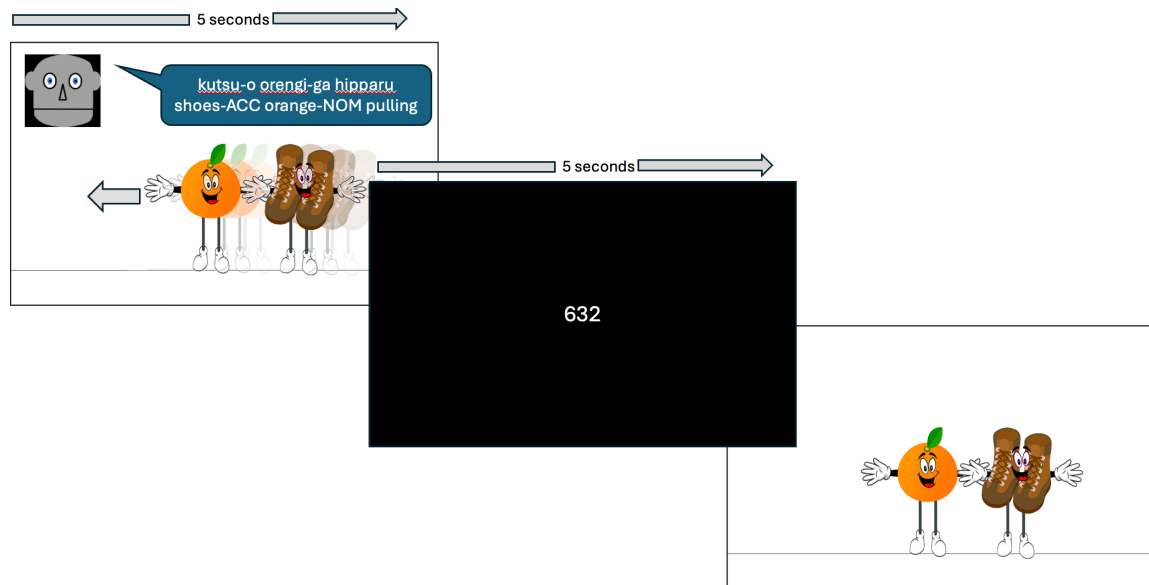


Figure 1. Structure of a test trial

The test trials followed the same procedure as the training trials, except that if the child did not attempt to repeat the target sentence, the experimenter used the prompt “Remember?” once only, before moving on to the next trial. All 40 test trials (20 target+nontarget noun pairs, each in one OSV and one SOV sentence) were presented in random order in a single session. The preregistration set out in the Registered Report specified that we would exclude and replace any child for which >10% (i.e., five or more trials) are disrupted by a technical error, an experimenter error or interruption from a third party, but no such exclusions were necessary.

Animations were generated on the fly by the Processing script, which moves supplied image files in predetermined paths, corresponding to the relevant verb/action. Each image file depicts an inanimate object, corresponding to the relevant target or nontarget noun, anthropomorphized by the addition of a standardized feature set – identical for each object – comprising arms, legs, eyes and mouth. Anthropomorphization was necessary in order to ensure that the inanimate objects could reasonably be construed as performing actions on one another. Most of the actions chosen (as highly transitive, familiar to children and simple to animate) – *lift*, *pull*, *push*, *roll* and *turn over* – require hands to initiate, and so would have been very difficult to illustrate with non-anthropomorphized characters. All nouns (as opposed to just the AGENT/SUBJECT nouns) were anthropomorphized in this way, in order to ensure that anthropomorphization was not a cue to participant roles.

Pilot studies

In June 2018, we completed an initial pilot study with five children aged 2;6-3;6 (the original target age). This version of the task was as described above, but used a pre-final set of noun stimuli, and ten different verbs (the five above, plus *drop*, *kick*, *pat*, *put down*, and *throw*), and also included a 3-digit repetition task as a distractor before the elicited sentence repetition. This version proved much too difficult for children of the target age, and all terminated the experiment early, with very few successful repetitions.

In August 2018, we completed a second pilot study with six children aged 2;6-3;6 (the original target age), replacing the distractor task with a 1-second delay, retaining the pre-final set of noun stimuli and ten verbs. These data can be downloaded from <https://osf.io/j7azp/>. In summary, all participants produced at least one scorable sentence repetition containing a *ga*-for-*o* or *o*-for-*ga* substitution (P1=7, P2=13, P3=15, P4=10, P5=6, P6=6), and all completed all 40 trials. We did not conduct any statistical analyses relating these errors to the input predictor. As a result of this successful second pilot, we settled on the method outlined in the original Registered Report, which is identical except for the use of a different set of noun stimuli, and of five rather than ten verbs: *Drop*, *kick*, *pat*, *put down*, and *throw* were dropped as the verbs that seem to cause the most difficulty for children, presumably because they are relatively difficult to illustrate in animations with characters that do not have moving limbs (necessary in order to allow animations to be generated on the fly by the Processing script). However, as noted above, this version was abandoned due to restrictions necessitated by the Covid 19 pandemic, and replaced with the online version described above.

Coding

The experimenter recorded children's responses by hand, and subsequently checked the transcriptions against audio recordings. In order to be retained as a scorable response, a repetition attempt had to include the two nouns (target and nontarget), each marked with *-ga* or *-o*, and the target verb in the same order as in the experimenter's target sentence (i.e., OSV or SVO), with no intervening linguistic material (other than fillers such as "umm"). Additional material produced before or after the target sentence (e.g., "Well...", "Here...", "It was...") was ignored.

The requirement that both nouns be marked with either *-ga* or *-o* in order for the utterance to be retained in the statistical analysis is a vital one, given the optionality – in informal spoken speech – of the relevant markers. This requirement ensures that we included only those trials for which children understand that the "aim of the game" is to produce utterances in which both nouns bear a case marker. The strong test of the input-frequency prediction at issue relates to children's production of *-ga*

for *-o* substitution *errors*. Any by-noun correlation between the input and children's performance in mere *provision versus omission* of *-o* would not demonstrate a lexical input effect: Such a correlation could arise simply because children and adults have a similar understanding of the types of nouns for which case marking is typically (not/) required for discourse-pragmatic reasons. For this reason, it is crucial to calculate the input-bias predictor in terms of *o*-vs-*ga* marking, rather than provision versus omission of *-o* (we thank an anonymous reviewer of the original Registered Report for helping to clarify this point). Nevertheless, because we were concerned that some readers might view this criterion as too stringent, we also added – after the data were collected – a non-preregistered exploratory analysis with correct provision of *-o* (1) versus all error types, including omission (0) as the dependent variable.

