# Comparing generativist and constructivist accounts of the use of the past tense form in early child Japanese*

Tomoko Tatsumi

*University of Liverpool*

*Tokyo University of Foreign Studies*

Julian M. Pine

*University of Liverpool*

*ESRC International Centre for Language and Communicative Development (LuCiD)*

*Address for correspondence: Tomoko Tatsumi, Institute of Psychology, Health and Society, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, England. EMAIL: Tomoko.Tatsumi@liverpool.ac.uk

RUNNING HEAD: RI-analogue and input-driven accounts

Keywords: Language Acquisition; Morphological development; Japanese

# Comparing generativist and constructivist accounts of the use of the past tense form in early child Japanese


## Abstract

The present study investigated children's early use of verb inflection in Japanese by comparing a generativist account, which predicts that the past tense will have a special default-like status for the child during the early stages, with a constructivist input-driven account, which assumes that children's acquisition and use of inflectional forms reflects verb-specific distributional patterns in their input. Analysis of naturalistic data from 4 Japanese children aged 1;5 to 2;10 showed that there was substantial by-verb variation in the use of inflectional forms from the earliest stages of verb use, and no general preference for past tense forms. Correlational and partial correlational analyses showed that it was possible to predict the proportional frequency with which the child produced verbs in past tense versus other inflectional forms on the basis of differences in the proportional frequency with which the verb occurred in past tense form in the child's input, even after controlling for differences in the rate at which verbs occurred in past tense form in input averaged across the caregivers of the other children in the sample. When taken together, these results count against the idea that the past tense has a special default-like status in early child Japanese, and in favour of a constructivist input-driven account of children's early use of verb inflection.

## Introduction

How children learn the system of verb inflection in their language is a long-standing question in language acquisition research (e.g. Brown, 1973; Hoekstra & Hyams, 1998; MacWhinney, 1978; Pinker, 1984; Pizzuto & Caselli, 1992; Shirai & Andersen, 1995; Tomasello, 2000; Wexler, 1994; 1998). However, work from different theoretical perspectives has tended to answer this question in very different ways. Generativist research has tended to emphasise the low frequency of inflectional errors in children's speech, and the fact that those errors that do occur tend to involve the incorrect use or over-use of one particular default-like form. Constructivist research has tended to emphasise the limited productivity of children's early use of verb inflection and the fact that children's early use of particular inflections tends to mirror the frequency statistics of the input.

The aim of the present study is to use data from early child Japanese – an agglutinative language, which shows substantial by-verb variation in the frequency with which verbs occur in different forms – to compare these two kinds of account. In order to test the first kind of account, we look for evidence that the past tense form, which has been reported to occur very early in the speech of Japanese children, has some special default-like status for the Japanese-learning child. In order to test the second kind of account, we look for evidence that by-verb variation in the children's tendency to use forms with the past tense inflection is related to by-verb variation in the relative frequency with which those verbs occur with the past tense inflection in the input. Since many previous studies have not distinguished adequately between specific effects of the input to which the child is exposed and more general effects of the semantic-distributional properties of the language being learned, we look for specific relations between the child's language use and the child's own input by partialling out the relation between each child's language use and input averaged across the caregivers of the other children in the sample.

*Generativist accounts*

Many generativist accounts of the development of verb inflection argue for early knowledge of inflection on the basis of the low frequency of inflectional errors in children's early speech (e.g. Hoekstra & Hyams, 1998; Wexler, 1998). According to this view, young children's largely correct use of verb morphology reflects underlying knowledge of inflection, and those errors that do occur reflect a tendency to produce Optional Infinitive (OI) or Root Infinitive (RI) errors (i.e. non-finite forms in contexts in which a finite form is required in the adult language; see also Rizzi, 1993/1994).

Some theories of the RI stage attempt to explain why RI errors occur in some languages and not in others. For example, according to Wexler (1998), RI errors reflect a Unique Checking Constraint (UCC), which interacts with the type of language being learned to result in RI errors in obligatory subject languages such as English, Dutch French and Swedish, but not in INFL-licensed null subject languages such as Spanish and Italian. However, others have attempted to extend the idea that children make errors involving the incorrect use of a 'tenseless' verb form to languages in which RI errors do not occur. For example, Salustri and Hyams (2003) argue that, in Italian, the imperative is an RI analogue, and Grinstead, De la Mora, Vega-Mendoza and Flores (2009) argue that, in Spanish, the 3sg present tense form is an RI analogue. The concept of an RI analogue has also been extended to non-Indo-European languages. For example, Kim and Phillips (1998) argue that, in early child Korean, children's overuse of a particular mood marker is analogous to the use of RIs, and Murasugi, Nakatani and Fuji (2009) argue that, in early child Japanese, the past tense form can be viewed as an RI analogue (see also Murasugi, 2015). The key idea in all of these analyses is that there is an early stage in development, during which TENSE can be underspecified, or fail to merge with V, in the underlying representation of the sentence, and that, during this stage, children

learning non-RI languages use a particular form of the verb as a 'tenseless' form in the same way that children learning RI languages use the infinitive.

The RI-analogue approach has the potential to provide a unified account of data across a wide range of different languages. However, it is subject to two potential problems. The first is that it assumes that early child language exhibits 'inflectional imperialism' (Slobin, 1985; 1216): the tendency to make errors that involve 'defaulting' to a single dominant pattern in the language. However, as Dressler (2005) points out, there is wide typological variation in the extent to which languages exhibit this property. Thus, although a common feature of weakly inflecting fusional languages such as English, it is much less common in agglutinating languages such as Turkish and Hungarian. This observation raises doubts about how far it is possible to extend the RI-analogue approach – and, in particular, about the extent to which the approach can be extended to agglutinative languages.

