

ARTICLE

Can prosody encode recursive embedding? Children's realizations of complex NPs in Japanese

Manami HIRAYAMA^{1*}, Laura COLANTONI², and Ana Teresa PÉREZ-LEROUX²

¹Seikei University, Japan and ²University of Toronto, Canada

*Address for correspondence: Manami Hirayama Department of English Faculty of Humanities Seikei University 3-3-1 Kichijoji Kitamachi Musashino-shi, Tokyo Japan 180-8633, E-mail: hirayama@fh.seikei.ac.jp

(Received 17 September 2018; revised 10 June 2019; accepted 25 May 2020;
first published online 29 June 2020)

Recursive NPs are difficult to produce and late to emerge. We compare prosodic and syntactic abilities in Japanese-speaking five- and six-year-olds ($n = 28$) and adults ($n = 10$). It is reported that syntactic structure in Japanese is prosodically marked via downstep and metrical boost. Results of an elicited imitation task suggested that children had acquired the lexical prosody (contrast between accented and unaccented words), a pre-requisite for downstep realization. While downstep, the prosodic phrasing involved in the complex NPs in this study, was established, children showed interspeaker variation with the metrical boost, a feature that distinguishes recursively embedded NPs from non-recursive NPs. However, variability was also found in adults, indicating that, in contrast to previous results, prosodic encoding of syntax is generally unreliable in adult speech. Finally, the magnitude of metrical boost was not correlated to children's ability to produce recursive possessives, suggesting that prosody does not help bootstrap Japanese children's recursive phrases.

Keywords: prosody; Japanese; complex NP

Introduction

Prosody plays a key role in the organization of language, serving to structure words, phrases and sentences. It also correlates with the distinctions of meaning associated with sentence types, information structure, and phrasing. Phrasal structure is often correlated with intonational cues (e.g., changes in the fundamental frequency), and serves, albeit inconsistently, to disambiguate structural ambiguities (Snedeker & Trueswell, 2003). In this study, we focus on the mastery of prosody associated with a particular type of structural ambiguity that arises when two modifiers appear within a noun phrase. Double modification can either be recursive, as in (1a) and (1b), when a noun is restricted by a noun phrase (NP) modifier that is itself restricted by another nominal modifier, or non-recursive, as in (1c), when two modifiers restrict the same head NP.

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

- (1) a. [[Elmo's] sister's] ball. (Recursive modification)
 b. The baby [with the woman [with the flowers]]. (Recursive modification)
 c. The plate [under the table] [with oranges]. (Double non-recursive modification)

Recursively embedded NPs are relatively delayed in child language. Data from English shows that three-year-olds can easily produce three prosodically integrated coordinated NP constituents, as in (2b) (Pérez-Leroux, Castilla-Earls, Bejar & Massam, 2012), and most can produce NPs with one modifier, as in (2a). In contrast, production of recursive NP embedding emerges much later. Three-year olds cannot produce them, and even five-year-old children often fail to produce these targets.

- (2) a. Elmo's ball. (Single-level modification)
 b. The doll, the ball, and the girl. (Coordinated NPs)

Moreover, recursively modified NPs (1a, b) are produced at much lower rates and learned later than lexically comparable NPs containing double non-recursive modifiers (1c) (Pérez-Leroux, Peterson, Bejar, Castilla-Earls, Massam & Roberge, 2018). This suggests that beyond the inherent complexity associated with planning production of noun phrases with modifiers, additional levels of embedding such as those found in recursive configurations impact children's limited sentence-planning abilities (McDaniel, McKee & Garret, 2010).

Can prosody boost the acquisition of recursive NPs? We could hypothesize that there is a positive correlation between marking recursion prosodically and syntactically; namely, if children can mark recursion in prosody, they may also be able to produce recursively embedded NPs in an elicitation task. To test this general hypothesis, we investigate the association between mastery of prosody and ability to produce recursive NPs in Japanese-speaking children. Indeed, Japanese offers a unique opportunity to explore the acquisition of the syntax-prosody interface, since both its lexical phonology (accented vs. unaccented) and phrasal phonology interact to mark syntactic structure.

Let us start by examining the first phenomenon discussed here: *SYNTAGMATIC DOWNSTEP* or *CATATHESIS*, as exemplified in (3a) and illustrated in Figure 1a. In (3a), we have a sequence of three lexically accented words, whose f₀ contour (Figure 1a) shows a stepwise lowering of pitch register (e.g., Beckman & Pierrehumbert, 1988; Kubozono, 1989; Poser, 1984; Selkirk & Tateishi, 1991). Note that in Tokyo Japanese, words can either be accented displaying a high-low (HL) f₀ contour (see the contour of NP1 in (3a)) or unaccented, and thus showing no such steep fall but rather a plateau (NP1 in (3b)). The staircase pattern shown below is restricted to domains larger than the prosodic word, such as the Major Phrase (Beckman & Pierrehumbert, 1988; Ito & Mester, 2013; Féry, 2016; Ishihara, 2016). Thus, if accented words are grouped or phrased together, we expect to see downstep. However, Japanese expresses a phonological contrast between initially-accented phrases (3a) and their unaccented counterparts (3b). In the latter case, illustrated in Figure 1b, we see that the fall from the peak in the first NP to the second NP is practically negligible, displaying the absence of downstep. Then, when we compare the downward f₀ movements in a sequence of accented words (Figure 1a) with those observed in sequence of unaccented-accented words (Figure 1b), we are talking about another

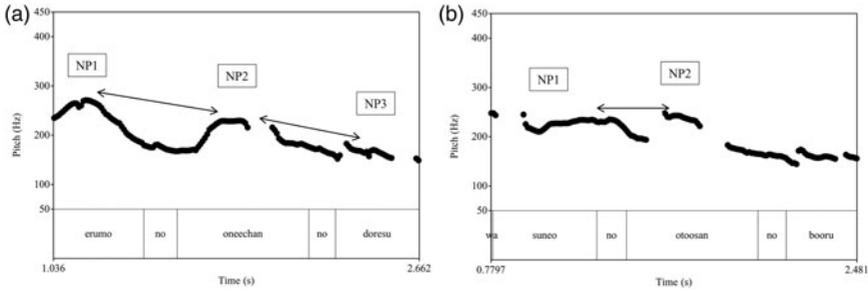


Figure 1. Pitch tracks illustrating downstep with accented words (a), and absence of downstep (b) when the initial word is unaccented (female adult speaker)

understanding of downstep known as PARADIGMATIC DOWNSTEP (e.g., Ishihara, 2016; Ito & Mester, 2013).

- (3) a. Érumo-no onéechan-no dóresu (AAA)¹
 Elmo-no sister-no dress
 ‘Elmo’s sister’s dress’
- b. Suneo-no otóosan-no booru (UAU)
 Suneo-no father-no ball
 ‘Suneo’s father’s ball’

The second phenomenon we will explore, METRICAL BOOST, is exemplified in the contrast between (4a) and (4b), whose prosodic realization is illustrated in Figure 2. In a recursive structure consisting of a sequence of three accented words (4a), we expect to see a sharp pitch fall from the first to the second NP (Figure 2a). When we compare this with a non-recursive structure (4b and Figure 2b), we see a different f0 pattern: here the second peak is almost at the same height as the first, a phenomenon Kubozono (1989) has dubbed METRICAL PITCH BOOST. Thus, the magnitude of pitch boost signals the difference between recursive and non-recursive structures.

- (4) a. Left-branching, Recursive structure:
 Érumo-no onéechan-no dóresu
 [[[elmo-no] sister-no] dress]
 ‘Elmo’s sister’s dress’
- b. Right-branching, Non-recursive structure:
 Érumo-no orénji-no dóresu.
 [[elmo-no] [orange-no] dress]
 ‘Elmo’s orange dress’²

Metrical pitch boost has been reported (Kubozono, 1989) as a property of Tokyo Japanese, a typical left-branching language. In Kubozono’s analysis, which was based on extensive reading materials collected by a single male speaker representative of

¹The acute accent mark indicates lexical accent.

²In Japanese, many color words are nouns, so the phrasal status of all modifiers in (4) is identical.

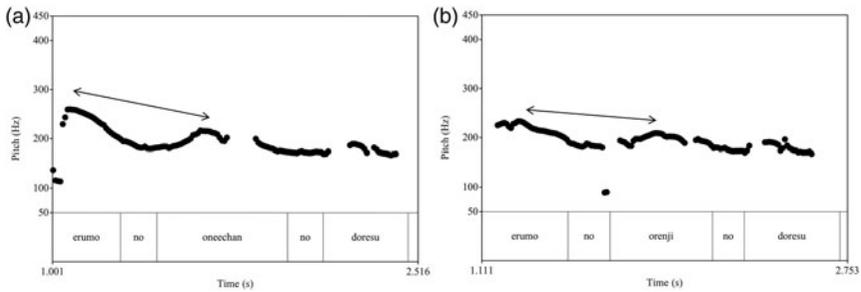


Figure 2. Pitch track and text grid illustrating recursive (4a) and non-recursive (4b) NPs (female adult speaker)

the dialect, he reported pitch differences between left- and right-branching structures (low and high phrasal attachment, respectively). In sequences of three accented phrases, the f_0 peak of the second phrase is higher in a right-branching configuration (as in Figure 2b), where the first phrase is not embedded within the second, than in a recursive (left-branching) configuration (Figure 2a). As stated by Féry (2010): “[...] quite an impressive literature has shown that Japanese needs recursion of its higher prosodic domains, namely the ones which are mapped to morpho-syntactic structures”.