For the main, preregistered analysis, scorable responses were scored as correct (1) if the target OBJECT/PATIENT noun was marked with *-o* and incorrect (0) if the target OBJECT/PATIENT noun was marked with *-ga*, regardless of whether the nontarget SUBJECT noun is marked with *-ga* or *-o* (though recall that it had to be marked with one or the other for the utterance to be retained as scorable).

Analysis Plan

The analysis plan in R syntax (unchanged from the Registered Report version) can be downloaded from <https://osf.io/j7azp/> (**Simulation5_Resub.R** in the zip file **Submission2 Simulations.zip**). We verified this syntax by running the analyses over simulated data from 100 participants, assuming, on the basis of previous studies (as discussed in more detail below), an effect size of $M = 0.31$, $SE = 0.08$ for the input-bias predictor.

Main analysis. The main analysis was a Bayesian mixed-effects model for OSV sentences only, with SOV sentences treated as fillers. All predictors were scaled into SD units and centred. Fixed effects were the input bias predictor (with normal prior of [0,0.33] as discussed below), and control predictors (all with normal prior of [0,1]) of age in months, noun length in mora, and trial number, with random effects of participant and target noun on the intercept. The model included by-participant random effects of the slope of input bias, noun length and trial number, and a by-target-noun random effect of the slope of age. This model was implemented using *brms* (Bürkner, 2017), with the following syntax:

```
prior = c(set_prior("normal(0,0.33)", class = "b", coef = "InputBias")
set_prior("normal(0,1)", class = "b"))
```

```
fit1=brm(formula = Response ~ InputBias + AgeMonths + LengthMora + TrialOrder +
InputBias*AgeMonths + (1+InputBias+LengthMora+TrialOrder|Participant) +
(1+AgeMonths|Noun), data=subset(Data, WordOrder=="OSV"), family=bernoulli(),
prior=prior, cores=4, warmup = 2000, iter = 5000, chains = 4, control = list(adapt_delta
= 0.99), sample_prior = TRUE)
```

For each population-level effect (Intercept, InputBias, AgeMonths, LengthMora, TrialOrder, InputBias*AgeMonths), we report the Estimate, the Estimated Error, the lower and upper 95 Credible Intervals, the Effective Sample Size and R^2 (potential scale reduction factor), as well as direction-corrected P_{MCMC} values.

Unlike their frequentist equivalents, Bayesian mixed-effects models are highly robust to convergence failure (we did not encounter any convergence failure in our simulated data). In the event of convergence failure – diagnosed as by a Gelman-Rubin convergence diagnostic (R -hat) value of > 1.1 – we planned to simplify the model according to the procedure outlined in Barr et al. (2013) for frequentist models; i.e., iteratively removing the random effect that explains the least variance until convergence is achieved. However, this turned out not to be necessary.

With regard to the predictor of interest (InputBias), Bayes Factors (BF10; Kass & Raftery, 1995) were calculated using the Savage-Dickey method (Wagenmakers et al., 2010), with a prior of $M = 0$, $SD = 0.33$, half-normal distribution, assuming a positive effect of input bias on correct production (we do not report Bayes Factors for the control predictors). This prior is based on the best estimate from four previous studies (see Table 3) that, on the surface, investigated a very different phenomenon to the present studies (Japanese children's acquisition of verb as opposed to noun markers), but, at an underlying level, share important conceptual similarities. In all four studies (two each reported in Tatsumi, et al., 2017, 2018), input-bias measures based on the Chi-square statistic (equivalent to the present Fisher-Yates exact measure) were used to predict the degree to which Japanese children would supply correct (1) or erroneous (0) morphological markers. These studies are also similar to the present study with regard to (a) the use of an extreme-groups approach, (b) the number of items per child and (c) the structure of the statistical models (random intercepts for participant and item, and by-participant random slopes for the input-bias predictor). We chose the prior from Tatsumi et al (2018), Study 1, since this is the larger, and hence more conservative prior (e.g., Jeon & De Boeck, 2017) of the two that used the same units as the present study.

Table 3. Findings from four previous studies used to set the Bayesian prior

Study	Age	N	Trials per Child	Missing data	β^*	SE(β)	Chi ²	p (Chi ²)
2017, S1	3;3-4;3	28	20	30%	0.36	0.29	1.58	0.21
2017, S2	3;5-5;3	30	20	23%	0.38	0.15	5.44	0.02
							95 CrI	pMCMC
2018, S1	3;2-5;8	22	40	12%	0.33	0.10	[0.15, 0.54]	<0.001
2018, S2	2;7-4;11	26	40	25%	0.31	0.08	[0.17, 0.47]	<0.001

*For the first two studies in this table, this beta value is unstandardized (units of natural-log-transformed chi-square input-bias predictor). For the final two studies, this value is standardized into *SD* units.

We did not make a specific prediction with regard to the interaction of age by input bias. One possibility is that the effect of input bias will decrease with age as children's representations become more abstract. However, at least some versions of a constructivist account (e.g., Bybee, 1985; Chandler, 2010; Croft, 2000) eschew stored abstractions in favour of an exemplar model whereby even adults' knowledge of inflectional morphology consists of knowledge of exemplars of individual inflected forms, with novel or low-frequency forms inflected on the basis of phonological analogy to these stored forms. Such accounts would not necessarily predict a developmental decrease in the effect of input bias, and may even predict an increase, as exemplar storage increases with age.

In order to estimate the number of participants that would be required in order to provide a Bayes factor at least six times in favour of the experimental hypothesis over the null or vice versa (our original intention as set out in the Registered Report), we used the R package *simr* (Green & MacLeod, 2016) to generate simulated data for 100 participants. For the main predictor of interest, input bias, we assumed a conservative effect size of $M = 0.31$, $SE = 0.08$ (the smaller of the two reported in Tatsumi et al, 2018). We also assumed a similar effect of age ($M = 0.31$, $SE = 0.06$), and a smaller negative interaction of input bias by age ($M = -0.155$, $SE = 0.04$), but effects of zero for verb length and trial order. We then submitted these data, and smaller randomly selected samples, to the Bayesian procedure outlined above. At each sample size, we ran 20 models, in order to obtain a mean Bayes Factor as well as an associated standard error. On the basis of these simulations, a sample size of $N = 100$ would yield a Bayes Factor (BF10) of 48.64 ($SE = 2.38$), $N = 80$ of 47.70 ($SE = 8.88$), $N = 60$ of 36.31 ($SE = 11.75$), $N = 40$ of 23.75 ($SE = 5.41$) and $N = 20$ of 6.14 ($SE = 1.19$). Recall that, in practice, we were able to include data from only 38 participants. Nevertheless, assuming a positive input-frequency effect of a magnitude similar to that observed in previous verb studies (Table 3), this sample size would yield a Bayes Factor of around 20 (i.e., 23.75 for $N = 40$).