The second is that it is often difficult to distinguish empirically between an RI-analogue account of children's early use of verb morphology and a constructivist account that emphasises the role of input frequency (e.g. Tomasello, 2000). This is because the hypothetical RI analogue is often the most frequent form in the input. For example, Grinstead et al. (2009) argue that, in Spanish, the 3sg present tense form is an RI analogue. However, in Spanish, the 3sg present tense form is the highest frequency form in the input (Aguado-Orea & Pine, 2015), and is also homophonous with the imperative for regular verbs. This makes it impossible to tell whether the overuse of the 3sg present tense form by Spanish children reflects its linguistic status as an RI analogue or a tendency to default to the most common form of the verb when the child is unable to generate the correct form in a particular morphological context.

In view of these considerations, the first aim of the present study is to evaluate the plausibility of the RI-analogue approach by focusing on the early use of verb inflection in

Japanese — an agglutinative language which shows substantial by-verb variation in the frequency with which verbs occur in different forms. This will be done by testing Murasugi et al.'s (2009) claim that the past tense form is an RI analogue for the Japanese-learning child. More specifically, we will investigate: 1) whether there is any evidence of a special preference for past tense forms in Japanese children's early verb vocabularies; and 2) whether the errors produced by Japanese children tend to involve the over-use of past tense forms.

**Constructivist accounts**

Constructivist accounts of the development of verb inflection have tended to emphasise the limited productivity of children's early use of verb morphology and the fact that children's early use of particular inflections tends to mirror the frequency statistics of the input (e.g. Pizzuto & Caselli, 1992; Tomasello, 2000). According to this view, young children's correct use of verb morphology reflects knowledge that is initially embedded within particular lexically-restricted patterns, and those errors that do occur reflect a tendency to use the form of the verb that occurs most frequently in the input language.

There is already considerable evidence that children's early verb learning and use is related to the frequency with which particular verbs occur in the input. For example, Smiley and Huttenlocher (1995), Naigles and Hoff-Ginsberg (1998) and Theakston, Lieven, Pine and Rowland (2004) all report significant relationships between the frequency with which particular verbs occur in the input language and the order in which they emerge in the child's speech (see Ellis (2002) and Ambridge, Rowland, Theakston & Kidd (2015) for reviews of the data on frequency effects in language learning). Evidence has also recently emerged that it is possible to explain the relative frequency with which young children produce particular forms of the verb in terms of the relative frequency of particular forms in the input language. For example, Freudenthal, Pine and Gobet (2010) have shown that it is possible to explain by-verb differences in the rate at which children make RI errors in a range of languages in

terms of the relative frequency of infinitive versus finite forms of the verb in English, Dutch, French, German and Spanish, and Räsänen, Ambridge and Pine (2014) have shown that it is possible to explain by-verb differences in the rate at which children produce RI errors in English in terms of the relative frequency of bare stem versus 3sg forms in the input.

However, a particular challenge facing constructivist analyses is how to disentangle effects of input frequency on children's learning and effects of sampling and/or of the semantic-distributional properties of the language being learned. With respect to the first of these issues, Tomasello and Stahl (2004) point out that analyses based on naturalistic speech samples tend to confound order of acquisition with frequency in the language, such that forms that occur with high frequency in the language are likely to be sampled earlier than forms that occur with lower frequency in the language, even if both forms were acquired by the child at the same time. This problem makes it difficult to distinguish between true effects of input frequency on learning and spurious effects that actually reflect differences in the likelihood that particular items will be sampled in naturalistic data.

A similar point can be made about measures of the relative frequency with which verbs occur in particular forms in the children's speech. Thus, it may be tempting to take correlations between the relative frequency with which particular verbs are used in particular forms by the child and by the caregiver as evidence of input-driven learning. However, as Yang (2013) points out, some words are more likely to be used in some contexts rather than others because of the semantic distributional properties of the language. For example, in English, one is more likely to use the word *bath* with the indefinite article *a* than the definite article *the* because baths are things that one *has* or *takes*, and more likely to use the word *bathroom* with the definite article than the indefinite article, because *bathrooms* are locations which one tends to *go to* or *visit*. This kind of patterning makes it difficult to distinguish between real effects of input frequency on learning and spurious effects that simply reflect

the fact that all speakers of the language are more likely to use some words in one way rather than another.

In view of these considerations, the second aim of the study is to test an input-driven account of Japanese children's use of past tense forms while controlling for the effects of sampling and/or the semantic-distributional properties of the language being learned. This will be done by looking for dyad-specific effects of the relative frequency of past tense and other inflectional forms in the input, while controlling for the effects of relative frequency in data averaged across the other caregivers in the sample. The rationale for this approach is that it has the potential to identify effects of the child's input over and above effects of input from the average mother — effects which therefore cannot be explained in terms of general differences in the likelihood of particular verbs occurring in the past tense as opposed to other inflectional forms in the input language. Note that Japanese is particularly well suited to this kind of analysis because both the past and nonpast verb forms occur frequently in both Japanese child-directed speech and in Japanese children's early production data (c.f. Clancy 1985; Shirai, 1998). This makes it possible, at least in principle, to distinguish between by-verb variation that reflects the general properties of child-directed Japanese and by-verb variation that reflects the specific properties of the speech to which particular children are exposed. We now provide a brief sketch of Japanese syntax and morphology.