In sum, evidence shows that Japanese clearly marks phrasing (syntactic grouping) using downstep in sequences of accented phrases, and marks recursion by a paradigmatic distinction between recursive structures. Hypothetically, these prosodic cues assist children in the acquisition of nominal recursion. Thus, the first goal of our study is to determine whether Japanese children have acquired the prosodic patterns that mark the distinction between sequences of accented and non-accented phrases, as well as the prosodic marking that distinguishes recursive and non-recursive NPs. Before we proceed to our study, we briefly review what is known about the acquisition of syntactic recursion.

Children’s acquisition of nominal recursion

Children learn the basic relations of possession and modification quite early. However, initial studies on the emergence of NPs containing recursive modification show that children generally do not produce such complex structures until age four. Between the ages of four and six, most children start producing recursive NPs, but do so at far lower rates than adults. This is true of English (Pérez-Leroux *et al.*, 2012; 2018), French (Roberge, Pérez-Leroux & Frolova, 2018), Spanish (Pérez-Leroux, *in press*), and Mandarin (Giblin, Zhou, Bill, Shi & Crain, 2019).

In Japanese, possessors and other modifiers are linked by a simple particle *no*, which children start using by the age of two, and tend to overgenerate by the age of three (Clancy, 1986; Murasugi, Nakatani & Fuji, 2012). There is no previous data on when Japanese children start combining multiple instance of *no*-modification into recursive NPs, but there are a few comprehension studies, which suggest that Japanese children are no different from others (Fujimori, 2010). As children in other languages (Limbach & Adone, 2011; Roeper, 2011), they often give partial or coordinated interpretations (Limbach & Adone, 2011; Roeper, 2011). Recent work by Terunuma and Nakato (2018) suggests an initial stage where children only

comprehend a single possessor (*Midori-no inu* ‘Midori’s dog’). Subsequently they succeed with multiple (two or more) possessors (e.g., *Orenji-chan-no inu-no fuusen* ‘Orenji’s dog’s balloon’). By the age of five, comprehension of double possessors is very high (>90%) (Terunuma & Nakato, 2018; Terunuma, Isobe, Nakajima, Okabe, Inaba, Inokuma & Nakato, 2017).

As such, the second goal of our study is to test to what extent Japanese kindergarten age children can produce recursive possessive NPs and, if so, to explore to what extent mastery of the prosodic contrast predicts their ability to produce recursive NPs.

Children’s acquisition of prosody

A robust body of work shows that children use prosodic cues in the speech signal (intonation, duration, rhythm) in the acquisition of grammar, supplementing and enhancing the distributional cues available to them. This is known as the prosodic bootstrapping hypothesis (Christophe & Dupoux, 1996; Christophe, Millotte, Bernal & Lidz, 2008), or, more recently, the phonological bootstrapping hypothesis (van Heugten, Dautriche & Christophe, 2014). Prosodic cues are sufficiently robust in the input to children (e.g., Athanasopoulou & Vogel, 2016; Soderstrom, Blossom, Foygel & Morgan, 2008), and children rely on these cues in word learning and for learning functional morphology (Christophe *et al.*, 2008; Jusczyk, Hirsh-Pasek, Nelson, Kennedy, Woodward & Piwoz, 1992; Christophe, Mehler & Sebastián-Gallés, 2001; Soderstrom, Seidl, Kemler Nelson & Jusczyk, 2003). For instance, de Carvalho, Dautriche, and Christophe (2016) found that French-speaking children as young as 3;05 use prosodic information to disambiguate pairs of noun/verb homophones.

At the word level, studies document early sensitivity to the role of lexical stress (e.g., Curtin, 2009; Archer & Curtin, 2016), although production of lexical stress seems to lag (Ota, 2016, p. 70; Kehoe, 1997; Lleó & Arias, 2006; but see Kehoe, Stoel-Gammon & Buder, 1995; Hochberg, 1988; Prieto, Estrella, Thorson & Vanrell, 2012). In languages that use pitch at the lexical level, like Japanese and Mandarin, studies have reported variation among ages of acquisition. For Mandarin, which has a rich tonal inventory, mastery is late, by the age of six or later (Ching, 1984; Wong, 2009; Wong & Leung, 2018). In Japanese, where lexical accent is realized by a single tonal pattern (H*L), children reach adult levels in disyllabic words by the age of two (Ota, 2003, 2006). Studies report dialectal variability in the acquisition of accentuation rules in compounds by preschoolers (Shirose, 2001; Shirose, Kakehi & Kiritani, 2001).

As we move beyond lexical prosody to sentence intonation, the empirical picture becomes less coherent. Although there is agreement that pitch control stabilizes in the first year (Vihman, 1996, p. 212), some studies claim that children have mastered most of the intonational contours of the language around the two-word stage (e.g., Snow, 2007; Snow & Balog, 2002; Balog & Snow, 2007; Prieto *et al.*, 2012; Astruc, Payne, Post, Vanrell & Prieto, 2013; Vihman, 1996), while others argue for a protracted period of development, possibly continuing until pre-puberty or even puberty (Filipe, Peppé, Frota & Vicente, 2017; Yang & Chen, 2018). Yang and Chen (2018) find that eleven-year-olds resemble adults in their use of durational cues but differ in their use of pitch to mark focus in Mandarin Chinese. A study of eight high-level prosodic categories in European Portuguese shows mastery is late and non-uniform across structures. Children first learn to signal the end of a turn (around age nine), but prosodic marking of focus only settles in late teens. Data for Spanish-speaking children are similar (Filipe *et al.*, 2017, p. 1063). Contrastingly,

English-speaking children acquire focus marking much earlier (by the age of five), while still lagging in comprehension (Wells, Peppé & Goulandris, 2004). Furthermore, children acquire sensitivity to the overall contours of their language early (Frota, Butler & Vigário, 2013), but only master the functions of such contours later (Ota, 2016; Saindon, Cirelli, Schellenberg, van Lieshout & Trehub, 2017).

The picture that emerges from the discussion above is that children take time to master sentence-level intonation patterns. Can children use the prosodic properties of words to disambiguate phrasal structure? De Carvalho *et al.* (2016) found that young French-speaking children can distinguish pairs of noun/verb homophones on the basis of prosody. For Korean, Choi and Mazuka (2003) reported that, while the prosodic marking of word-segmentation is acquired around the ages of three and four, six-year-olds cannot yet use prosody to sort out phrasal chunking. Compounds and coordination offer important evidence. In English, the stress shift that signals the contrast between compounds (*'hot,dog*) and phrasal patterns (*hot 'dog*) is not fully acquired until after nine years, and, even then, remains greatly affected by a bias towards compounding (Vogel & Raimy, 2002). In other studies, English-speaking children are able to differentiate compounds and noun phrases (*fruit-salad and milk* vs. *fruit, salad and milk*) by either the age of five (Wells *et al.*, 2004) or later (Katz, Beach, Jerouri & Verma, 1996). In Portuguese, Filipe *et al.* (2017) report that it takes until adolescence to master correct use of prosodic cues to signal these chunking differences. Spanish-speaking children may reach adult levels earlier, by the age of ten (Martínez-Castilla & Peppé, 2008).

Few studies target syntactic structure (chunking) without involving lexical prosody; those deal primarily with coordination. Cruttenden (1985) considered structural ambiguities in the ellipsis context, finding that ten-year-olds performed poorly in distinguishing constituent structure in ellipsis environments on the basis of prosody (*She [[dressed and fed] the baby]* vs. [*She dressed] and [fed the baby]*). In other studies, five-year-olds were able to use prosody for comprehension of constituency in coordinated sequences of adjectives (e.g., [*pink and green] and white]* vs. [*pink and green and white]*) (Beach, Katz & Skowronski, 1996), but failed to mark this contrast in production (Katz *et al.*, 1996). One common observation in all these studies is substantive individual variation (Wells *et al.*, 2004).

A separate body of evidence on the relationship between phrasal structure and prosody has emerged from the psycholinguistics literature on online resolution of syntactic ambiguities. Those studies focus specifically on the incremental processing of PPs that could potentially be interpreted as VP adjuncts (with high attachment, [*tap [the frog] [with the flower]*]) or as modifiers to direct objects (with low attachment, as in [*tap [the frog [with the flower]*]). The initial study by Snedeker and Trueswell (2003) found that English-speaking preschoolers were not sensitive to this difference. Subsequent studies found sensitivity once researchers controlled for lexical biases, which emerged when children had to process verbs typically used with an instrument (Snedeker & Yuan, 2008). The patterns of gaze fixation in four- to six-year-olds show an effect, but in a later temporal window than in adults.

To sum up, the question of children's mastery of the association between intonation and syntax has not received a consistent answer: results vary across language and structures. For English, results indicate that five-year-olds can use prosody to disambiguate the attachment of a PP modifier, and to disambiguate the scope of adjectives in coordination. However, they fail to use prosody to disambiguate clausal coordination involving ellipsis (admittedly, a more complex structure). Japanese

children learn lexically related pitch fall patterns early. Whether they can consistently distinguish between different configurations of phrases is an open question. Our first goal is to investigate whether Japanese children master the prosodic realization of a sequence of accented phrases (downstep) and the structural embedding of multiple modifiers (metrical boost).