Supplementary analysis. We also preregistered a supplementary analysis that includes both OSV and SOV sentences, in order to explore any differences between them. The statistical model was identical to the main analysis, except that target word order and its interaction with the input-bias predictor was included as an additional fixed-effect (and by-participant random slope). Because we did not have a firm basis for making a prior estimate, we report only *pMCMC* values and credible intervals, but no Bayes Factor analysis. Since this analysis revealed no effects with 95% credible intervals that did not span zero, we do not report it further here (the full model output can be found at <https://osf.io/j7azp/>).

Results

Figure 2 plots, on the y axis, the proportion of correct *-o* responses versus *-ga-for-o* substitution errors (recall that all other responses were excluded as missing data), against the input-bias predictor (x axis).

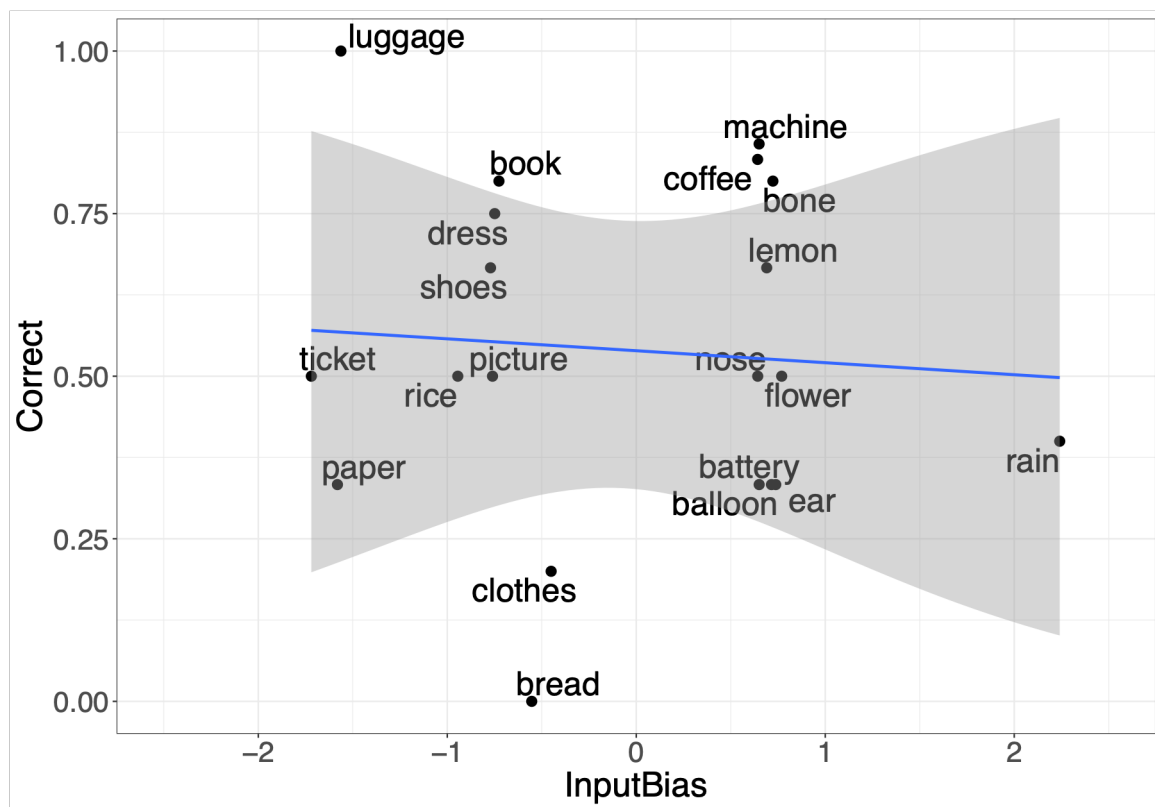


Figure 2. Proportion of correct *-o* responses versus *-ga-for-o* substitution errors (y axis) against the input-bias predictor (positive values = *ga*-bias, negative values = *o*-bias)

Because the input-bias predictor is coded such that negative values indicate that a given noun is biased towards *o-* (and away from *-ga*), the input-frequency prediction is of a negative correlation (i.e., a line sloping down to the right). Visual inspection of Figure 2 suggests that, although the slope is in the predicted direction, it is almost flat, and there is little evidence of any input effect.

Indeed, inspection of the model summary output (Table 4) reveals that the input-bias predictor has an estimated effect centred on zero; neither do any of the other fixed effect terms have credible intervals that exclude zero (the full output for this and all models can be found in the online appendix at the project OSF site).

Table 4. Output of the Bayesian model (Main analysis)

	Estimate	SE	95% CI Lower	95% CI Upper	R²
Intercept	1.00	0.98	-0.79	3.20	1
InputBias	0.00	0.31	-0.60	0.61	1
AgeMonths	0.89	0.67	-0.45	2.21	1
LengthMora	0.57	0.61	-0.66	1.81	1
TrialOrder	-0.71	0.66	-2.01	0.64	1
InputBias:AgeMonths	-0.44	0.70	-1.79	1.01	1

The Bayes Factor (BF₁₀) was 0.96⁶, indicating that the observed data are almost exactly equally likely under (a) the null hypothesis of no effect and (b) the alternative hypothesis of the expected input-bias effect (i.e., they are inconclusive). Thus, it is important to stress that the failure to find the predict effect here must not incorrectly be taken (as is so often the case), as positive evidence for the null hypothesis of no effect. The data are simply inconclusive.

