Exploiting Pitch Accent Information in Compound Processing: A Comparison between Adults and 6- to 7-Year-Old Children

Yuki Hirose & Reiko Mazuka

To cite this article: Yuki Hirose & Reiko Mazuka (2017): Exploiting Pitch Accent Information in Compound Processing: A Comparison between Adults and 6- to 7-Year-Old Children, Language Learning and Development, DOI: 10.1080/15475441.2017.1292141

To link to this article: http://dx.doi.org/10.1080/15475441.2017.1292141

Published online: 10 Mar 2017.

Submit your article to this journal

View related articles

View Crossmark data
Exploiting Pitch Accent Information in Compound Processing: A Comparison between Adults and 6- to 7-Year-Old Children

Yuki Hirose\textsuperscript{a} and Reiko Mazuka\textsuperscript{b}

\textsuperscript{a}Graduate School of Arts and Sciences, The University of Tokyo; \textsuperscript{b}Laboratory for Language Development, RIKEN Brain Science Institute

\textbf{ABSTRACT}

A noun can be potentially ambiguous as to whether it is a head on its own, or is a modifier of a Noun + Noun compound waiting for its head. This study investigates whether young children can exploit the prosodic information on a modifier constituent preceding the head to facilitate resolution of such ambiguity in Japanese. Evidence from English suggests that young speakers are not sensitive to compound stress in distinguishing between compounds and syntactic phrases unless the compound is very familiar (Good, 2008; Vogel & Rainey, 2002). This study concerns whether children in general have such limited capability to use prosodic cues to promptly compute a compound representation without the lexical boost, or whether they might show greater sensitivity to more categorical compound prosody such as that associated with the Compound Accent Rule (CAR) in Japanese. A previous study (Hirose & Mazuka, 2015) demonstrated that adult Japanese speakers can predict the compound structure prior to the head if the prosodic information on the modifier unambiguously signals that the CAR is being applied. The present study conducted the same online experiment with children (6- to 7-year-olds) and compared the time course of the effects with that of adults using permutation-based analysis (Maris & Oostenveld, 2007). The results reveal that children are sensitive to pitch accent information that facilitates the quicker processing of the compound or the single head noun representation compared to when such prosodic signals are less apparent, depending on the type of the lexical accent of the noun in question.

\textbf{Introduction}

\textit{Incremental processing and prosodic disambiguation in children}

In language processing at the sentence level, the syntactic structure of upcoming material can be predicted from the prosodic information that accompanies an earlier part of the input (Kjelgaard & Speer, 1999; Nakamura, Arai, & Mazuka, 2012; Snedeker & Trueswell, 2004). Sentence processing studies investigating children have provided a body of evidence indicating that young children (6–7 years of age or younger) make an early commitment to a syntactic analysis based on the syntactic information available in the language. Their ability to use linguistic or non-linguistic information to override the default preference and their efficiency in reacting to subsequent (and more definitive) information, however, may differ from that of adults (PP attachment ambiguity in English: Trueswell, Sekerina, Hill, & Logrip, 1999; Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000; Snedeker & Trueswell, 2004; Kidd & Bavin, 2005; Kidd, Stewart, & Serratrice, 2011; RC attachment ambiguity in English: Felser, Marinis, & Clahsen, 2003; location/destination PP ambiguity in Korean: Choi & Trueswell, 2010; wh-dependency resolution in English and Japanese:
Omasi, White, Goro, Lidz, & Phillips, 2014; RC gap ambiguity in Turkish: Özge, Marinis, & Zeyrek, 2015; measure expressions; Syrett, 2010). Most of these studies share a basic understanding that children do make early decisions but oftentimes dismiss potentially useful cues such as referential information and/or lexical-semantic information. In addition, when their initial commitment is falsified, young children can have more difficulty reanalyzing the structure compared to adults. The present study aims to demonstrate that young children are indeed capable of using prosodic information to expedite the construction of a morpho-syntactic structure for the input they receive. Studies on young children’s ability to utilize prosodic information to resolve structural ambiguity have had rather mixed results. A relatively small number of studies have produced positive evidence. For example, de Carvalho, Dautriche, and Christophe (2016) demonstrated that children as young as 3; 6 can exploit prosodic boundaries for syntactic disambiguation in French. In English, Snedeker and Yuan (2008) reported that 4- to 6-year-olds can use information about the location of an iP boundary (a level of prosodic juncture typically marking major syntactic breaks, signaled by the presence of a pause and lengthening of phrase-final words) to resolve PP attachment ambiguity, with a delay of about 500 ms compared to adults. Snedeker and Yuan argued that the children’s performance was sensitive to the experimental design: children tested on the same constructions did not exhibit the same level of sensitivity to the prosodic break if they saw all the prosodic conditions in a within-subject design, while they showed a strong effect of lexical bias. Choi and Mazuka (2003) demonstrated that Korean-speaking children (3- to 5-year-olds) failed to resolve structural boundary ambiguity even with robust prosodic information, but were relatively successful in using comparable sets of acoustic cues (i.e., the size of pauses) to resolve word segmentation ambiguity.