*A Brief Sketch of Japanese Syntax and Morphology*

Japanese is a typical head-final language, with features such as verb-final word order, attributive-noun word order, and the use of postpositional case markers. The basic word order is SOV, but the order of nominal arguments is pragmatically conditioned and relatively free in spoken discourse — a phenomenon often referred to as 'scrambling' (see Shibatani, 1990). Nominal arguments are also frequently left unexpressed. This can make it difficult for the researcher to infer speakers' communicative intentions from spoken discourse, especially in

the case of young language-learning children. It also means that it is more difficult to identify obligatory contexts for particular verb inflections in Japanese than it is in English.

Japanese verbs have agglutinative morphology with extensive use of suffixation to mark inflectional distinctions. Finite verbs are always inflected for tense. However, unlike many Indo-European languages, Japanese does not have subject-verb agreement. A basic form consists of a stem and a tense marker as in *tabe-ru* 'eat-NONPAST' and *tabe-ta* 'eat-PAST'. However, verb forms can also be more complex with morphological processes such as derivation, inflection and concatenation. Note that past versus nonpast is the only tense distinction in Japanese, with past tense forms being used to refer to past events and nonpast forms being used to refer to both present and future events. Bare verb roots or stems are not well formed as words in Japanese, and are never produced as such. Nor does Japanese have infinitive forms like those in European languages such as German, French and Spanish. Table 1 shows some basic inflectional distinctions and the suffixes that encode them for two major morphological classes: verbs where the stem ends in a consonant and verbs where the stem ends in a vowel. The former show stem alternation, which is dependent on whether or not they are followed by an auxiliary. Note that Table 1 does not provide a comprehensive

Table 1 about here

description of Japanese verb inflection. It simply provides the reader with some information about the most common forms that Japanese speakers produce. Children are therefore typically exposed to a wider range of forms than those in Table 1 (see Shibatani (1990) for a more detailed description of Japanese verb morphology).

Previous studies of verb inflection in early child Japanese have reported that Japanese-speaking children typically use a number of different inflections by the age of two years, although some of these inflections may not be handled productively (e.g. Clancy, 1985;

Rispoli, 1981). Clancy lists the inflectional forms that children tend to acquire early. These are: V-*te* imperative, V-*ta* past, V-*teru* nonpast progressive, V-*ru* nonpast, V-*chatta* completed past, V-*nai* negative nonpast, V-*tai* desiderative nonpast (Clancy, 1985: 426, with glosses modified by the current authors). Shirai (1998) and Shirai and Miyata (2006) report frequent use of both the past and the nonpast inflection during the early stages. However, they also report individual variation in acquisition profiles, which is consistent with the idea that it may be possible to find dyad-specific input effects on Japanese children's early use of verb morphology. Murasugi et al. (2009), on the other hand, argue that there is an initial stage in early child Japanese in which children only use past tense verb forms in their speech.

*The present study*

The aim of the present study is to investigate children's early use of verb inflection in Japanese by comparing the idea that the past tense form has a special default-like status for the Japanese-learning child with the idea that Japanese children's use of past tense forms reflects the relative frequency with which particular verbs occur in past tense form in their input. The RI-analogue account will be tested by looking for evidence that Japanese-speaking children show a particular preference for past tense forms during the early stages, and a specific tendency to overuse past tense as opposed to other inflectional forms of the verb in their early speech. The input-driven account will be tested by looking for by-verb effects of the relative frequency of past tense and other inflectional forms in the child's input on the relative frequency with which they use past tense forms in their speech, while controlling for general differences in the relative frequency with which verbs occur in past tense form in Japanese child-directed speech.

## Method

### *Corpora*

The data used in the present study were those of 4 Japanese-speaking children and their caregivers: three children (Aki, Ryo and Tai) from the Miyata corpus (Miyata, 1992; 1995; 2000) and one child (Jun) from the Ishii corpus (Ishii, 1999). Both of these corpora are available in the CHILDES database (MacWhinney, 2000). The Miyata corpus consists of transcripts of parent–child interaction (predominantly mother-child interaction) recorded in the normal home environment. The children were recorded at weekly intervals, except for Aki's recordings between 1;5 and 1;11, which were conducted at monthly intervals. Each recording session lasted between 30 and 60 minutes. The Ishii corpus consists of transcripts of parent-child interaction recorded twice a month between 0;6 and 3;8, with each recording lasting approximately 15 minutes for the first half of the corpus and approximately 60 minutes for the second half.

The data for the present study were taken from a 12-month period starting from the point at which the first verb was observed in each child's corpus. Aki's first verb appeared at 1;8, Ryo's first verb appeared at 1;10 and Jun's first verb appeared at 1;8. Tai was already using verbs in his first recording at 1:5. Unintelligible utterances, imitations and repetitions were excluded from the analysis. Unintelligible utterances were defined as utterances in which any portion of the utterance was marked as unintelligible (i.e. 'xxx') in the transcript. Imitations were defined as utterances that consisted entirely of material that had occurred in one of the two immediately preceding adult utterances. Repetitions were defined as utterances that consisted entirely of material that had occurred in one of the two immediately preceding child utterances. This process was designed to minimise direct effects of the child's verb use on the parent's very use. However, it should be noted that it does not allow us to rule out all possible alternative explanations of potential input effects. For example, it does not rule out the

possibility that they may reflect more subtle effects of the child on the caregiver, such as lexical priming effects, indirect effects of parental sensitivity to the child's interests, or commonalities in the pragmatic contexts from which the child and cartegiver speech samples were drawn. After these exclusions, the total number of child utterances in each dataset was 14,417 for Aki, 10,002 for Ryo, 14,575 for Tai and 17,611 for Jun and the total number of utterances that included a verb was 2,510 for Aki, 2,772 for Ryo, 4,760 for Tai and 3,702 for Jun. The input data used in the analysis were all drawn from the speech of the children's caregivers. For Aki, Ryo and Tai, this was the mother. However, in Jun's case, input data from both the mother and the father were included as the father was the child's main interlocutor in many of the transcripts.