Before moving on from the main analysis, it is important to note that children showed a relatively low rate of overall performance (57% correct); though the error-rate did differ substantially between children (by-participant intercept $SD = 2.40$, $SE = 1.47$), and indeed between nouns (by-participant intercept $SD = 1.56$, $SE = 1.03$); just not in a way that was related to the input-bias predictor. Presumably the reason that children found this task difficult is that correct performance with PATIENT-first OSV

⁶ Recall that this prior was based on the best estimate taken from previous studies (0.33). In order to check that the inconclusive Bayes Factor was not a result of selecting an overly narrow prior, we conducted a sensitivity check by re-running the analyses with an Input Bias prior of 0.54 (upper bound of 95CI from previous studies) and 1 (an arbitrary value). The Bayes Factors (see the OSF site for the full model output) were respectively 0.84 and 0.68 (i.e., fractionally more evidence for the null hypothesis, but still essentially inconclusive).

sentences requires children to override the processing patterns that they have developed across the (roughly) 95% of Japanese sentences that follow AGENT-first SOV order. Indeed, this rate of correct performance is not particularly surprising given that, even in (on-the-face-of-it easier) comprehension tasks, children aged 3;4-10;0 show only around 66% correct performance with noncanonical sentences crosslinguistically (Kolak et al., 2025).

Non-preregistered exploratory analyses

Recall from the Method section that, because we were concerned that some readers might view the exclusion criteria for the main analysis as too stringent, we also conducted – after the data were collected – a non-preregistered exploratory analysis with correct provision of *-o* (1) versus all error types including omission (0) as the dependent variable. (Recall that for the main, preregistered analysis, utterances that did not include either *ga-* or *o-* on both nouns were excluded). Again, this analysis yielded no evidence for an effect of input bias ($M = 0.13$, $SE = 0.20$, $95\% CI = [-0.27, 0.52]$, $BF = 0.75$), with the data fractionally more likely (at best, “anecdotal” evidence) under the null than the experimental hypothesis (i.e., the observed data are roughly 1.3 times as likely under the null than the experimental hypothesis).

One might also argue that, for this new analysis, the more appropriate dependent measure is the simple input frequency of *-o* forms, rather than the *o*-vs-*ga* bias measure used in the main analysis; since this new analysis looks at mere correct/incorrect provision of *-o* marking rather than incorrect *ga*-for-*o* substitutions. We therefore ran a final analysis with the (scaled and centred) input frequency of the relevant *-o* form (and, as a control *-ga* form) in place of the input-bias predictor. Again, this analysis yielded no evidence for an effect of input frequency, of either the *-o* form ($M = -0.22$, $SE = 0.22$, $95\% CI = [-0.65, 0.22]$, $BF = 1.09$) or the *-ga* form ($M = -0.18$, $SE = 0.23$, $95\% CI = [-0.64, 0.25]$, $BF = 0.91$).

Finally, we ran three further models with new operationalizations of the input-bias predictor developed in response to review suggestions (detailed in the ***Additional non-preregistered exploratory predictors*** subsection of the Method section). The full model outputs for these models can be found in the following files on the project OSF site, with the main effect for the input-bias predictor shown in brackets.

InputBias2.txt: New (AGENT-SUBJECT-only) *-ga* counts; original (raw number) *-o* counts ($M = 0.01$, $SE = 0.31$, $BF = 1.07$).

InputBias3.txt: Original (raw number) *-ga* counts; New (clause-initial-OBJECT-only) *-o* counts ($M = 0.02$, $SE = 0.31$, $BF = 1.05$).

InputBias4.txt: New (AGENT-SUBJECT-only) *-ga* counts; New (clause-initial-OBJECT-only) *-o* counts ($M = 0.02$, $SE = 0.31$, $BF = 1.01$).

Thus, these different operationalizations did not change the pattern of results, with inconclusive Bayes Factors very close to 1, indicating that the observed data are almost exactly equally likely under the null hypothesis of no effect and the alternative hypothesis of the expected input-bias effect.

Discussion

In the present study, 38 native Japanese speaking children aged 4;0-4;11 were asked to repeat OBJECT SUBJECT VERB (OSV) sentences which – because they reverse typical Japanese SOV order – unambiguously require *wo...ga* marking in order to convey the correct meaning. Constructivist accounts of language acquisition predict that – across nouns – the frequency with which children produce correct OBJECT-*o* marking versus incorrect OBJECT-*ga* marking (with all other responses discarded as missing data) will be related to the relative frequency of each target noun with *o-vs-ga* marking in child-directed speech (e.g., the relative input frequency of *banana-o* vs *banana-ga*). In fact, the findings provided no support for this prediction. Neither did a Bayes Factor analysis provide any positive evidence against this prediction, which could arguably have been taken as support for rival generativist accounts. Rather, the data were inconclusive; almost exactly as likely under the experimental hypothesis and the null hypothesis of no effect.

When a study fails to find a predicted or expected effect, several explanations are possible. In what follows, we consider each of the following in turn, taking particular care to do so systematically and even-handedly, rather than falling into the common trap of leaping immediately to the first, “most interesting”, possible conclusion:

1. The effect is underlyingly present in the world, but there is something about the specific sample (e.g., age, language background, SES), method (e.g., production vs comprehension, stimuli, number and/or nature of trials, sensitivity of the dependent measure) operationalization (e.g., corpus and calculations used for the input-bias predictor; definition of equi-biased verbs) or analysis which prevented us from detecting this effect.
2. The effect is underlyingly present in the world and can in principle be detected with the specific sample, method, operationalization and analysis used but was not observed due to nothing more than sampling error; the luck of the draw.
3. The effect is not underlyingly present in the world or is present but with much smaller magnitude than the relevant theory and/or previous research would have us believe; and to the extent that the effect was expected on the basis of previous research, this is because previous research overestimates the magnitude of this effect.

Considering the first possibility, determining exactly whether there is something about the particular sample, method, operationalization or analysis that prevented us from observing an underlyingly-present effect would of course require running many different versions of the study. In the meantime, it is worth reemphasizing that the present study is similar in most respects to the studies of Tatsumi et al (2017, 2018), which used a similar input-bias measure and a similar (but, in fact, smaller) corpus, and which did find the predicted effect.

The best candidate for an important methodological difference between the present study and those of Tatsumi et al. (2017, 2018) is the age of the children. Although the age-range of the present study, 4;0-4;11 sits within the age range of three of the four studies reported in Tatsumi et al. (2017, 2018), lies towards the upper end of the range, with those of the previous studies ranging as low as 3;3, 3;5, 3;2 and 2;7 (see Table 3).

Were the children in the present study “too old” to show a lexical input frequency effect? Would children aged 4;0-4;11 even be *predicted* to show this type of input frequency effect under a constructivist account? This is a difficult question to answer, since constructivist accounts come in many varieties. The exemplar account of Abbot-Smith and Tomasello (2006) and Ambridge (2020a,b) argues explicitly that the exemplars that children use to build linguistic representations are never erased or effaced but are in some sense stored forever. This account therefore clearly predicts that (Ambridge et al., 2015: 241):

frequency effects are ubiquitous in every domain of child language acquisition and ...any apparent null finding simply reflects a failure to conceptualize frequency appropriately, to find a sufficiently sensitive dependent measure, or to hold constant other relevant factors.