Research questions and hypotheses

The goal of this study is two-fold. First, the prosodic component investigates how Japanese children prosodically realize complex nominal phrases. Then we examine the interaction between prosody and syntax. For these purposes, we ask the following questions:

1. Do Japanese children have an adult-like production of prosody in complex NPs (downstep)? (Figure 1a vs. Figure 1b)
2. Do Japanese children mark phrasal attachment with prosody, i.e., the metrical boost or pitch reset pattern described by Kubozono (1989)? (Figure 2a vs. Figure 2b)
3. Is there an association between prosodic ability, as described above, and the ability to produce recursive NPs?

To answer the questions above, we begin by examining whether children display the expected patterns at the lexical level: that is, whether they demonstrate appropriate pitch range in lexically accented and unaccented words. This is an important preliminary step since the target-like realization of lexical pitch accents is a precondition for downstep. Second, we analyze the f_0 peaks to determine if downward pitch contour patterns in recursive sequences of accented nominals (as in Figure 1) appear as expected. Question 1 is answered through the comparison of the pitch contour in AAA sequences, with their expected staircase pattern, and in UAU recursive sequences. Third, to address Question 2, we keep the accentual properties fixed by examining AAA structures only, and test for differences between recursive and non-recursive structures such as pitch reset or metrical boost. Finally, to answer Question 3, we examine the association between children's mastery of prosodic contrast between recursive and non-recursive structures and their overall ability to produce recursive possessive NPs in a referential elicitation task.

Methods

Participants

To answer our questions, we compare a group of 10 participants from Tokyo or nearby areas (5 women; age range: 21 to 36; mean 27;04; SD 5;08) to 28 kindergarten-aged children from the same areas. One child who did not complete the session was removed from the analysis. The mean age of the remaining children ($n = 27$; 16 girls) was 5;10.3 (SD = 5.2 months), ranging from 5;00 to 6;07.³ We targeted this range because the literature shows that, by age five, children are beginning to produce recursive NPs productively in other languages, and already comprehend recursive possessives in Japanese, as mentioned in the introduction. Moreover, at this age, children are expected to have acquired the distinction between accented and unaccented words.

³See data analysis section for more details on the participants.

Procedure

After parental consent was obtained, children were tested at their homes or their friends' homes. For adults, recordings took place in a room with sound-attenuated walls. Individual sessions were recorded with a Marantz digital recorder (PDM 661), using a sampling rate at 44.1k Hz and 24-bit quantization, and an Audio Technica cardioid condenser lavalier microphone (AT831b, frequency response: 40–18,000 Hz).

Task and materials

This study includes two tasks. We used a context-based imitation task (the SENTENCE REPETITION TASK below) to elicit the appropriate prosodic structures. Imitation tasks enable control of both lexical items and interpretation, and do not tax children's production capacities. The second task was a referential elicitation task designed to determine whether participants could produce the syntactic structures under analysis.

Sentence repetition task

Identical, lexically-controlled tokens needed for prosodic analysis were elicited by presenting participants with a set of PowerPoint slides (see Figure 3), each containing a set of simple pictures with a Robot image. The robot spoke with a flat voice, and the children were asked what the robot had said. Stimuli consisted of eight target pairs of sentences (see Appendix), matched as closely as possible for phonological patterns, including the number of moras and accent patterns, and two additional simple sentences (a 'noun-*no* noun' phrase and a transitive sentence), which served as practice trials at the beginning of the session.

The sentence repetition task contained four structures, controlled according to constituent analysis (high or low attachment) and locus of structural ambiguity (NP internal or NP/VP), for a total of 16 trials. VP trials are not reported here. The current study focuses on four pairs of NP sentences, which contained sequences of three nouns linked by the particle *no* (see Figure 4 for a summary of our stimuli design).

The stimuli were presented as an audio recording of a flat-pitched robot voice. Similar tasks have been successfully used to study the acquisition of different prosodic structures (e.g., focus) by four-year-olds and older (e.g., Arnhold, Chen & Järvi­kivi, 2016; Chen & Höhle, 2018; Sauermann, Höhle, Chen & Järvi­kivi, 2011). The flat-pitched stimuli were created by composing the test phrases from independently read morphemes (e.g., nouns and particles). These were recorded by the first author, a female speaker from Tokyo, aiming to produce the same pitch for each morpheme as much as possible. The recording was done using a digital recorder (Marantz PMD 661), with the sampling rate at 44.1 kHz and 24 bits quantization, and a unidirectional condenser headset (Countryman Isomax; frequency response 20–20,000 Hz). The sound files obtained were then manipulated so that each mora was 250 ± 10 ms long (i.e., 240–260 ms), with the exception of syllables with a long vowel ((C)VV), which were manipulated to be 500 ± 20 ms. This duration manipulation was done using Praat (Boersma & Weenink, 2014; ver. 5.3.64). Full cycles of the acoustic signal were targeted in the manipulation so that there was no unnatural transition in the file. The morphemes were then concatenated, leaving 300 ms of silence between them. The target phrases were put in the carrier phrase *sore-wa ___ desu* 'that is ___', which was also created with a flat pitch following the same procedure described above. Finally, the full utterance

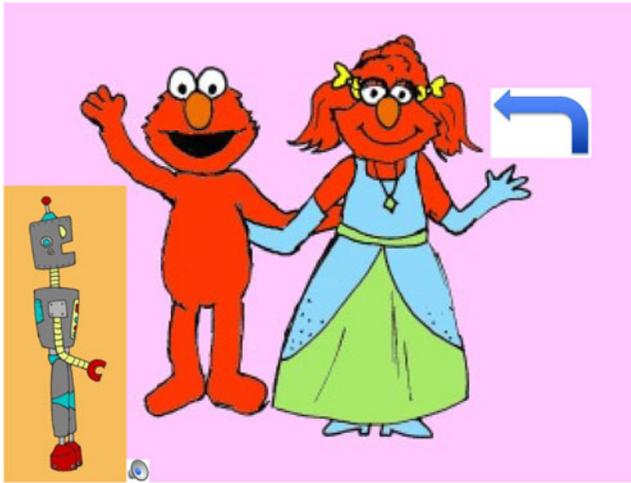


Figure 3. Computer screen for *sore-wa érumo-no onéechan-no dóresu desu* 'That is Elmo's sister's dress'

(*sore-wa NP1-no NP2-no NP3 desu*) was manipulated to sound as if pronounced by a robot. To achieve this, the following functions and settings were employed in Audacity (ver. 2.0.5): Noise removal (10 dB reduction, sensitivity 0.00, 150 Hz, 0.15 sec), Change pitch (−15%), Echo (Delay time 0.015, Decay factor 0.65), Change tempo (−10%), and Echo (same settings as above). The resulting sentence was decidedly robot-like and artificial, and monotonous. Indeed, the mean pitch range of the four sentences (i.e., the phrases in Figure 4 in the carrier phrase) was 20.5 Hz (SD: 0.57).

The image was first presented on the screen, then the robot voice was played. Repetitions were elicited with the following prompt, issued by the experimenter:

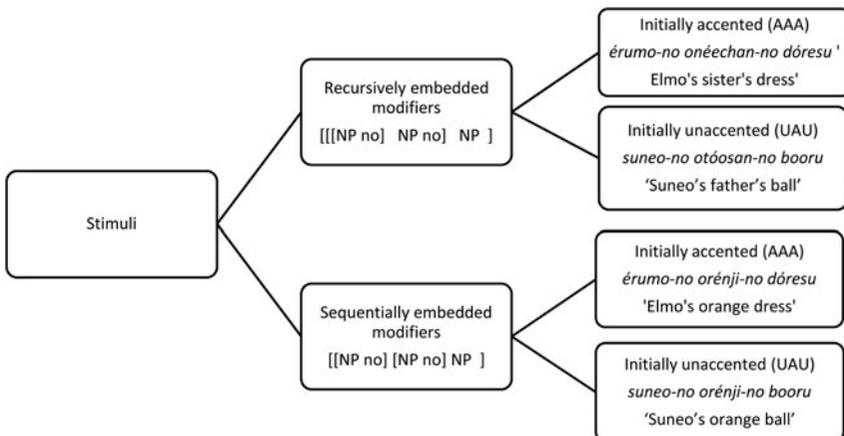


Figure 4. Stimuli design in the sentence repetition task

- (5) *róbotto-wa nán.te it-ta-kana?*
 robot-TOPIC what say-PAST-QUESTION
 ‘What did the robot say?’

Participants could listen again to the robot if they were unable to repeat the first presentation of the stimuli. They were instructed to repeat the sentences with a normal voice. A couple of practice images were used to familiarize participants with this task.