*Analyses*

The corpus data described above were analysed in the following ways. All verb forms, with the exception of subsidiary verbs (e.g., *miru* in *itte miru* go-CONNECTIVE see-NONPAST '(try to) go')) and irregular verbs (e.g., *nai* 'not be' and *kudasai* 'please'), were extracted from the corpora. In the case of the Miyata corpus, this was done by searching for verb forms on the morphological coding tier of the transcripts. In the case of the Ishii corpus, which is not morphologically coded, this was done by directly coding the transcripts themselves. These data were then used to identify the first 10 verb forms in Aki, Ryo and Jun's speech and to calculate the proportion of these that were past tense forms. Tai's data were not analysed in this way because he was already using verbs in his first recording sessions. They were also used to identify the first 50 verb roots in each of the children's speech and to calculate the proportion of both child and caregiver uses of each of these verb roots that was marked for past tense. The children's verb uses were also coded for morphological errors. This was done by hand-coding each of the child's verb uses in context, and identifying cases in which the child used a particular form of the verb in an inappropriate morphological

context (e.g. a past tense form in a nonpast tense context or vice versa). These measures were then used to test the RI-analogue account and the input-driven account as follows.

The RI-analogue account was tested by:

1) Identifying the first 10 verb forms that appeared in Aki, Ryo and Jun's data and calculating the proportion that were past tense forms;
2) Identifying any inflectional errors made by the children, and calculating the proportion that involved the inappropriate use of a past tense form and the proportion that involved the inappropriate use of other inflectional forms.

These analyses are designed to establish whether the past tense form had a special default-like status for the Japanese-learning child.

The input-driven account was tested by:

1) Calculating the rate at which each of the child's first 50 verbs occurred in past tense as opposed to some other form in the child's speech;
2) Calculating the rate at which each of these verbs occurred in past tense as opposed to some other form in the caregiver's speech;
3) Computing simple correlations between the by-verb rate of past tense forms in the child and caregiver's speech and partial correlations that controlled for the average rate at which past tense forms occurred in the speech of the other 3 caregivers. In each case, the relevant proportional measures were arcsine transformed to ensure that they met parametric testing assumptions.

This approach was designed to allow us to identify dyad-specific effects of the proportion of past tense forms in the child's input, controlling for effects of the general frequency statistics of Japanese child-directed speech. It should be noted that, since Japanese has agglutinative verb morphology, and it is unclear whether children distinguish between past tense inflections in simple past (i.e. verb root + past tense) and complex past tense forms (i.e. verb

root + other suffix(es) + past tense), no distinction was made between simple and complex past tense forms in any of the above analyses.

## Results

### *Testing the RI-analogue account*

The key prediction of the RI-analogue account is that the past tense form will have a special default-like status for the Japanese-learning child. One way of operationalizing this prediction is to focus on the earliest inflectional forms that children produce. For example, the RI-analogue account would seem to predict a preponderance of past tense forms in the children's early verb vocabularies. A second is to look at the kind of inflectional errors that children produce. For example, the RI-analogue account would seem to predict that the majority of inflectional errors made by the child will involve the over-use of past tense forms.

### *Earliest inflectional forms*

Table 2 provides details of the first 10 verb forms to appear in Aki, Ryo and Jun's corpora. It can be seen from Table 2 that, although, for all these children there are some past tense forms

Table 2 about here

amongst these first 10 verb forms, there is no real evidence of a particular preference for past tense forms in the children's data, with all of the children producing a range of different inflectional forms, including past tense forms, nonpast tense forms, and imperatives, and with past tense forms only accounting for between 4/10 and 5/10 forms in each case.

These results are consistent with those reported by Shirai (1998) in an analysis that included one of the same 3 children (Aki). What they seem to reflect is not a verb-general preference for the past tense form, but a tendency to acquire high frequency forms that are pragmatically useful from the child's point of view. For example, the past tense form of the verb *ar-* 'be' appears very early in all of the children's data. This form means 'be-

PAST=was', and can occur in utterances such as *hon-wa tsukue-no ue-ni atta* (book-TOPIC desk-GENITIVE top-LOCATIVE be-PAST) 'The book was on the table.' However, it can also be produced as a single-word utterance: *Atta!* (be-PAST) 'was' in object-finding contexts, where it is the equivalent of the English utterance 'There it is!' The early appearance of this form in all 3 children's speech therefore probably reflects its usefulness for drawing the caregiver's attention to interesting objects.

A similar point can be made about the imperative forms that appear in the children's data. The verbs that appear in imperative form in the children's data tend to be action verbs, such as *tot-te* (take-IMPERATIVE), which can be translated as 'Pass me (that),' and *doi-te* (step.aside-IMPERATIVE), which can be translated as 'Get out of the way'. These forms are common in the input, and their early appearance in the children's speech probably reflects the fact that they are useful for manipulating other people's behaviour.