Other constructivist accounts seem to assume, at least implicitly, that the original exemplars that gave rise to linguistic abstractions (e.g., NOUN+*o*; NOUN+*ga*) are in some sense “replaced” by these abstractions, and thus that we would not expect to see input-frequency effects beyond the age at which this abstraction is deemed to have happened. The truth may lie somewhere in between, with very young infants relying almost exclusively on individual exemplars, and adults almost completely on emergent abstractions, but in a continuous and gradient fashion, with no “line in the sand”, no point at which “full abstraction has now happened”.

Thus, it is impossible to say, in any meaningful way, whether or not the participants of the present study were too old to show the expected input-frequency effect, having formed “fully abstract” NOUN+*o* and NOUN+*ga* schemas. Certainly, any definitive claim that this is the case risks looking like post-hoc rationalization, given that older

children, and indeed adults, show input-frequency effects for other well-entrenched schemas such as the English past-tense [*VERB*]+*ed* and transitive-causative (e.g., [SUBJECT] [*VERBed*] [*OBJECT*]) constructions (e.g., Bidgood et al, 2021; Blything et al., 2018). Neither does the particular domain of children's first language acquisition of noun case marking seem to be immune to frequency effects (e.g., Granlund et al., 2019; Saviciuite et al., 2018).

Another possibility (emphasized by an anonymous reviewer) is that the lack of an apparent input frequency effect (and children's relatively low overall performance) could be attributed to processing and performance difficulties (e.g., Omaki & Lidz, 2015; Phillips & Ehrenhofer, 2015). This is absolutely possible. Certainly, children found the task difficult (note the overall correct performance rate of around 50%), presumably due to the need to override the AGENT-first heuristic (e.g., Abbot-Smith et al., 2017; Huang & Arnold, 2016; Huang et al., 2013; MacDonald, 2013) that applies to around 95% of Japanese sentences (Kuno, 1973), and which may even have its roots in nonlinguistic cognition (though see Sano, 2020, for some evidence that Japanese may constitute something of an exception). Indeed, recall that in the crosslinguistic meta-analysis of Kolak et al. (2025), older children and even adults had difficulty processing OBJECT-first sentences, even in comprehension.

Perhaps, then, one reason why we fail to see an effect of input-frequency is due to a type of floor-effect introduced by processing difficulties. Perhaps, for example, children found the task so difficult that they gave up trying to regenerate the target sentences from scratch and relied instead on attempting to recall the relevant to-be-repeated sentence by rote. On the other hand, performance and processing limitations can in principle be used to explain away frequency effects. Suppose that children had shown the predicted input frequency effect. It would be possible to argue that 4-year-old children, in principle, have complete mastery of *-o* marking in OSV sentences, but rely on recalling individual *NOUN-o* chunks when subject to performance and processing difficulties, as is no doubt the case in the present, clearly relatively difficult task.

It is important, then, to be aware of one's own biases when discussing processing and performance difficulties. Given our overall theoretical approach – as well as the findings of our own previous studies (e.g., Tatsumi et al., 2017; 2018) – we clearly expected to find a frequency effect here. Thus, it would all too convenient for us to cite performance and processing difficulties (or, conversely, mastery of *-o* marking that is masked by these difficulties) as the reason for the present null or inconclusive finding.

This brings us to second possibility set out above: that the effect is present in the world, and in-principle detectable using the present setup, but was not observed due to nothing more than sampling error; the luck of the draw. Perhaps because it is

theoretically uninteresting, this possibility is rarely discussed by authors of studies with null or inconclusive findings. But it is a very real one. Inspect the forest plot of almost any meta-analysis that yields conclusive evidence for an effect, and you will find studies that, despite being conducted along very similar lines to the studies that did find the effect, failed to do so, for no reason other than sampling error. Indeed, this banal fact underlies the publication-biases analyses that can be found in any reputable meta-analysis. If there is no publication bias, we would expect to see just as many studies that underestimate as overestimate the true effect size; and the smaller the study, the greater the degree of under/overestimation. It would not be at all surprising then if the present study, with its relatively small sample of $N=38$, significantly underestimated an in-principle detectable effect (to the extent that its estimate is close to zero), due to nothing more than sampling error.

The third and final possibility to consider, then, is that the theory that gave rise to the present experimental prediction is incorrect; that input-frequency effects are not inherently present in child language acquisition. If any finding of an input frequency effect is to be taken as evidence in support of constructivist accounts, then any finding of a genuinely null effect of input frequency (as opposed to a merely inconclusive effect) must be taken as evidence against constructivist accounts. Furthermore, any such finding must be taken as evidence in support of accounts that seem to predict the *absence* of frequency effects, at least beyond the point at which abstractions have been formed (e.g., Clahsen et al., 1992; Fodor, 1998; Gertner et al., 2006; Hoesktra & Hyams, 1998; Pinker, 1999; Rispoli et al. 2009; Schuler et al., 2016). We are happy to put on the record here that we consider any future finding of positive evidence for the null hypothesis of no frequency effect to constitute probabilistic evidence against constructivist accounts and in support of early abstraction/generativist accounts⁷. But it bears repeating that the present study did *not* find positive evidence for either the null hypothesis of no frequency effect or the alternative hypothesis of a frequency effect; hypotheses which – according to the present Bayes Factor analysis – were almost exactly equally consistent with the data.

In other words, the present study leaves the state of the evidence essentially unchanged. So, what was the existing state of the evidence prior to the present study? This is of course a matter of debate and interpretation. But, on our reading of the pre-existing literature – which of course comprises only *published* data – input frequency effects are ubiquitous in the field. To take just one example, a target article in a special issue of the *Journal of Child Language* that marshalled evidence for this claim (Ambridge et al., 2015) received, in addition to over 500 citations, nine peer commentary articles, all of which, while challenging many of the details and interpretations, broadly accepted the main thesis.