The guidance effect of prosody is not limited to the selection of syntactic structure based on durational cues. Word accent (e.g., stress or pitch), focus accent, and phrasal/sentential intonation also play important roles in language comprehension. Ito, Jincho, Minai, Yamane, and Mazuka (2012) demonstrated that children are sensitive to pitch accent in contrast resolution in Japanese, although in other languages children’s sensitivity to contrastive pitch accent seems rather limited (English: Arnold, 2008; Russian: Sekerina & Trueswell, 2012). Studies looking at children’s ability to prosodically disambiguate the interpretation of focus particles (such as only in English), based on either stress or pitch accent, are not conclusive either (an example of negative evidence comes from Costa & Szendroán I, 2006; for Portuguese, while positive conclusions have been reported by Höhle, Berger, Müller, Schmitz, & Weissenborn, 2009, for German; and Zhou, Su, Crain, Gao, & Zhan, 2011; Zhou, Su, Gao, and Zhan, 2012, for Mandarin Chinese). The present study focuses on children’s real-time processing of compound prosody, which is realized by distinct accentual prominence in many languages.

**Compound processing in children**

Different languages have varying degrees of availability of prosodic cues indicating the compound status of a series of two lexical items (such as N + N or A + N). A study on Korean showed that children had difficulty distinguishing a Noun + Noun (N + N) compound from a series of Noun arguments (where case markers are dropped), presumably because the phonological marking of compounds in Korean is not always reliable or distinct (Jun, 1996). In English, compound stress (e.g., bláck board vs. black bóard) appears to provide a clear distinction between compounds and Adjective + Noun (A + N) syntactic phrases. English-speaking children can understand the basic semantic concept of unfamiliar compounds by the age of 3 or 4 years (Clark, Gelman, & Lane, 1985; Gottfried, 1997) and assign the appropriate compound stress pattern in production by approximately age 5 (Wells, Peppé, & Goulandris, 2004). In comprehension, children at age 3 start to develop sensitivity to compound prosody in identifying a compound representation (as opposed to a list of two nouns or an A + N phrase) for compounds that are familiar to them. In contrast, studies to date agree that children in general (young and old) are not sensitive to compound stress in comprehension (Good, 2008; Vogel & Raimy, 2002) unless the compound is already a well-known lexical entry
to them. In fact, Atkinson-King (1973) reported that children up to age 12 had difficulty distinguishing familiar expressions such as *hot dog* vs. *hót dog* (whether all the compound items were familiar to the child participants is unknown, however, as Vogel & Raimy, 2002, pointed out). The difficulty in perceiving compound stress may be partly because there is no way to tease compound stress apart from the focus stress marking new information status with which the modifier of a compound is typically associated (Vogel & Raimy, 2002). In fact, Gamache (2013) reported that even adults failed to associate compound stress with a compound representation when processing novel compounds.

Children’s relatively poor performance in identifying compound prosody when comprehending novel/unfamiliar compounds allows for at least three possible accounts. One possibility is that children in general are simply unable to use compound prosody productively to compute a compound structure without support from their still-developing lexical knowledge. Their ability to compute word-internal syntactic structure may also still be underdeveloped. Highly familiar compounds are understood via the lexicon, without the need for a computation process. A second explanation concerns language-specific characteristics of English compound prosody. After all, in English (and presumably many other languages), prosodic prominence on a first constituent (e.g., primary vs. secondary/tertiary stress in English) may be difficult to evaluate without comparison to the subsequent head (Kunter, 2011). Thus, stress change on the first constituent may be relatively uninformative and of little use in making a prompt decision about its morpho-syntactic status. A third possibility is that children do not actually acquire compound stress in the first place: their apparent success in producing compounds correctly, with the primary stress on the first constituent, is actually due to the fact that the first modifier constituent usually carries new information. That is, children are only marking information status by assigning stress prominence to the modifier (Vogel & Raimy, 2002). In the next section, we will show how Japanese can be useful to elaborate on this question and possibly eliminate the first and third alternatives, by helping to explain the apparent confusion about child compound processing/production in English.