In short, the data presented in Table 2 are consistent with the idea that what determines the nature of the children's earliest inflectional forms is not a general preference for the past tense form, but a tendency to acquire the inflectional form of the verb that is most pragmatically salient in the input. These data thus provide no real support for the RI-analogue account.

*Inflectional errors*

Table 3 provides details of the rates of inflectional errors in each of the 4 children's corpora. It can be seen from Table 3 that inflectional errors are extremely rare in all of the children's

Table 3 about here

data. These very low error rates are at least partly a reflection of the fact that it is more difficult to identify obligatory contexts for particular verb inflections in Japanese than it is in English. They also collapse across relatively long periods of development and may therefore

hide higher error rates during the early stages. Nevertheless, it is clear from Table 3, that the few inflectional errors that it is possible to identify in the children's speech include at least as many instances of the overuse of nonpast tense forms as they do of past tense forms.

These results are consistent with those presented in a similar analysis by Kato, Sato, Chikuda, Miyoshi, Sakai and Koizumi (2003), which included two of the same children (Ryo and Tai), and provide no real support for the claim that the past tense form has a special status for the Japanese-learning child. On the contrary, what the inflectional errors produced by the children seem to reflect is the use of a high frequency form of the verb (whether it be a past tense form, a nonpast tense form, or some other form) in a context in which a lower frequency form is required. Thus, on the one hand, Aki produced the past tense form *not-ta* (get.on-PAST) '(he) got on' when he was told by his mother that his brother was on the train. The most natural form to use in this context is *not-te-ru* (get.on-ASPECTUAL-NONPAST) 'he is (riding) on the train', but the past tense form *not-ta* is more frequent in Aki's mother's input (accounting for 92 of a total of 295 tokens) than the target form *not-te-ru* (41 out of 295 tokens). On the other hand, Aki also produced the opposite kind of error. For example, he produced the nonpast tense form (*iw-u* say-NONPAST) in the utterance *nani Suuze iwu?* (what Suuze (person name) say-NONPAST) 'What does Suuze say?' instead of using the past tense form *it-ta* (say-PAST) when he asked his mother what Suuze had said a moment before. In this case, the nonpast tense form *iw-u* is the most frequent form of the verb in Aki's mother's data (accounting for 41 of a total of 81 tokens), whereas the target form *it-ta* is much less frequent (9 out of 81 tokens). These examples are consistent with the idea that what determines the nature of the children's early inflectional errors is not a general preference for the past tense form, but a tendency to overuse the inflectional form of the verb that occurs most frequently in the input.

To summarise, neither the data on the children's earliest inflected forms, nor the data on their inflectional errors or the proportion of past tense forms in their early speech provide any real support for the RI-analogue account. What these data do seem to suggest is that the children's early verb use is influenced by the relative frequency with which different forms of the verb occur in the children's input. It is therefore to an analysis of the relation between children's early verb use and the input that we now turn.

*Testing the input-driven account*

Table 4 presents data on the proportional frequency with which each of the children and their caregivers produced each of the children's first 50 verbs in past tense form. It is clear from

Table 4 about here

Table 4 that past tense forms make up a substantial proportion of both the adults' and the children's uses of these verbs (.20 to .28 and .22 to .42, respectively). However, it is also clear that they make up a higher proportion of the children's than the adults' uses in all 4 dyads, with the difference ranging from .02 in the case of Ryo to .14 in the case of Aki.

These differences were analysed using paired-sample t-tests, which revealed significant differences for Aki and Tai (both *t*s > 2.04 both *p*s < .05), but not for Ryo and Jun (both *t*s < .82, both *p*s > .42). These results therefore provide some support for the idea that Japanese children may be particularly sensitive to past tense forms in the input (and are hence arguably consistent with an RI-analogue account). However, it is also clear from Table 4 that there is considerable by-verb variation in the proportion of past tense forms in the children's and the adult's verb use, with some verbs never occurring in past tense form and others occurring in past tense form 100% of the time.

Table 5 presents simple correlations between the proportional frequency of past tense forms across the first 50 verbs, in the speech of the children, their caregivers and the average

caregiver. It can be seen from Table 5 that there are significant correlations for all of the children and their respective caregivers (all *r*s > .33, all *p*s < .02, two-tailed), suggesting that

Table 5 about here

a key determinant of the extent to which children use verbs in past tense form is the relative frequency with which those verbs occur in past tense form in the input language. However, it can also be seen from Table 5 that there are significant correlations for 3 of the 4 children with the average caregiver (all *r*s > .34, all *p*s < .02, two-tailed), the exception being Jun ($r = .17$, $p = .24$). This pattern of correlations underlines the need to control for general by-verb differences in the relative frequency of past tense forms in the target language before taking significant correlations between their relative frequency in the children and their caregivers' speech as evidence of direct effects of the input on the children's early verb use.

In order to control for this kind of general effect of the target language, partial correlations were computed between the proportional frequency of past tense forms across verbs in the children and their caregivers, while controlling for the average proportional frequency of past tense forms across verbs in the other 3 caregivers. These partial correlations are presented in

Table 6 about here

Table 6, from which it can be seen that there are dyad-specific effects in all 4 cases (all *r*s > .31, all *p*s < .04, two-tailed).

These results suggest that, rather than reflecting by-verb variation in the likelihood that particular verbs will occur in past tense form in the target language, the relationship between the children and their caregivers' use of past tense forms reflects a direct effect of the relative frequency with which different verbs occur in past tense form in the children's input. They thus provide strong support for an input-driven account of Japanese children's early use of

verb inflection.