Elicited production task

A referential elicitation task was adapted for Japanese from Pérez-Leroux *et al.* (2012; 2018) to target the elicitation of doubly modified NPs. Each trial started with a brief narrative, accompanied by pictures, which contained multiple items of the same kind. The story served to draw attention to the target referent and introduce the relevant properties that distinguished the target from the other elements. For instance, in a scenario where four characters are going to play baseball, the characters are Elmo and his sister, and Kermit and his sister. Each girl carries a ball. One of the balls is flat. In Japanese, simple possession was used to describe the kinship relationship (e.g., *érumo-no onéechan*). A prompt question of the form *which x is y?* served to elicit a recursive possessive, since both kinship and ownership were required to unambiguously describe the damaged ball. An example is given in (6):

- (6) Prompt: *náni-ga kúuki-ga nukechat-ta*
 what-NOM air-NOM flatten-PAST
 ‘What is flat?’
 Target: *érumo-no onéechan-no booru.*
 Elmo-no sister-no ball
 ‘Elmo’s sister’s ball’

There were five different conditions: four recursive conditions and one non-recursive. There were six trials per condition, plus additional distractors for a total of thirty-two elicited production items. The present article reports only the one condition that was a direct analog of the intonation task: the recursive possessives condition.

Data analysis

Only a subset of children produced repetitions that could be prosodically analyzed. While all adults’ repetitions were successful, some of the child data could not be included for acoustic analysis. Some children seemed unable to understand some of the words in the robot voice, while others fully imitated the robot-like flat pitch. Those trials were not included among successful repetitions in the prosodic analysis. Instances of segmental alteration and/or incrementations which did not alter the accentuation of the phrase (e.g., syllable substitution as in *émbo-no* instead of *érumo-no*, or inclusion of a diminutive as in *suneokun-no* instead of *suneo-no*) were included for analyses. However, productions which had missing words or included words that altered the phrasal accent pattern (e.g., *orenjiro-no* ‘orange colour’ (U) instead of *orénji-no* (A)) had to be eliminated from the analysis. Therefore, prosodic analyses were conducted only on the subsample of children who produced the

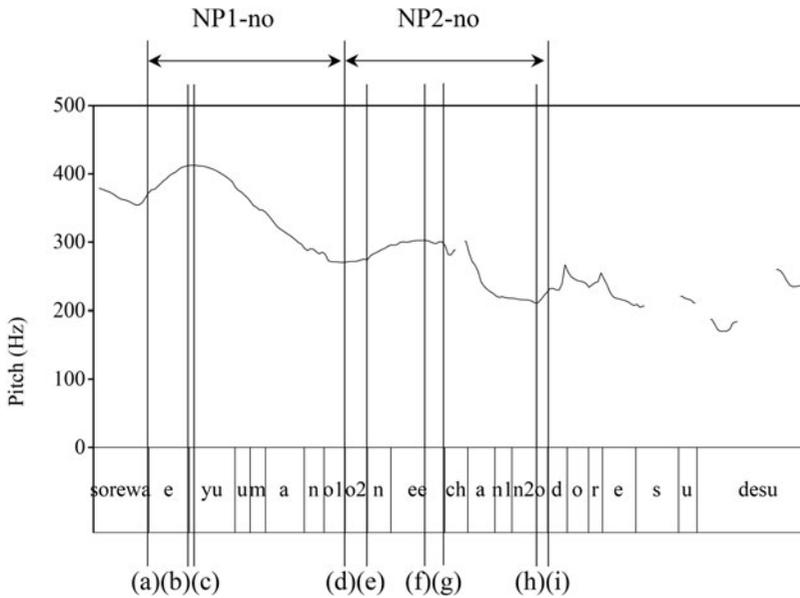


Figure 5. Example of acoustic measurements used in this paper. For each NP, the figure shows the mark inserted at the beginning of the rise, at the peak, and at the end of the rise.

relevant comparisons (12 girls and 7 boys). Since not all 19 participants successfully repeated the items to test both structures, each analysis reports on the subset of children that produced analyzable repetitions.

The first step of the acoustic analysis involved an examination of word prosody in order to determine whether children had acquired the phonological contrast between accented and unaccented words. As the reader may recall, this is a prerequisite for the acquisition of downstep and metrical pitch boost. Thus, we calculated the pitch range in the initial NP ($f_{0\max} - f_{0\min}$, i.e., f_0 at (c) minus f_0 at (d) in Figure 5). If the contrast has been acquired, we expect a larger pitch range in accented words than in unaccented words.

In order to investigate downstep and metrical boost, we analyzed the pitch fall between peaks, i.e., the difference between the f_0 peak associated with the left most noun in the sequence (NP1) and the peak associated with the subsequent noun (NP2) (difference between f_0 at point (c) and f_0 at point (f) in Figure 5). F_0 values were measured in semitones (ST). For example, a difference of 3ST indicates that the f_0 peak in NP1 is 3STs higher than the peak f_0 in NP2. F_0 values were measured in STs in order to facilitate comparisons between pitch movements in female and male speakers, as well as to allow us to determine whether the differences obtained were perceptually relevant (t'Hart, Collier & Cohen, 1990, p. 23–24). Stevens (1998: 228) suggests that differences smaller than 1.5 ST could be perceptually relevant for vowels. Thus, we set our threshold at 1 ST, treating differences of 1 ST or larger between the first and second peak as perceptually meaningful. As we show below, differences were indeed larger than this threshold.

Table 1. Measurements taken in the Sentence Repetition task by prosodic property

Property	Stimuli Selected	Measurements
Accent realization (pitch range)	All NP1s: AAA/UAU Recursive/ Non-recursive	f ₀ peak (f _{0max}) & f ₀ valley (f _{0min}) associated with accented phrase Pitch range (f _{0max} -f _{0min}) Expected: AAA range > UAU range
Downstep (syntagmatic)	AAA Recursive	f _{0max} in NP1 and NP2 Pitch change = NP1 max-NP2 max Expected: downstep NP1 > NP2
Downstep (paradigmatic)	AAA Recursive vs. UAU Recursive	f _{0max} in NP1 and NP2 Pitch change = NP1 max-NP2 max Expected: downstep AAA > UAU
Metrical boost	AAA Recursive vs. AAA Non-Recursive	f _{0max} in NP1 and NP2 Pitch change = NP1max-NP2max Expected: boost Rec > Non-rec

Table 1 summarizes the properties under study in the Sentence Repetition task, the measurements taken, and the expected results given available descriptions of adult speech in Japanese.

We conducted four separate sets of statistical analyses on the prosodic data. The first analysis explored whether children had acquired the lexical prosody of Japanese. This consisted of examining the realization of accented and unaccented words in terms of pitch range (the difference between pitch peaks and valleys, i.e., f_{0max} and f_{0min}). The second analysis explored the presence of syntagmatic downstep in each group by testing the significance of magnitude of f₀ differences in the accented sequences of recursive structures. The third analysis compared f₀ differences between AAA and UAU structures in recursive stimuli, i.e., paradigmatic downstep, for each group. When group differences were found, we tested whether the effect of accent type was comparable across groups. The fourth analysis considered the effect of syntactic condition (recursive/non-recursive) for each group, i.e., the metrical boost.

One additional analysis examined the association between prosodic abilities and syntactic productive abilities (Task 2: Elicited Production). Each possessive phrase was transcribed, and results were labeled as either target (recursive) or non-target, which could include incomplete realizations, alternative phrasing, etc. Then, we entered the proportion of target responses in the Elicited Production tasks as an indicator of syntactic ability, and we compared this with children's metrical boost scores, which were treated as indicators of their ability to encode recursion prosodically.

Given the unbalanced and small sample sizes, all children were collapsed into a group, and we employed the Wilcoxon Sign Rank test in R(3.5.0) (R Core Team, 2018) in all comparisons, as this non-parametric test makes no assumptions on the distribution of the data (Hollander & Wolfe, 1973). Since each analysis proposed above simultaneously examines the measures for children, adults, and the comparison between the two groups, we used a Bonferroni correction with a family-wise error rate of 0.05, so that for each individual test α is set at $0.05/3 = 0.0167$. To calculate the interactions between condition and group, we used two vectors of individual differences in f₀ peak in NP1 minus f₀ peak in NP2 (one for children and one for adults). We then used the Mann Whitney test to calculate

Table 2. Descriptive summary of repetition accuracy in children: Mean number of correctly repeated nouns per trial (max=3), number of omitted nouns, and number of nouns containing segmental errors (n=27)

Condition	Average correct nouns per trial	Omitted nouns	Nouns with Segmental errors
Non-Recursive	2.68	10	26
Recursive	2.61	16	28

significance of difference on these two vectors. This vector of differences in f_0 represents the effect of condition (either metrical boost as in our fourth analysis, or recursive syntax as in our final analysis) on the f_0 contour. If the effect size is comparable in child and adult groups, we infer an absence of interaction. Finally, we measured the various correlations between prosodic measures, children's age in months, and success in the elicitation task using a non-parametric method (Kendall's tau).

Results

Sentence repetition task: repetition success

Sentences were first analyzed by repetition success, and then included or not in the acoustic analyses, according to the criteria mentioned above. Segmental errors were twice as frequent as NP omissions across conditions, as shown in Table 2. The recursive condition yielded slightly fewer targets than the non-recursive condition.

We observed a clear effect of age. Older children were more likely to produce successful NP repetitions; the correlation between age and average of successful repetition was significant (Tau = .426, $p = .004$).

Realization of pitch accents: f_0 peak alignment and pitch range

The target-like realization of pitch accents is a prerequisite for the realization of downstep patterns. Data from speakers who cannot realize the expected patterns of accented and unaccented words cannot be used to study downstep. Thus, we compared the realization of the pitch accents of accented and unaccented words (NP1s and NP2s). If children have mastered the accentual patterns, we expect to observe a larger pitch range in accented words than in unaccented words. Recall that previous research indicated one-year-old infants possessed adult-like f_0 fall in disyllabic words. Here we consider the same property in more complex structures.