⁷ Like most researchers, at least in practice, we do not assume a Popperian falsificationist view of science, but one based on probabilistic updating of the evidence

It is almost certain, however, that the prevalence and magnitude of these frequency effects have been overestimated. Over the past decade or so, it has becoming increasingly apparent that the scientific record has long been distorted by four biases (e.g., De Vries et al., 2018): (1) *publication bias*, whereby researchers “file drawer” studies with null or inconclusive results, or see them rejected by “selective” journals; (2) *outcome-reporting bias*, whereby researchers drop groups, conditions or sub-studies that fail to show a clear and/or desired effect; (3) *spin*, whereby researchers talk-up findings that show the predicted effect, while seeking to explain away null or inconclusive results as anomalous and (4) *citation bias*, whereby researchers fail to cite null or inconclusive findings. Consider, for example, drugs used to treat depression (De Vries et al., 2018). On the basis of the clinical trial register maintained by the US Food and Drug Administration only around 50% of studies find a positive effect. The scientific literature, however, tells a very different story, with 95% of published studies reporting (and/or spinning) such an effect.

Similar findings have been reported outside of the clinical arena. For example, in a survey of Psychology journals, Scheel et al. (2021) found that 96% of studies reported positive effects. Closely mirroring the antidepressants study of De Vries et al. (2018), the true rate of null results is likely to be around 50%: When Scheel et al. (2021) restricted the analysis to Registered Reports (a format in which studies are “accepted in principle” by journals before data collection begins), the rate of positive results dropped from 96% to 44%. We are not aware of any such analysis conducted specifically for first language acquisition research, though similar findings have been reported in the neighbouring fields of education (Chow & Ekholm, 2018) and in the context of the debate over the possible cognitive benefits of bilingualism (e.g., Gunnerud et al., 2020). Thus, there is no reason to expect that studies of child language acquisition in general, or of input frequency effects in particular, are likely to be immune to the problems of publication bias, outcome-reporting bias, spin and citation bias.

On the contrary, there is at least some reason to think that empirical investigations of input frequency may be *particularly* susceptible to publication bias. Although we know of no formal analysis to support this assertion (which would be an interesting research project in its own right) our perception of the field that most studies that test for effects of input frequency are run by researchers aligned with theories that predict such effects. It is likely, then, that the file-drawers of such researchers contain many null or inconclusive frequency effects. Indeed, in a 2025 cross-discipline survey of 11,000 researchers (Springer Nature, 2025), two thirds agreed with the statement “Null results are unlikely to be accepted for publication by journals”, and – rather shockingly – a third with the statement “I did not know it was possible to publish null results in a journal”. Notably, several studies of input-frequency effects by (at least on our reading of the literature) theoretically oriented sceptics, have found null effects

of input frequency, or at least broader effects of apparent statistical insensitivity (e.g., Hudson Kam & Newport, 2009; Gagliardi & Lidz, 2014; Gagliardi et al., 2017).

For this reason, then, we consider it important to report the present inconclusive effect, particularly given that it was observed in a context in which, on the face of it, a positive effect would have been expected, at least with a well-powered study. Although the inconclusive findings mean that the study was underpowered with regard to both the experimental and null hypotheses, in our view, this makes it particularly crucial that this study forms part of the scientific record. That is, we agree with Hernán (2022: 203) that “It is preferable to have multiple studies with imprecise estimates than having no study at all. After several studies become available, we will meta-analyze them and provide a more precise pooled effect estimate”. It is only by publishing all such null and inconclusive results that the field can build an understanding of the true magnitude of input frequency effects and – assuming that their distribution is not purely random – of the conditions under which such effects are and are not observed. While inconclusive on their own, the present findings would certainly serve to dilute the pooled input-frequency effect in any future meta-analysis. Indeed, on the assumption that this domain is no more immune to biases in the scientific record than domains that have been more closely studied, we can reasonably assume that only around half of *conducted* studies found the predicted effect, as opposed to around 95% of *published* ones.

Given that – at least given our informal knowledge of the literature – published null and inconclusive findings for input frequency are relatively rare, it is far too early to draw conclusions as to whether there are particular conditions under which we are more or less likely to see input frequency effects. At this point, we can only speculate.

Our speculation, for what it is worth, is that input-frequency effects are, in principle, largest at the earliest stages of acquisition, when children are most reliant (though probably never *exclusively* reliant) on the use of rote-learned chunks. (We say “in principle” because experiments with children are inherently noisy, and more-often-than-not significantly underpowered.) Input-frequency effects are, in principle, smallest for adults – to the extent that they can often be observed only using fine-grained measures such as eye-tracking or reaction-time data – who make the most use (though probably never *exclusive* use of in some sense emergent abstractions). Thus, the extent to which an input frequency effect will be observed in any given study (setting aside sampling error) depends on the position of the participants on that continuum – from mostly exemplars to mostly abstractions – with regard to the phenomenon under investigation (e.g., here *NOUN-o* marking); and, crucially, how this interacts with the precise design of the study and processing/performance limitations.

Testing this speculation will require not just a handful of studies, but – for each domain of interest – multiple meta-analyses. And, of course, all of these studies will need to be honestly and transparently reported, whether the results are as expected, null or inconclusive. Indeed, we will make real progress on the key theoretical questions facing the field only once we have reached the point at which null and inconclusive findings are just as likely to be published (and cited and included in meta-analyses) as are positive findings. We are probably a very long way from that point (we cannot be certain, as the whole problem is that unpublished studies are, in Donald Rumsfeld’s infamous taxonomy, “unknown unknowns”). Our hope is that by publishing the present unanticipated null finding as prominently as possible, we can play our own small part in nudging the field – if imperceptibly at first – in that direction.

References

- Abbot-Smith, K., Chang, F., Rowland, C., Ferguson, H., & Pine, J. (2017). Do two and three year old children use an incremental first-NP-as-agent bias to process active transitive and passive sentences? A permutation analysis. *PLOS ONE*, 12(10), e0186129. <https://doi.org/10.1371/journal.pone.0186129>
- Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23, 275–290. <https://doi.org/10.1515/TLR.2006.011>
- Akhtar, N. (1999). Acquiring basic word order: evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26, 339–356. <https://doi.org/10.1017/S030500099900375X>
- Ambridge, B. (2020). Against stored abstractions: A radical exemplar model of language acquisition. *First Language*, 40(5-6), 509–559. <https://doi.org/10.1177/0142723719869731>.
- Ambridge, B. (2020). Abstractions made of exemplars or You’re all right and I’ve changed my mind. Response to commentators. *First Language*, 40(5-6), 640–659 <https://doi.org/10.1177/0142723720949723>.
- Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511975073>
- Ambridge, B., & Lieven, E. V. M. (2015). A constructivist account of child language acquisition. In B. MacWhinney and W. O’Grady (Eds.), *Handbook of Language*