**Discussion**

The aim of the present study was to investigate the early use of verb inflection in Japanese by testing the hypothesis that the past tense form is an RI analogue, and contrasting it with the alternative hypothesis that children's use of past tense versus other inflectional forms reflects the distributional patterning of past tense and other inflectional forms in the input. On the one hand, our results provide little support for the RI-analogue account. Thus, all of the children used a range of forms during the earliest stages of acquisition, and all of the children produced errors involving the overuse of a range of inflectional forms. On the other hand, our results do provide support for an input-driven account, with all of the children showing by-verb effects of the proportion of past tense versus other inflectional forms in their input on the proportion of their first 50 verbs that appeared in past tense form even after controlling for differences in the proportion of past tense versus other inflectional forms in input averaged across the caregivers of the other children in the sample.

These results have a number of implications for our understanding of early morphological development. First, they raise doubts about recent attempts to explain children's early use of verb morphology in non-RI languages in terms of a hypothetical RI-analogue stage. For example, Salustri and Hyams (2003) argue that, in Italian, the imperative is an RI analogue; and Grinstead et al. (2009) argue that, in Spanish, the 3sg present tense form is an RI analogue. Within such accounts, the choice of default form is assumed to be linguistically motivated. However, the imperative in Italian is also a relatively high frequency form that is homophonous with the 3sg or 2sg present tense form depending on the conjugation of the verb; and the 3sg present tense in Spanish is both the highest frequency form in the input (Aguado-Orea & Pine, 2015), and also homophonous with the imperative for regular verbs. These facts make it difficult to distinguish empirically between an RI-analogue account of

children's early use of verb morphology and a constructivist account that emphasises the role of input frequency, since they mean that the hypothetical RI analogue is likely to be the most frequent form of most, if not all, of the verbs in the input. Focusing on a language like Japanese, however, offers a way of distinguishing more clearly between these two types of account, since the relatively balanced frequency of past and nonpast tense forms in the input results in significantly more by-verb variation in the form in which individual verbs occur most frequently, with some verbs occurring more often in past tense form, and others occurring more often in nonpast tense form. What the results of the present study suggest is that when one does focus on such a language, one does not find the verb-general pattern of use predicted by the RI-analogue account, but the verb-specific pattern predicted by a constructivist account. The implication is that, to the extent that children do default to a particular form in their language, this may have more to do with the relative frequency of that form in the input than with its linguistic status as an RI analogue within the child's system. Räsänen, Ambridge and Pine (in press) present evidence from a sentence elicitation study in favour of exactly this kind of account of the pattern of defaulting errors in early child Finnish. One way of extending the findings of the present study would be to use a similar elicitation paradigm to look for verb-specific patterns of defaulting error in early child Japanese. The prediction would be that Japanese-speaking children would default from nonpast tense to past tense forms for verbs that are more frequent in past tense form in the input, and from past tense to nonpast tense forms for verbs that are more frequent in nonpast tense form in the input.

Second, our results underline the need to control for general effects of the frequency and semantic distributional properties of the language to be learned when investigating the relation between input frequency and children's early acquisition and use of particular forms in naturalistic speech samples. Many previous studies of this relation (e.g. Freudenthal et al.,

2010; Naigles & Hoff-Ginsberg, 1998; Rowland, Pine, Lieven & Theakston, 2003; Smiley & Huttenlocher, 1995; Theakston et al., 2004) have taken correlations between the frequency of particular forms or sequences in the input and the order in which those forms or sequences appear in children's speech, or between the relative frequency of different forms in the input and the relative frequency of those forms in the child's language, as evidence of direct effects of input frequency on children's acquisition and use of language. However, it has become increasingly clear in recent years that measures of order of acquisition based on naturalistic speech samples are confounded with the frequency of words and sequences in the language, and measures of the relative frequency with which children use particular forms are confounded with the semantic distributional properties of the language that they are learning. These problems imply the need to control for general differences in the frequency of words and sequences in the language and for the semantic distributional properties of the language being learned before taking relations between input frequency and order of emergence or between relative frequency in the input and relative frequency in the child's language as evidence for direct effects of input frequency on language learning.

Third, our results illustrate how it is possible to control for these potential confounds by looking for specific effects of the child's own input over and above the effects of input data averaged across caregivers. Note that this approach does not rule out all possible alternative explanations of potential input effects. For example, it does not rule out the possibility that they might reflect more subtle effects of the child on the caregiver. However, it does allow us to isolate dyad-specific relations and hence to control for general frequency differences and for general semantic-distributional properties of the language being learned. It may also be particularly useful in the study of agglutinative languages, which do not tend to show the 'inflectional imperialism' typical of languages such as English (Dressler, 2005), and in which

there is therefore likely to be considerable by-verb variation in the forms in which particular verbs are most likely to occur.

To conclude, the findings of the present study provide evidence against an RI-analogue account and in favour of an input-driven account of Japanese children's use of verb inflection. Although there are a number of factors that are likely to influence children's early verb use, our results suggest that Japanese-learning children are sensitive to the relative frequency of different inflectional forms in their parents' speech rather than just the semantic-distributional properties of the language being learned. Future work should attempt to rule out alternative explanations of these results by replicating them experimentally, as well as investigating the influence of additional factors such as semantic and pragmatic salience on the relative frequency with which particular forms of the verb are used.

## References


Aguado-Orea, J. & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS One* **10**(3), e0119613.

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**, 239-273.