Measurements in Table 3 included the linking particle *no*. Adults have overall slightly greater pitch ranges for NP1 than children. According to the Wilcoxon test of differences, these adult-child differences in pitch range are not statistically significant for unaccented words (Adult $Mdn = 1.41$, Child $Mdn = 1.16$; $W = 281$, $p = 0.11$), but are significant for accented words (Adult $Mdn = 5.74$, Child $Mdn = 4.52$; $W = 453$, $p = 0.021$).

Central to our pursuit, the contrast between initially accented and unaccented words was categorical for both syntactic conditions. All participants had a larger pitch range for accented syllables than for unaccented syllables, and the smallest individual difference was 1 ST. The average magnitude of individual contrast in pitch range

Table 3. Mean pitch fall in NP1 (in ST) for accented (AAA) and unaccented (UAU) nouns for children and adults for each type of syntactic structure.

		AAA		UAU	
		Mean	(SD)	Mean	(SD)
Adults (n = 10)	Non-Recursive	6.04	(3.04)	1.53	(0.69)
	Recursive	6.50	(2.87)	1.77	(0.91)
Children (n = 19)	Non-Recursive	4.38	(1.39)	1.24	(0.51)
	Recursive	5.28	(2.04)	1.68	(1.47)

across lexical accent types was 4.6 ST for the adults and 3.4 ST for the children. This suggests that children, despite their narrower pitch range, show the contrast between accented and unaccented words in their production.

Downstep and the effect of accent type

We begin this subsection by analyzing the patterns of f_0 fall in AAA recursive sequences. If there is syntagmatic downstep, we expect to see a sharp fall from the f_0 peak in NP1 to the peak in NP2 (see Table 1). Thus, we take the differences between the f_0 peaks in the first two noun phrases in the accented recursive condition (AAA). Results showed (Figure 6) that, for adults, the median difference for recursive AAA NPs (NP1-NP2 differences) is 2.56 semitones. A Wilcoxon test (one-tailed) revealed that these values are significantly greater than zero ($V = 55$, $p = .002$). Similarly, for children the median difference is also significantly greater than zero ($Mdn = 2.16$ semitones, $V = 184$, $p < .001$). A Mann-Whitney test showed no significant differences between adults and children in the overall magnitude of pitch fall in recursive AAA phrases ($W = 110$, $p = .512$).

Although group differences are not significant, several speakers (2/10 adult participants and 5/15 children) did not have a difference larger than 1 ST between the f_0 peaks of the first two NPs in the recursive structure. The absence of this type of downstep could be attributed to different production strategies. One observed pattern was either a pause or creaky voice between NP1 and NP2, which leads to pitch reset in NP2. In other cases, we observed the presence of phrase-final tones (often rising) at the end of NP1. These could be analyzed as the insertion of a phonological boundary that blocks downstep, with the pitch range being reset for the NP2. One adult speaker appeared to focalize the NP2, which would raise the peak f_0 of that constituent (Ishihara, 2016).

As explained above, downstep is not expected to occur when the first word is unaccented. Thus, to evaluate the effect of the initial accent, we compare the magnitude of pitch fall in the recursive AAA condition to the magnitude of pitch fall in the recursive UAU phrases (Figure 6). Both children and adults showed significant effects of accent type, with a larger pitch fall in initially accented phrases than in their initially unaccented counterparts (Adult accent contrast $Mdn = 3.82$ ST, $V = 55$, $p < .001$; children's accent contrast $Mdn = 2.03$ ST, $V = 118$, $p < .001$). The effect of accent is thus apparent in both groups. While adults show somewhat larger contrast and less dispersion, a Mann-Whitney test comparing the two vectors of differences reveals that group differences ($W = 115$, $p = .026$) are not significant, given that α was set at 0.0167.

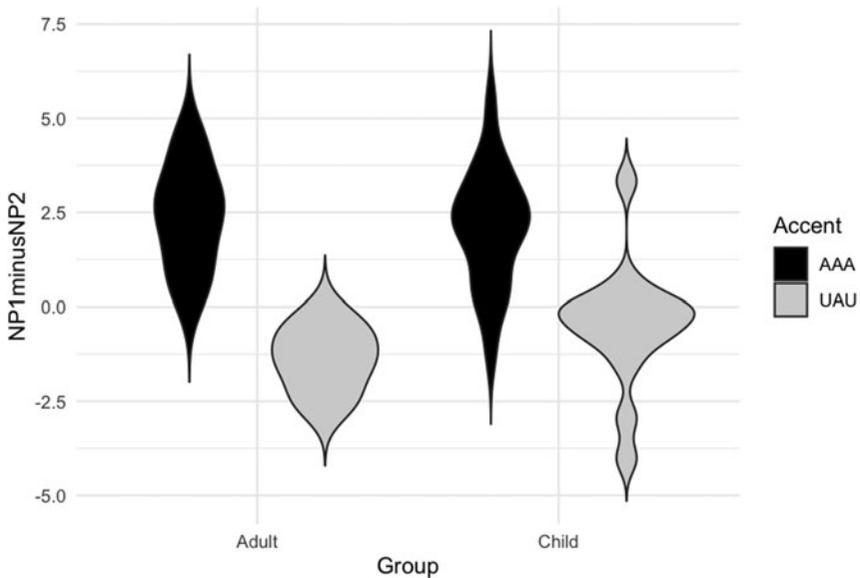


Figure 6. Violin plot of f_0 differences (NP1-NP2) in the AAA vs. UAU recursive conditions, by group (Adult, $n = 10$; Children, $n = 15$)

Recursive vs. non-recursive phrases (metrical boost)

As discussed above, the research literature on the prosody of Japanese (Kubozono, 1989) suggested that adults make a difference between recursive and non-recursive NPs, given a sequence of accented phrases. Thus, a larger f_0 fall is expected in the former when compared to the latter. Our results confirmed this hypothesis. Adults in our study showed a slight but reliable effect of metrical boost ($Mdn = .704$ ST, $V = 52$, $p = .004$). Children had even smaller and non-significant differences between the two structures ($Mdn = .145$ ST, $V = 88$, $p = .06$). However, the Mann-Whitney test of the vector of differences (which aims to test the interaction between group and condition) shows non-significant differences ($W = 88$, $p = 0.495$). As shown in Figure 7, children's realizations of recursive and non-recursive phrases are not distinctive.

Given the absence of significant group differences, we conducted an individual analysis (see Figure 8) to provide a clearer answer to the question of whether individual children differed from adults in their sensitivity to syntactic structure. Each point in Figure 8 represents the individual participant's magnitude of metrical boost, and age group is represented by color. For example, the adult speaker who had the highest metrical boost score is shown in the right-hand top corner of the plot. For that participant, the difference between the recursive and non-recursive magnitude of f_0 fall between NP1 and NP2 was approximately 4 semitones. All participants are ranked by their metrical boost score. A quick inspection of Figure 8 reveals that while adults (represented by black dots) show a trend to appear above children, it is also clear that individual speakers are equally spread above and below the line of perceptibility of 1 ST. The same is true for both younger and older children. Our data shows no development over the age span studied. A Kendall correlation between children's age and magnitude of their metrical boost score was not significant ($\text{Tau} = -.181$, $p = 0.362$).

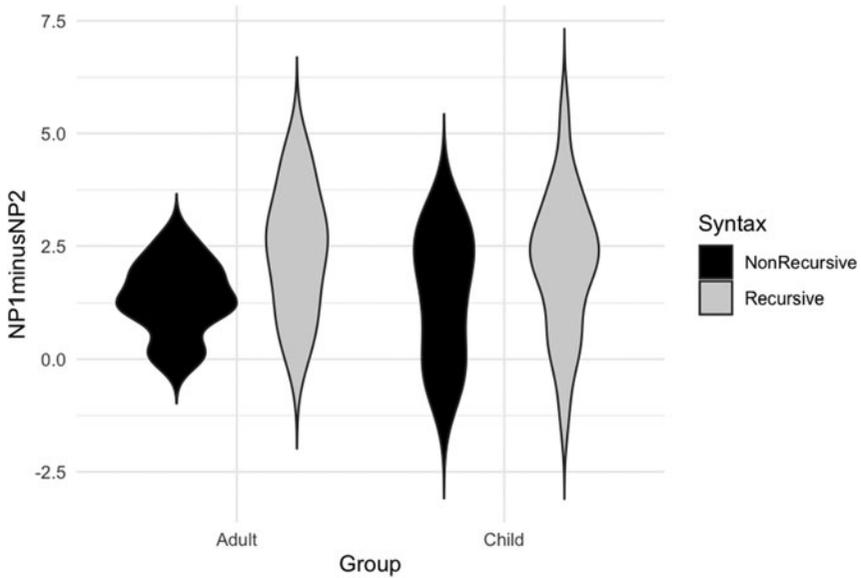


Figure 7. Violin plot comparing the magnitude of f_0 differences (NP1-NP2) for the recursive and non-recursive AAA conditions by group (Adult: $n = 10$; Children: $n = 15$)