- Emergence* (pp. 478-510). Wiley Blackwell. <https://doi.org/10.1016/B978-0-08-101107-2.00023-3>
- Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 149-168. <https://doi.org/10.1002/wcs.1215>
- Ambridge, B., Rowland, C. F., Theakston, A. L. & Kidd, E. J. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(2), 239-7. <https://doi.org/10.1017/S030500091400049X>
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19, 241-248. <https://doi.org/10.1111/j.1467-9280.2008.02075.x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, E., & MacWhinney, B. (1987). Competition, variation and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Erlbaum.
- Bidgood, A., Pine, J., Rowland, C., Sala, G., Freudenthal, D., & Ambridge, B. (2021). Verb argument structure overgeneralisations for the English intransitive and transitive constructions: grammaticality judgments and production priming. *Language and Cognition*, 13(3), 397-437. <https://doi.org/10.1017/langcog.2021.8>
- Blything, R. P., Ambridge, B., & Lieven, E. V. (2018). Children's acquisition of the English past-tense: Evidence for a single-route account from novel verb production data. *Cognitive Science*, 42, 621-639. <https://doi.org/10.1111/cogs.12581>
- Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1-28. <https://doi.org/10.18637/jss.v080.i01>
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing. <https://doi.org/10.1075/tsl.9>
- Chandler, S. (2010). The English past tense: Analogy redux. *Cognitive linguistics*, 21(3), 371-417. <https://doi.org/10.1515/COGL.2010.014>

- Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, 37(6), 739-748. <https://doi.org/10.1037/0012-1649.37.6.739>
- Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, 30(3), 727-744. <https://doi.org/10.1007/s10648-018-9437-7>
- Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, 45(3), 225-255. [https://doi.org/10.1016/0010-0277\(92\)90018-D](https://doi.org/10.1016/0010-0277(92)90018-D)
- Comrie, B. (2013). Alignment of case marking of full noun phrases (Chapter 98). In Haspelmath, M., Dryer, M. S., Gil, D., & Comrie, B. (Eds.), *World Atlas of Linguistic Structure online*. Max Planck Digital Library. <https://wals.info/>
- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Pearson.
- De Vries, Y. A., Roest, A. M., de Jonge, P., Cuijpers, P., Munafò, M. R., & Bastiaansen, J. A. (2018). The cumulative effect of reporting and citation biases on the apparent efficacy of treatments: the case of depression. *Psychological Medicine*, 48(15), 2453-2455. [10.1017/S0033291718001873](https://doi.org/10.1017/S0033291718001873)
- Fodor, J. D. (1998). Unambiguous triggers. *Linguistic Enquiry*, 29(1), 1-36. <https://doi.org/10.1162/002438998553644>
- Gagliardi, A., Feldman, N. H., & Lidz, J. (2017). Modeling statistical insensitivity: Sources of suboptimal behavior. *Cognitive Science*, 41(1), 188-217. <https://doi.org/10.1111/cogs.12373>
- Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, 90(1), 58-89. <https://doi.org/10.1353/lan.2014.0010>
- Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, 17(8), 684-691. <https://doi.org/10.1111/j.1467-9280.2006.01767.x>
- Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Lieven, E.V.M., Pine, J.M., Theakston, A.L. & Ambridge, B. (2019). Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited production study of Polish, Finnish and

- Estonian. *Journal of Memory and Language*, 107, 169-194.
<https://doi.org/10.1016/j.jml.2019.04.004>
- Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493-498. <https://doi.org/10.1111/2041-210X.12504>
- Gunnerud, H. L., ten Braak, D., Reikerås, E. K. L., Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin*, 146(12), 1059–1083.
<https://doi.org/10.1037/bul0000301>
- Hakuta, K. (1982). Interaction between particles and word order in the comprehension and production of simple sentences in Japanese children. *Developmental Psychology*, 18(1), 62-76. <https://doi.org/10.1037/0012-1649.18.1.62>
- Hernán, M. A. (2022). Causal analyses of existing databases: no power calculations required. *Journal of Clinical Epidemiology*, 144, 203-205.
<https://doi.org/10.1016/j.jclinepi.2021.08.028>
- Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua*, 106(1-4), 81-112. [https://doi.org/10.1016/S0024-3841\(98\)00030-8](https://doi.org/10.1016/S0024-3841(98)00030-8)
- Huang, Y. T., & Arnold, A. R. (2016). Word learning in linguistic context: Processing and memory effects. *Cognition*, 156, 71–87.
<https://doi.org/10.1016/j.cognition.2016.07.012>
- Huang, Y. T., Zheng, X., Meng, X., & Snedeker, J. (2013). Children's assignment of grammatical roles in the online processing of Mandarin passive sentences. *Journal of Memory and Language*, 69(4), 589–606.
<https://doi.org/10.1016/j.jml.2013.08.002>
- Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, 59(1), 30–66.
<https://doi.org/10.1016/j.cogpsych.2009.01.001>
- Jeon, M., & De Boeck, P. (2017). Decision qualities of Bayes factor and *p* value-based hypothesis testing. *Psychological Methods*, 22(2), 340-360.
<https://doi.org/10.1037/met0000140>
- Kass, R. E. & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90 (430), 791. <https://doi.org/10.1080/01621459.1995.10476572>

- Kolak, J., Vihman, V., Engelmann, F., Granlund, S., Theakston, A., Lieven, E. V. M., Pine, J. M., Fazekas, J., & Ambridge, B. (2025). Why learners privilege word-order over case-marking: A cross-linguistic meta-analysis, new data from Estonian, Finnish and Polish, and a discriminative learning model. *Psychological Review*. <https://doi.org/10.1037/rev0000560>
- Kuno, S. (1973). *The structure of Japanese*. Cambridge: MIT Press.
- Lust, B., Flynn, S., and Foley, C. (1996). What children know about what they say: Elicited-imitation as a research method for assessing children's syntax. In D. McDaniel, C. McKee and H. S. Cairns, eds., *Methods for assessing children's syntax*. MIT Press.
- MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226. <https://doi.org/10.3389/fpsyg.2013.00226>
- MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press. <https://doi.org/10.4324/9781315805641>
- Matsuo, A., Kita, S., Shinya, Y., Wood, G. C., & Naigles, L. (2012). Japanese two-year-olds use morphosyntax to learn novel verb meanings. *Journal of Child Language*, 39(3), 637–663. <https://doi.org/10.1017/S0305000911000213>
- Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2007). French children's use and correction of weird word orders: A constructivist account. *Journal of Child Language*, 34(2), 381–409. <https://doi.org/10.1017/S030500090600794X>
- Minashima, H. (2001). On the deletion of accusative case markers in Japanese. *Studia Linguistica*, 55(2), 176–191. <https://doi.org/10.1111/1467-9582.00078>
- Miyagawa, S. (2003). A-movement scrambling and options without optionality. In S. Karimini (Ed.), *Word order and scrambling*. Blackwell. <https://doi.org/10.1002/9780470758403.ch8>
- Murao, A., Ito, T., Fukuda, S. E., & Fukuda, S. (2017). Grammatical case-marking in Japanese children with SLI. *Clinical Linguistics & Phonetics*, 31(7–9), 711–723. <https://doi.org/10.1080/02699206.2017.1310929>
- Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17(2), 357–374. <https://doi.org/10.1017/S0305000900013817>