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

Clancy, P. (1985). The acquisition of Japanese. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition* Vol. 1. Hillsdale, NJ: Erlbaum.

Dressler, W. U. (2005). Morphological typology and first language acquisition: Some mutual challenges. In G. Booij, E. Guevara, A. Ralli, S. Sgroi & S. Scalise (eds), *Morphology and Linguistic Typology: Online Proceedings of the Fourth Mediterranean Morphology Meeting*.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition* **24**, 143-188.

Freudenthal, D., Pine, J. M. & Gobet, F. (2010). Explaining quantitative variation in the rate of Optional Infinitive errors across languages: A comparison of MOSAIC and the Variational Learning Model. *Journal of Child Language* **37**, 643-669.

Grinstead, J., De la Mora, J., Vega-Mendoza, M. & Flores, B. (2009). An elicited production test of the Optional Infinitive stage in Child Spanish. In J. Crawford, K. Otaki & M. Takahashi (eds). *Proceedings of the 3rd Conference on Generative approaches to Language Acquisition North America*. Somerville: Cascadilla Press.

Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua* **106**(1), 81-112.

Ishii, T. (1999). *The JUN Corpus*, unpublished.

Kato, S., Sato, Y., Chikuda, Y., Miyoshi, R., Sakai, Y. & Koizumi, M. (2003). "Root infinitives": Nihongo karano kensho ["Root infinitives": From the perspective of Japanese]. *Tohoku University Linguistics Journal* **12**, 113-127. Sendai: Tohoku University.

Kim, M. & Phillips, C. (1998). Complex verb constructions in child Korean: Overt markers of covert functional structure. In A. Greenhill, M. Hughes, H. Littlefield & H. Walsh (eds), *Proceedings of the 22nd annual Boston University Conference on Language Development*. Somerville, MA. Cascadilla Press.

MacWhinney, B. (1978). The acquisition of morphophonology. *Monographs of the society for research in child development* **43**, 1-123.

MacWhinney, B. (2000). *The CHILDES project: tools for analysing talk*. Mahwah, NJ: Erlbaum.

Miyata, S. (1992). Wh-questions of the third kind: The strange use of Wa-questions in Japanese children. *Bulletin of Aichi Shukutoku Junior College* **31**, 151–155.

Miyata, S. (1995). The Aki corpus: Longitudinal speech data of a Japanese boy aged 1.6-2.12. *Bulletin of Aichi Shukutoku Junior College* **34**, 183–191.

Miyata, S. (2000). The Tai Corpus: Longitudinal speech data of a Japanese boy aged 1;5.20-3;1.1. *Bulletin of Shukutoku Junior College* **39**, 77-85.

Murasugi, K., Nakatani, T., & Fuji C. (2009). The roots of root infinitive analogues: The surrogate verb forms common in adult and child grammars. *Proceedings of the 34th annual Boston University Conference on Language Development. Online Supplement*.

Murasugi, K. (2015). Root infinitive analogues in Child Japanese. In M. Nakayama (ed.). *Handbook of Japanese Psycholinguistics* Vol. 9. Boston: Walter de Gruyter,

Naigles, L. R. & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* **25**, 95-120.

Pinker, S. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: implications for models of language development. *Journal of Child Language* **19**, 491-557.

Räsänen, S. H. M., Ambridge, B. & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest frequency form? *Journal of Child Language* **41**, 756-779.

Räsänen, S. H. M., Ambridge, B. & Pine, J. M. (in press). An elicited-production study of inflectional verb morphology in child Finnish. *Cognitive Science*.

Rispoli, M. J. (1981). *The emergence of verb and adjective tense-aspect inflections in Japanese*. Master's Thesis, Graduate School of Arts and Sciences, University of Pennsylvania.

Rizzi, L. (1993/1994). Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* **3**, 371-393.

Rowland, C. F., Pine, J. M., Lieven, E. V. M. & Theakston, A. L. (2003). Determinants of acquisition order in wh-questions: Re-evaluating the role of caregiver speech. *Journal of Child Language* **30**, 609-636.

Salustri, M. & Hyams, N. (2003). Is there an analogue to the RI stage in the null subject languages. In B. Beachley, A. Brown & F. Conlin (eds), *Proceedings of the 27th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Shibatani, M. (1990). *The languages of Japan*. Cambridge University Press.

Shirai, Y. (1998). The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language* **18**, 281-309.

Shirai, Y., & Andersen, R. W. (1995). The acquisition of tense-aspect morphology: A prototype account. *Language* **71**, 743-762.

Shirai, Y., & Miyata, S. (2006). Does past tense marking indicate the acquisition of the concept of temporal displacement in children's cognitive development? First *Language* **26**, 45-66.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (eds), *Studies of child language development*. New York: Holt, Rinehart & Winston.

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition* Vol. 2. Hillsdale NJ: Erlbaum.

Smiley, P. & Huttenlocher, J. (1995). Conceptual development and the child's early words for events, objects and persons. In M. Tomasello & W. Merriman (eds.), *Beyond names for things: The acquisition of verbs*. Hillsdale, NJ: Erlbaum.

Theakston, A. L., Lieven, E. V., Pine, J. M. & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language* **31**, 61-99.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* **4**(4), 156-163.

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough? *Journal of Child Language* **31**, 101-122.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivation in child grammar. In N. Hornstein & D. Lightfoot (eds), *Verb movement*. Cambridge, England: Cambridge University Press.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* **106**, 23-79.

Yang, C. (2013). Ontogeny and phylogeny of language. *Proceedings of the National Academy of Science*s **110**, 6324-6327.