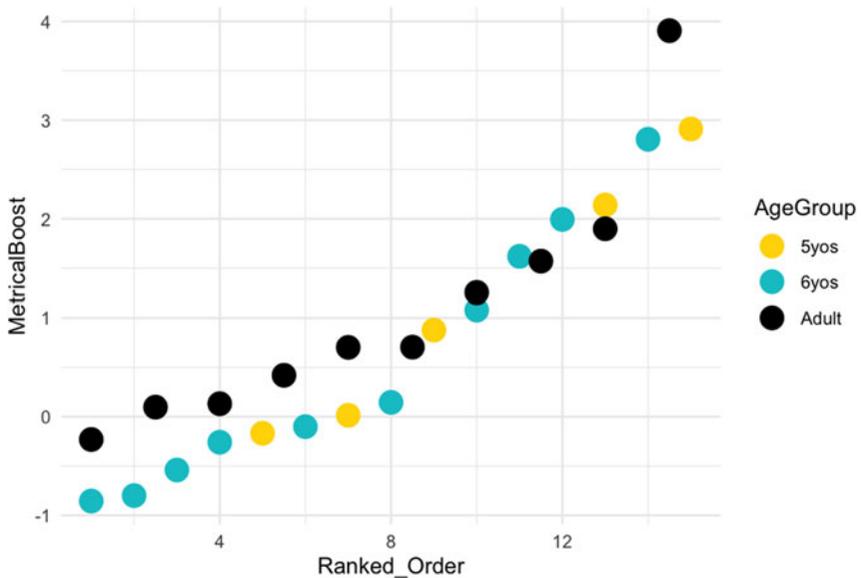


Figure 8. Individual analysis of metrical boost. Participants are ranked by magnitude of metrical boost. Shading is used to indicate age groups (five-year-olds, six-year-olds, and adults).

Relationship between prosodic and productive syntactic abilities

Previous sections reported the results of the Sentence Repetition task, where constituent structure is fixed by meaning, as determined by context and utterance content. These

Table 4. Mean proportion of target recursive possessor per age group

Group	Mean target recursion	(SD)
5yos (n = 14)	0.328	0.258
6yos (n = 13)	0.589	0.111
Adult (n = 10)	0.914	0.120

results allowed us to determine if children resembled adults in their pitch patterns in recursive and non-recursive structures. However, these results do not tell us much about the participants' actual capacity to produce the target syntactic structures spontaneously. If prosodic ability is needed to scaffold syntactic productive ability, we expect a positive correlation. Table 4 shows that production of the target recursive possessor in the elicited production task lies within the expected range for children this age, given previous results in English and French (Pérez-Leroux *et al.*, 2012; 2018; Roberge *et al.*, 2018). These studies suggest that most have started producing recursive possessives by age five, but their success is still limited. Children, and to a lesser degree adults, sometimes produce structurally simpler but pragmatically appropriate responses to these referential questions (i.e., "The one in the middle", etc.). The present observations are compatible with previous work. Three of the Japanese children failed to produce a single recursive possessive. Six-year-olds were significantly better than five-year-olds, but their performance lagged behind that of adults, which was close to ceiling.

To test for developmental correlations between syntactic and prosodic abilities, we entered into the analysis individual children's proportions of target production of recursive possessives and their metrical boost scores. The analysis shows no association between the two abilities (Kendall's tau = -0.228, $p = .26$). Figure 9 plots children's proportion of target responses in the production task as a function of metrical boost, with our results grouped by age (five- and six-year-olds). Six-year-olds have relatively high production of recursive NPs, but metrical boost scores vary greatly.

Figure 9 suggests that the realization of the metrical boost is not a prerequisite to the ability to produce recursive possessives. Most children are able to produce some recursive possessives, and those who do are roughly divided between those who could produce intonational contrasts above one semitone, and those who could not. In other words, children whose realizations are across both sides of the perceptibility boundary represent the range of productive syntactic abilities.

Discussion

Summary of observations

To summarize, our results showed that children have acquired aspects of the realization of lexical accent in Japanese, exhibiting clear and reliable distinctions between accented and unaccented words in terms of pitch range. This is consistent with previous observations (e.g., Ota, 2016).

Since the Japanese children in this study had mastered the contrast between accented and unaccented words, we were able to test for the realization of downstep patterns. To answer our first research question concerning syntagmatic downstep, we can conclude that children have acquired the downward pitch contour in AAA sequences. Our results

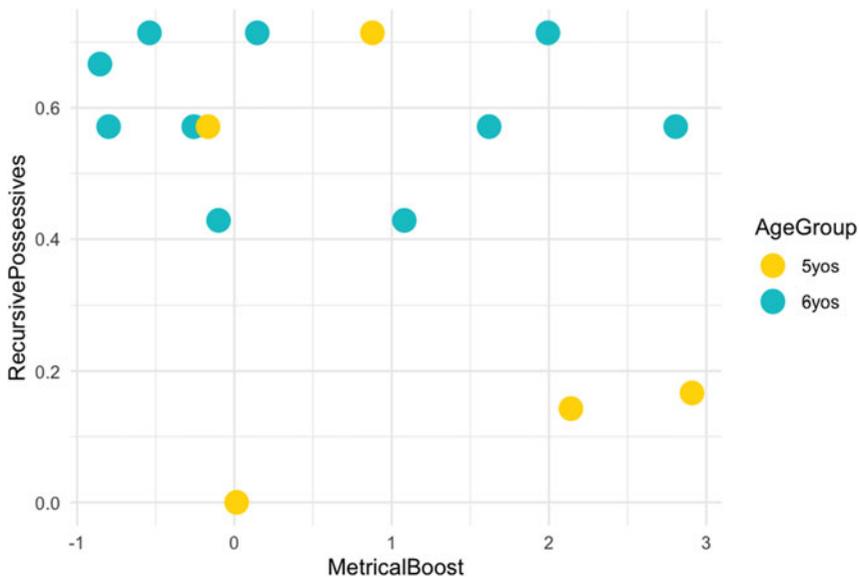


Figure 9. Scatterplot of individual children's mean target production of recursive possessives in the elicitation task as a function of metrical boost. Shade in the points in the scatterplot represents age groups (five- and six-year-olds). ($n = 15$)

indicated that both adults and children have a significant pitch fall from the peak in NP1 to the peak in NP2. Moreover, there were no significant differences in the magnitude of pitch fall for AAA recursive NPs between the two groups. We also observed that both adults and children showed a clear difference in pitch fall for the two accent type conditions, i.e., AAA vs. UAU. Given the same syntactic configuration (recursive NPs), both groups show significantly higher pitch fall from the first NP to the second in the accented condition than in the unaccented condition. Although adults showed a more robust effect of accent type, the difference between groups did not reach significance. Thus, we infer that children have acquired the downstep pattern associated with two sequentially accented NPs.

Regarding our second research question – namely, whether Japanese children are sensitive to the effect of syntax in prosody – our results revealed that, while adults have significant effects of metrical boost (i.e., larger pitch fall for the recursive condition than for the non-recursive condition), this was not the case for children, who demonstrated a great degree of variability for the non-recursive condition. As a result, the overall contrast between the two groups failed to reach significance. Our analysis of individual patterns allows us to make a further observation: participants across age groups were evenly distributed between those who produced differences above the perceptibility range and those who did not (Figure 8). In other words, both child and adult populations included individuals who produced intonational contrasts to reflect syntactic contrast and individuals who did not.

Finally, we explored whether prosodic ability was associated with syntactic productive ability. Our production data attests to substantive growth in the period of study. Five-year-olds produced target recursive possessives in over 30% of the trials, whereas six-year-olds succeed at almost double those rates. These figures are fully

comparable to data in Spanish and English; and similar to the five-year-olds' data in Roberge's *et al.* (2018) French study, which reports no increase for recursive possessives by age six. However, we found no association between production of recursive possessives and realization of a prosodic contrast between recursive and non-recursive configurations. It is worth noting that, from the subset of children who provided a metrical boost score, only one had no recursive ability, and two more had low ability. So, it is possible that our lack of correlation stems from having relatively high production scores, and that a study of younger children would have yielded a different picture. Within the bounds of our evidence, we must conclude that children's ability to realize metrical boost appears unrelated to their ability to produce recursive possessives in the elicitation task.

Acquisition of prosody and syntax

Our results indicate that five- and six-year-old Japanese children have acquired the prosodic patterns in recursive accented and unaccented NPs. This contrast is expressed by the difference between a sharp and staircase-like f_0 pattern and a less pronounced falling contour or even slight rise in UAU sentences. Although the downstep pattern is a specific characteristic of Japanese, previous literature has established that falling contours are acquired earlier than other types of contours (e.g., Balog & Snow, 2007; Prieto *et al.*, 2012; Filipe *et al.*, 2017). Given that Japanese children seem to have acquired these contours before having acquired the use of pitch to signal syntactic structure, the patterns we have observed resemble those reported in the literature.

At the age range examined in this study, Japanese children do not consistently use pitch to mark syntactic structure. Previous studies focusing on adult speakers have claimed that recursive sequences of accented words differ from non-recursive sequences in the scaling of the f_0 peaks. Whereas a sharp descent is expected in the former, a lesser degree of descent is expected in the latter – this is the metrical boost (Kubozono, 1989; Ishihara, 2016). Our results have shown that children did not make a pitch distinction between the two structures, showing a great degree of variability, particularly in non-recursive structures. However, in this they were not that different from adults, who also showed substantive variability. Indeed, when analyzing the individual results, we observed that 6/10 adults were well below the 1 ST threshold. Similar variability in adult speech in recursive structures in Japanese was observed by Ishihara (2016). Indeed, this is also true of English: Wells *et al.* (2004) found variability in perception and in production in the grouping of noun sequences.