- Noble, C. H, Rowland, C. F, Pine, J. M. (2011). Comprehension of argument structure and semantic roles: Evidence from infants and the forced-choice pointing paradigm. *Cognitive Science*, 35(5), 963-982.
<https://doi.org/10.1111/j.1551-6709.2011.01175.x>
- Omaki, A., & Lidz, J. (2015). Linking parser development to acquisition of syntactic knowledge. *Language Acquisition*, 22(2), 158-192.
<https://doi.org/10.1080/10489223.2014.943903>.
- Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic Approaches to Bilingualism*, 5(4), 409-453.
<https://doi.org/10.1075/lab.5.4.01phi>
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. MIT Press. <https://doi.org/10.7551/mitpress/4158.001.0001>
- Pinker, S. (1999). *Word and Rules*. Basic Books.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38(3), 265-2.
<https://doi.org/10.1006/jmla.1997.2546>
- Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, 10(2), 178-192.
<https://doi.org/10.1037/1082-989X.10.2.178>
- Rispoli, M., Hadley, P. A., & Holt, J. K. (2009). The growth of tense productivity. *Journal of Speech, Language, and Hearing Research*, 52(4), 930-944.
<https://doi.org/10.1044/1092-4388>
- Sano, T. (2020). On the generality of the agent-first strategy. In M. M. Brown & A. Kohut (Eds.) *Proceedings of the 44th Boston University Conference on Language Development* (pp. 503-507). Cascadia Press.
<https://www.lingref.com/buclid/44/BUCLD44-40.pdf>
- Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: lexical and structural priming of syntactic constructions in young children. *Developmental Science*, 6(5), 557-567. <https://doi.org/10.1111/1467-7687.00312>

- Saviciute, E., Ambridge, B., & Pine, J. M. (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language*, 45(3), 641-672.
<https://doi.org/10.1017/S030500091700037X>
- Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, 4(2), 25152459211007467. <https://doi.org/10.1177/25152459211007467>
- Schönbrodt, F. D., & Wagenmakers, E. J. (2017). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 1-15.
<https://doi.org/10.3758/s13423-017-1230-y>
- Schuler, K. D., Yang, C., & Newport, E. L. (2016). Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 38.
<https://escholarship.org/uc/item/1cn5n52p>
- Shigenaga, Y. (2014). *Processing and acquisition of scrambled sentences by learners of Japanese as a second language*. Doctoral dissertation, University of Arizona.
- Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12(3), 229-265.
<https://doi.org/10.1016/0010-0277>
- Springer Nature (2025). *The state of null results: Insights from 11,000 researchers on negative or inconclusive results*. White paper.
<https://stories.springernature.com/the-state-of-null-results-white-paper/index.html>
- Stefanowitsch, A., & Gries, S. T. (2003). Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209-243. <https://doi.org/10.1075/ijcl.8.2.03ste>
- Takano, Y. (2003). Nominative objects in Japanese complex predicate constructions: A prolepsis analysis. *Natural Language & Linguistic Theory*, 21(4), 779-834. <https://doi.org/10.1023/A:1025545313178>
- Tatsumi, T., Ambridge, B., & Pine, J. M. (2017). Disentangling Effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science*.
<https://doi.org/10.1111/cogs.12554>

- Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Testing an input-based account of children's errors with inflectional morphology: An elicited production study of Japanese. *Journal of Child Language*. <https://doi.org/10.1017/S0305000918000107>
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Tomasello, M., & Brooks, P. J. (1998). Young children's earliest transitive and intransitive constructions. *Cognitive Linguistics*, 9(4), 379-395. <https://doi.org/10.1515/cogl.1998.9.4.379>
- Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3), 158-189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, 106(1-4), 23-79. <https://doi.org/10.1016/S0024-3841>
- Yano, M., & Koizumi, M. (2018). Processing of non-canonical word orders in (in)felicitous contexts: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 33(10), 1340-354. <https://doi.org/10.1080/23273798.2018.1489066>

Data, Code and Materials Availability Statement

All stimuli, data and analysis code can be downloaded from the Project OSF Site <https://osf.io/j7azp/>. Note that all files with a date-stamp earlier than January 2025 relate to the original Registered Report version (including the video **JapaneseStudy.mov**, which shows a complete run through the original version of the experiment). All data and analysis code files relating to the current version can be found in the folder **Zoom_Version_2025 > LDR_SUB2.zip**.

Ethics Statement

The study was approved by the ethics committee of the University of Liverpool. Since Japanese universities generally do not have ethics committees for non-medical research, an independent local ethics review was generously conducted by a Professor of Ethics at a Japanese University. All participants' caregivers gave informed written consent before their children took part in the study.

Authorship and Contributorship Statement

Ben Ambridge conceived of the study, designed the study, analyzed the data, and wrote the first draft of the manuscript. **Motoki Saito, Samuel David Jones, Tomoko Tatsumi** and **Kumiko Fukumura** contributed to the design of the study (in particular, obtaining the corpus counts and selecting the stimuli) and revised the manuscript. **Tomoko Tatsumi** and **Ayuno Kawakami** collected the data and revised the manuscript. **Colin Bannard** assisted with data analysis and revised the manuscript. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 681296: CLASS). Ben Ambridge is Professor in the International Centre for Language and Communicative Development (LuCiD) at The University of Manchester. The support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged.

License

Language Development Research (ISSN 2771-7976) is published by TalkBank and the Carnegie Mellon University Library Publishing Service. Copyright © 2026 The Author(s). This work is distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits any use, reproduction and distribution of the work for noncommercial purposes without further permission provided the original work is attributed as specified under the terms available via the above link to the Creative Commons website.