Table 1. Some basic inflectional forms in Japanese

| | *kik-* 'listen' (consonant-ending verb) | *mi-* 'look at' (vowel-ending verb) |
|---|---|---|
| Nonpast | *kik-u* | *mi-ru* |
| Past | *kii-ta* | *mi-ta* |
| Conjunctive/imperative | *kii-te* | *mi-te* |
| Conditional | *kik-e-ba* | *mi-re-ba* |
| Hortative | *kik-oo* | *mi-yoo* |
| Passive Nonpast | *kik-a-re-ru* | *mi-rare-ru* |
| Causative Nonpast | *kik-a-se-ru* | *mi-sase-ru* |
| Potential Nonpast | *kik-e-ru* | *mi-(ra)re-ru* |
| Progressive Nonpast | *kii-te-ru* | *mi-te-ru* |
| Negative Nonpast | *kik-a na-i* | *mi na-i* |

Table 2. The first 10 verb forms in Aki, Ryo and Jun's corpora

| Aki | | | Ryo | | |
|---|---|---|---|---|---|
| Age | Verb form | Meaning | Age | Verb form | Meaning |
| 20 | *atta* | be-PAST | 22 | *atta* | be-PAST |
| | *miru* | see-NONPAST | | *dechatta* | get.out-COMPLETIVE-PAST |
| 22 | *iku* | go-NONPAST | | *deta* | get.out-PAST |
| 24 | *doite* | step.aside-IMPERATIVE | 23 | *yatte* | do-IMPERATIVE |
| | *aru* | be-NONPAST | | *haitta* | enter-PAST |
| | *mieta* | be.in.sight-PAST | | *i nai* | be-NEG-NONPAST |
| 25 | *tatte* | stand-IMPERATIVE | | *aru* | be-NONPAST |
| | *deta* | get.out-PAST | | *deru* | get.out-NONPAST |
| | *itchatta* | go-COMPLETIVE-PAST | | *haira nai* | enter-NEG-NONPAST |
| | *dete* | get.out-CONNECTIVE | | *tot-te* | take-IMPERATIVE |

Table 2. The first 10 verb forms in Aki, Ryo and Jun's corpora (Continued)

| Jun | | |
|---|---|---|
| Age | Verb form | Meaning |
| 20 | *atta* | be-PAST |
| 22 | *shita* | do-PAST |
| 23 | *itchatta* | go-COMPLETIVE-PAST |
| | *itte* | go-CONNECTIVE |
| | *shite* | do-CONNECTIVE |
| 25 | *ita* | be-PAST |
| | *aru* | be-NONPAST |
| | *mita* | see-PAST |
| 26 | *matte* | wait-CONNECTIVE |
| 27 | *chigau* | be.different-NONPAST |

Table 3. Rates of inflectional errors in the 4 children's data

| | Aki | Ryo | Tai | Jun |
|---|---|---|---|---|
| Past | 11/545 (2.02%) | 4/263 (1.52%) | 2/1281 (0.16%) | 3/612 (0.49%) |
| Nonpast | 26/1034 (2.51%) | 5/439 (1.14%) | 3/2390 (0.13%) | 4/925 (0.43%) |
| Connective | 8/114 (7.02%) | 3/68 (4.41%) | 3/753 (0.40%) | 1/104 (0.96%) |
| Imperative | 5/240 (2.08%) | 1/140 (0.71%) | 0/520 (0%) | 0/254 (0%) |
| Total | 50/1933 (2.59%) | 13/910 (1.43%) | 8/4944 (0.16%) | 8/1895 (0.42%) |

Table 4. Means and standard deviations for the proportion of past tense forms in the children and their caregivers' use of the child's first 50 verbs

| | Child | | | Caregiver | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Aki | 0.42 | 0.36 | 0-1 | 0.28 | 0.23 | 0-0.79 |
| Ryo | 0.29 | 0.26 | 0-1 | 0.28 | 0.26 | 0-1 |
| Tai | 0.32 | 0.28 | 0-1 | 0.27 | 0.25 | 0-0.87 |
| Jun | 0.22 | 0.26 | 0-1 | 0.20 | 0.22 | 0-0.85 |

Table 5. Simple correlations between children and caregivers' proportional use of past tense forms for each of the child's first 50 verbs

| | | Aki | Ryo |
|---|---|---|---|
| Child and child's caregiver | Pearson Correlation | .599 | .883 |
| | Sig. (2-tailed) | .001 | .001 |
| | df | 48 | 47 |
| Child and average input | Pearson Correlation | .392 | .346 |
| | Sig. (2-tailed) | .005 | .014 |
| | df | 48 | 48 |
| | | Tai | Jun |
| Child and child's caregiver | Pearson Correlation | .677 | .337 |
| | Sig. (2-tailed) | .001 | .018 |
| | df | 48 | 47 |
| Child and average input | Pearson Correlation | .511 | .173 |
| | Sig. (2-tailed) | .001 | .238 |
| | df | 48 | 46 |

Table 6. Partial correlations for each of the 4 child-caregiver pairs, controlling for relative frequency in the input averaged across the other 3 caregivers

| Correlation between | Correlation | Significance (2-tailed) | df |
|---|---|---|---|
| Aki and Aki's mother | 0.492 | <.001 | 47 |
| Ryo and Ryo's mother | 0.851 | <.001 | 46 |
| Tai and Tai's mother | 0.523 | <.001 | 47 |
| Jun and Jun's parents | 0.318 | .031 | 44 |