That children lag behind in the acquisition of prosodic cues to mark syntactic structures has also been observed for a variety of languages, such as English (Vogel & Raimy, 2002), Portuguese (Filipe *et al.*, 2017) and Spanish (Martínez-Castilla & Peppé, 2008), and a variety of structures (i.e., compounds, adjectival scope, syntactic grouping). Our data suggest that children are actually mirroring the variability found in the adult group. Here there are two alternative explanations to account for the differences between our findings and the previous findings regarding metrical boost in Japanese. First, we can hypothesize that the task we used favoured variability in both populations since participants were given the whole sentence and had to repeat it with a normal intonation, as opposed to previous studies where they were asked to read. If task difficulty were at the source

of this variability, we should expect similar comments in previous works using this task. Second, it may be the case that adult data are intrinsically variable, and that some previous studies may have failed to capture such variability. Some degree of variability was reported in Ishihara (2016), albeit regarding a different structure. In Kubozono (1989), a single speaker was used, offering no opportunity to examine inter-speaker variability. Variability in prosodic marking of syntactic structure, however, has been reported in a variety of studies on other languages; although further research is certainly needed for the adult population – indeed, adults variably use prosodic cues to mark ambiguous sentences, and might only reliably do so when made aware of the ambiguity as participants in highly controlled experiments (e.g., Snedeker & Casserley, 2010). In addition to ‘awareness’, other recent studies are beginning to uncover some of the subtle sources of this variability, which may be attributed to contextual factors (e.g., selection of lexical items) that trigger different individual interpretations (e.g., Klassen & Wagner, 2017) in the selection of anaphoric antecedents and the consequent prominence shifts.

Conclusion

This article has contributed to the previous literature on the acquisition of prosody and syntax by presenting new data on the acquisition of an understudied phenomenon in an understudied population. We have shown that, by ages five and six, Japanese children have acquired lexical prosody, in particular the contrast between accented and unaccented words, which is a prerequisite for the prosodic marking of phrasing (downstep) and recursion (metrical boost). We have also established that they have clear mastery of the Japanese pattern of downstep with accented phrases.

We have concluded that Japanese children did not consistently produce the expected metrical boost in accented non-recursive sequences. However, our data showed that adults are equally variable. Thus, although previous literature claimed that Japanese consistently marked recursion prosodically, our results showed that our control group did not consistently do so. Albeit new for Japanese, this result is consistent with previous literature from other languages suggesting that prosodic encoding of syntactic structure is not a reliable cue in sentence processing. Furthermore, if we compare our results on the acquisition of downstep and metrical boost, we can conclude that children have acquired the less variable patterns of prosodic downstep and that they match the variability of adults in marking or not marking recursion. Finally, as to the question of whether Japanese children who have acquired prosodic marking of recursion are better at producing recursive complex NPs, the answer seems to be negative. In our data, prosody does not seem to contribute to children’s emerging ability to produce complex, recursively embedded NPs.

Compared to previous research, our study has the advantage of using two semi-spontaneous tasks that independently targeted prosodic and syntactic development. Moreover, to explore prosodic development, we carefully controlled for type of word and syllable type. The usual disclaimers apply: our sample size is relatively small, and the number of tokens per condition limited. However, the current set of results offers a first step towards studying the development of the prosody-syntax interface in a pitch accent language, and thus studying acquisition of the unique patterns of downstep and metrical boost that characterize Japanese.

Acknowledgments. This study was conducted with the support of the Social Sciences and Humanities Research Council of Canada to A. T. Pérez-Leroux and Yves Roberge (SSHRC 435-2014-2000 Development of NP complexity in children). We also thank two anonymous reviewers and audiences of Intonation Workshop and International Child Phonology Conference for their comments. All errors are ours.

References

- Archer, S. L., & Curtin, S. (2016). Nine-month-olds use frequency of onset clusters to segment novel words. *Journal of Experimental Child Psychology*, 148, 131–141.
- Arnhold, A., Chen, A., & Järvikivi, J. (2016). Acquiring complex focus-marking: Finnish four- to five-year-olds use prosody and word order in interaction. *Frontiers in Psychology*, 7, 1886.
- Astruc, L., Payne, E., Post, B., Vanrell, M. M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech*, 56(2), 229–253.
- Athanasopoulou, A., & Vogel, I. (2016). Is the input for prosodic bootstrapping of word order reliable? The case of phrasal prominence in Turkish and French. *Speech Prosody*, 8, 687–691.
- Balog, H. L., & Snow, D. (2007). The adaptation and application of relational and independent analyses for intonation production in young children. *Journal of Phonetics*, 35, 118–133.
- Beach, C. M., Katz, W. F., & Skowronski, A. (1996). Children's processing of prosodic cues for phrasal interpretation. *Journal of Acoustical Society of America*, 99, 1148–1160.
- Beckman, M., & Pierrehumbert, J. B. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. <http://www.praat.org/>
- de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, 19(2), 235–250.
- Chen, A., & Höhle, B. (2018). Four- to five-year olds' use of word order and prosody in focus marking in Dutch. *Linguistics Vanguard*, 4(1), 1–9.
- Ching, T. Y. (1984). Lexical tone pattern learning in Cantonese children. *Language Learning and Communication/Zhongying Yuwen Jiaoxue*, 3(3), 317–334.
- Choi, Y., & Mazuka, R. (2003). Young children's use of prosody in sentence parsing. *Journal of Psycholinguistic Research*, 32(2), 197–217.
- Christophe, A., & Dupoux, E. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. *The Linguistic Review*, 13(3–4), 383–412.
- Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, 2(3), 385–394.
- Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, 51(1&2), 61–75.
- Clancy, P. (1986). The acquisition of Japanese. In D. I. Slobin (ed.), *The crosslinguistic study of language acquisition*, vol. 1 (pp. 373–524). Hillsdale, NJ, Lawrence Erlbaum Associates.
- Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, 12, 643–661.
- Curtin, S. (2009). Twelve-month-olds learn novel word–object pairings differing only in stress pattern. *Journal of Child Language*, 36, 1157–1165.
- Féry, C. (2010). Recursion in prosodic structure. *Phonological Studies*, 13, 51–60.
- Féry, C. (2016). *Intonation and prosodic structure*. Cambridge: Cambridge University Press.
- Filipe, M., Peppé, S., Frota, S., & Vicente, S. G. (2017). Prosodic development in European Portuguese from childhood to adulthood. *Applied Psycholinguistics*, 38, 1045–1070.
- Frota, S., Butler, J., & Vigário, M. (2013). Infants' perception of intonation: Is it a statement or a question? *Infancy*, 19(2), 194–213.
- Fujimori, C. (2010). *Acquisition of recursive possessives in Japanese*. Unpublished manuscript, University of Massachusetts, Amherst.
- Giblin, I., Zhou, P., Bill, C., Shi, J., & Crain, S. (2019). The Spontaneous eMERGEncE of Recursion in Child Language. In M. Brown & B. Dailey (Eds.), *Proceedings of the 43rd annual Boston University Conference on Language Development* (pp. 270–285). Somerville: Cascadilla Press.
- †Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental phonetic approach to speech melody*. Cambridge: Cambridge University Press.

- van Heugten, M., Dautriche, I., & Christophe, A. (2014). Phonological and prosodic bootstrapping. In P. Brooks, V. Kempe, & J. G. Golson (Eds.), *Encyclopedia of Language Development* (pp. 447–451). Thousand Oaks, CA: SAGE Publications Inc.
- Hochberg, J. G. (1988). First steps in the acquisition of Spanish stress. *Journal of Child Language*, 15, 273–292.
- Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods*. New York: John Wiley & Sons.
- Ishihara, S. (2016). Japanese downstep revisited. *Natural Language and Linguistic Theory*, 34, 1389–1443.
- Ito, J., & Mester, A. (2013). Prosodic subcategories in Japanese. *Lingua*, 124, 20–40.
- Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.
- Katz, W., Beach, C., Jenouri, K., & Verma, S. (1996). Duration and fundamental frequency correlates of phrase boundaries in productions by children and adults. *Journal of the Acoustical Society of America*, 99 (5), 3179–3191.
- Kehoe, M. (1997). Stress error patterns in English-speaking children's word productions. *Clinical Linguistics & Phonetics*, 11, 389–409.
- Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38(2), 338–350.
- Klassen, J., & Wagner, M. (2017). Prosodic prominence shifts are anaphoric. *Journal of Memory and Language*, 92, 305–326.
- Kubozono, H. (1989). Syntactic and rhythmic effects on downstep in Japanese. *Phonology*, 6, 39–67.
- Limbach, M., & Adone, D. (2011). Language acquisition of recursive possessives in English. In K. Franich, K. M. Iserman, & L. L. Keil (Eds.), *BUCLD 34: Proceedings of the 34th annual Boston University Conference on Language Development 34* (pp. 281–90).
- Leó, C., & Arias, J. (2006). Foot, word and phrase constraints in first language acquisition of Spanish stress. In F. Martínez-Gil & S. Colina (Eds.), *Optimality-theoretic studies in Spanish phonology* (pp. 470–496). Amsterdam: John Benjamins.
- Martínez-Castilla, P., & Peppé, S. (2008). Intonation features of the expression of emotions in Spanish: Preliminary study for a prosody assessment procedure. *Clinical Linguistics & Phonetics*, 22(4–5), 363–70.
- McDaniel, D., McKee, C., & Garret, M. F. (2010). Children's sentence planning: Syntactic correlates of fluency variations. *Journal of Child Language*, 37, 59–94.
- Murasugi, K., Nakatani, T., & Fuji, C. (2012). Three types of the “overgenerated no” in the acquisition of Japanese noun phrases. *Nanzan Linguistics*, 8, 69–87.
- Ota, M. (2003). The development of lexical pitch accent systems: An autosegmental analysis. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 48(3/4), 357–383.
- Ota, M. (2006). Phonological acquisition. In M. Nakayama, R. Mazuka, & Y. Shirai (Eds.) with P. Li (general editor), *The handbook of East Asian psycholinguistics, Volume II: Japanese* (pp. 41–47). Cambridge: Cambridge University Press.
- Ota, M. (2016). Prosodic phenomena: Stress, tone, and intonation. In J. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford handbook of developmental linguistics* (pp. 68–86). Oxford: Oxford University Press.
- Pérez-Leroux, A. T. (in press). A child's view of Romance modification. In G. Alboiu & R. King (Eds.), *Points of Convergence in Romance Linguistics 2018: Papers selected from the 48th Linguistic Symposium on Romance Languages*. Amsterdam/Philadelphia: John Benjamins.
- Pérez-Leroux, A. T., Castilla-Earls, A. P., Bejar, S., & Massam, D. (2012). Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition*, 19(4), 301–311.
- Pérez-Leroux, A. T., Peterson, T., Bejar, S., Castilla-Earls, A. P., Massam, D., & Roberge, Y. (2018). The acquisition of recursive modification in NPs. *Language*, 94(2), 332–359.
- Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese* (unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.
- Prieto, P., Estrella, A., Thorson, J., & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical and lexical development? Evidence from emerging intonation in Catalan and Spanish. *Journal of Child Language*, 39, 221–257.
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <<https://www.R-project.org/>>.

- Roberge, Y., Pérez-Leroux, A. T., & Frolova, A.** (2018). On recursive modification in L1 French. *Languages*, 3, 6. doi:10.3390/languages3010006
- Roeper, T.** (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics* 5, 57–86.
- Saindon, M. R., Cirelli, L. K., Schellenberg, E. G., van Lieshout, P., & Trehub, S. E.** (2017). Children's and adults' perception of questions and statements from terminal fundamental frequency contours. *Journal of the Acoustical Society of America*, 141(5), 3123–3131.
- Sauermann, A., Höhle, B., Chen, A., & Järvikivi, J.** (2011). Intonational marking of focus in different word orders in German children. In M. B. Washburn, K. McKinney-Bock, E. Varis, A. Sawyer, & B. Tomaszewicz (Eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics* (pp. 313–322). Somerville, MA: Cascadilla Press.
- Selkirk, E., & Tateishi, K.** (1991). Syntax and downstep in Japanese. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda* (pp. 519–543). Dordrecht: Kluwer Academic Publishers.
- Shirose, A.** (2001). Fukugo meishi akusento kisoku no kakutoku katei (Acquisition of compound accent rule in Japanese). *Onsei Kenkyu (Journal of the Phonetic Society of Japan)*, 5(2), 39–53.
- Shirose, A., Kakehi, K., & Kiritani, S.** (2001). Interference of accent system variation with the acquisition of compound accent rule: A cross-dialectal study. *Journal of Cognitive Science*, 2, 33–44.
- Snedeker, J., & Casserly, E.** (2010). "Is it all relative?" Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, 25(7–9), 1234–1264.
- Snedeker, J., & Trueswell, J.** (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48, 103–130.
- Snedeker, J., & Yuan, S.** (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58(2), 574–608.
- Snow, D.** (2007). Polysyllabic units in the vocalizations of children from 0;6 to 1;11: Intonation-groups, tones and rhythms. *Journal of Child Language*, 34(4), 765–797.
- Snow, D., & Balog, H.** (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025–1058.
- Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J.** (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4), 869–902. doi:10.1017/S0305000908008763
- Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W.** (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language* 49, 249–67
- Stevens, K. N.** (1998/2000). *Acoustic phonetics*. Boston: The MIT press.
- Terunuma, A., Isobe, M., Nakajima, M., Okabe, R., Inaba, S., Inokuma, S., & Nakato, T.** (2017). Acquisition of recursive possessives and locatives within DPs in Japanese. In M. LaMendola & J. Scott (Eds.), *Proceedings of the 41st Annual Boston University Conference on Language Development* (pp. 626–636). Somerville, MA: Cascadilla Press.
- Terunuma, A., & Nakato, T.** (2018). Recursive possessives in child Japanese. In L. Amaral, M. Maia, A. Nevins, & T. Roeper (Eds.), *Recursion across domains* (pp. 187–210). Cambridge: Cambridge University Press. doi:10.1017/9781108290708.012
- Vihman, M. M.** (1996). *Phonological development: The origins of language in the child*. Oxford: Blackwell.
- Vogel, I., & Raimy, E.** (2002). The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language*, 29(2), 225–50.
- Wells, B., Peppé, S., & Goulondris, N.** (2004). Intonation development from five to thirteen. *Journal of Child Language*, 31, 749–778.
- Wong, P.** (2009). *Development of lexical tone production in disyllabic words by 2- to 6-year-old Mandarin-speaking children* (Doctoral dissertation). Retrieved from Dissertation Abstracts International, B: Sciences and Engineering (69(9)).
- Wong, P., & Leung, C. T.-T.** (2018). Suprasegmental features are not acquired early: Perception and production of monosyllabic Cantonese lexical tones in 4- to 6-year-old preschool children. *Journal of Speech, Language, and Hearing Research*, 61, 1070–1085. doi:10.1044/2018_JSLHR-S-17-0288
- Yang, A., & Chen, A.** (2018). The developmental path to adult-like prosodic focus-marking in Mandarin Chinese-speaking children. *First Language*, 38(1), 26–46.

Appendix

Test phrases in sentence repetition task

1. NP recursion: NP1-no NP2-no NP3				
a. [A] [A] [A]				
Recursive:	erumo-no	oneechan-no	doresu	
	Elmo-NO	sister-NO	dress	
	'Elmo's sister's dress'			
Non-recursive:	erumo-no	orenji-no	doresu	
	Elmo-NO	orange(N)-NO	dress	
	'Elmo's orange dress'			
b. [A] [U] [A]				
Recursive:	erumo-no	otomodachi-no	zubon	
	Elmo-NO	friend-NO	trousers	
	'Elmo's friend's trousers'			
Non-recursive:	erumo-no	midori-iro-no	zubon	
	Elmo-NO	green(N)-NO	trousers	
	'Elmo's green trousers'			
c. [U] [U] [U]				
Recursive:	Suneo-no	otomodachi-no	booru	
	Suneo-NO	friend-NO	ball	
	'Suneo's friend's ball'			
Non-recursive:	Suneo-no	midori-iro-no	booru	
	Suneo-NO	green(N)-NO	ball	
	'Suneo's green ball'			
d. [U] [A] [U]				
Recursive:	Suneo-no	otoosan-no	booru	
	Suneo-NO	father-NO	ball	
	'Suneo's father's ball'			
Non-recursive:	Suneo-no	orenji-no	booru	
	Suneo-NO	orange(N)-NO	Ball	
	'Suneo's orange ball'			

2. VP high attachment vs. low attachment				
a.	watashi-wa	megane-de	panda-o	mi-ta
	I-NOM	glasses-with	panda-ACC	see-PAST
	'I saw the panda with glasses'			

b.	watashi-wa	megane-no	panda-o	mi-ta	
	I-NOM	glasses-NO	panda-ACC	see-PAST	
	'I saw a panda that was wearing glasses.'				
c.	watashi-wa	sofaa-no	tonari-ni	ranpu-o	oi-ta
	I-NOM	sofa-NO	next-NI	lamp-ACC	put-PAST
	'I left the lamp next to the sofa.'				
d.	watashi-wa	sofaa-no	tonari-no	ranpu-o	sawat-ta.
	I-NOM	sofa-NO	next-NO	lamp-ACC	touch-PAST
	'I touched the lamp next to the sofa.'				
e.	watashi-wa	nagano-de	ringo-o	hakon-da.	
	I-NOM	Nagano-DE	apple-ACC	carry-PAST	
	'I carried apples in Nagano.'				
f.	watashi-wa	nagano-no	ringo-o	hakon-da.	
	I-NOM	Nagano-NO	apple-ACC	carry-PAST	
	'I carried apples that came from Nagano.'				
g.	watashi-wa	booshi-de	otoko-o	tatai-ta.	
	I-NOM	hat-DE	man-ACC	hit-PAST	
	'I hit the man with a hat (instrumental).'				
h.	watashi-wa	booshi-no	otoko-o	tatai-ta.	
	I-NOM	hat-NO	man-ACC	hit-PAST	
	'I hit the man who was wearing a hat.'				

Cite this article: Hirayama M, Colantoni L, Pérez-Leroux AT (2021). Can prosody encode recursive embedding? Children's realizations of complex NPs in Japanese. *Journal of Child Language* 48, 429–453. <https://doi.org/10.1017/S0305000920000367>