**Predictive processing of the compound representation**

Compound prosody in Japanese is useful to investigate the role of information that could potentially enable the pre-head anticipation of a compound, because the prosodic information on the modifier alone can provide an absolute and dichotomous cue about its compound status (unlike English, German, and Korean as discussed above). In Japanese, a noun is either accented or unaccented in the lexicon, and if it is accented, the position of the accent is also lexically specified. For example, the lexically accented word *mi’kan*\(^1\) (tangerine) carries an accent nucleus (associated with a sharp pitch fall, HL) on the initial mora\(^2\) and exhibits the tonal pattern HLL. In contrast, a lexically unaccented word such as *ringo* (apple) shows the tonal pattern LHH (i.e., a word lacking a sharp HL pitch declination is considered unaccented). When two nouns form a nominal compound, the Compound Accent Rule (CAR) applies. The outcome of the CAR is conditioned by the number of morae and the type of lexical accent of the second constituent (C2). In the word combinations used in this study, the second constituents are two- or three-morae-long initially accented nouns. If the first constituent (C1) is lexically accented, then the CAR deletes its lexical accent. The new accent is re-assigned to, in this case, the initial mora of C2. For example, in (1a), C1 is *mi’kan* (tangerine) and C2 is *ju’usu* (juice), and when they become a compound their original accent is deleted and the compound accent is re-assigned to *ju’usu*. Effectively, the lexical accent of the first constituent changes if the noun was originally accented in isolation, thus providing an early cue that the noun in question is part of a compound. We therefore call this case “Unambiguous” with respect to whether the CAR has been applied or not. This change (HLL HLL → LHH HLL), however, does not occur when the first noun was originally unaccented in isolation (LHH), resulting in a temporary ambiguity as to

---

\(^1\)An apostrophe (’) indicates that the lexical accent falls on the following mora.

\(^2\)A mora (plural *morae*) is a phonological unit, usually comprising a vowel or a consonant-vowel combination, which is the basis of the sound/rhythm system of Japanese (in contrast to the syllable).
whether the noun is a head noun or the modifier of a noun-noun compound. Such a case, exemplified in (1b), is referred to as “Ambiguous” in this study.

In the experiment conducted by Hirose and Mazuka (2015), adult participants saw eight visual objects on a computer screen. The objects depicted were animals, fruits, or vegetables, or imaginary “combined creatures” that were variants of the animals with some of the attributes (e.g., color, shape, pattern, texture, etc.) of one of the fruits or vegetables, such as *mikan-ri’su* (tangerine-squirrel) or *ringo-ko’ara* (apple-koala) (see Figure 1 for an example display). Two and a half seconds after the presentation of the visual scene, the participants heard audio stimuli such as “Where is *mi’kan* /mikan-ri’su /ringo /ringo-ko’ara?” and were asked to click on the correct referent.

Immediately after hearing the compound modifier (which was segmentally ambiguous between a single head and modifier of a compound), listeners’ looks to the compound-denoting targets (Target Compound and Competitor Compound) increased when the initial noun was lexically accented and thus unambiguously underwent the CAR (e.g., “*mikan* . . .”), but not when the noun was lexically unaccented (e.g., “*ringo* . . .”). The pitch accent information also accelerated the process of eliminating the compound representation from the possible candidate set: single noun targets more rapidly attracted looks when the noun was lexically accented (unambiguously non-compound, e.g., “*mi’kan* . . .”) compared to when the noun was originally unaccented (ambiguous, e.g., “*ringo* . . .”). In the current study, we conduct the same

---

**Figure 1.** A sample visual display for the sentence *Mikan-ri’su wa do’re? “Where’s the tangerine-squirrel?”* (from Hirose & Mazuka, 2015).
experiment with 6- to 7-year-old children. The next section discusses our predictions and their implications in regard to the characteristics of compound prosody in Japanese.

**Predictive compound processing in children**

This study is intended to test whether the compound prosody in Japanese leads to a processing advantage at the interface between the lexical and the syntactic levels in young children. One possible prediction, extending Vogel and Raimy’s (2002) claim for English-speaking children, is that compound prosody will only facilitate the Japanese-speaking children’s comprehension of familiar/known compounds. If so, the prosodic manipulation of novel compounds in this study will result in no effect. On the other hand, the accented/unaccented distinction in Japanese is discrete, and directly signals whether the CAR is applied or not, which may lead to different results for Japanese than for English. In other words, the pitch accent indicating CAR application may allow children to immediately disambiguate between a compound modifier and a single head, regardless of whether a given compound representation exists in their lexicon. If this is the case, we should expect evidence for facilitated processing of both compounds and single head nouns when there is an explicit signal of (non-)application of the CAR, prior to the compound head. Such a finding in turn would lead us to further questions: (i) whether such a processing advantage is predictive in nature in a similar way for children and for adults (i.e., whether the choice between the compound and the single head status is made prior to the bottom-up evidence for presence/absence of the head noun); and (ii) how much slower the children’s eye-movement responses triggered by the prosodic information are in comparison to the adults’ eye-movement responses.

**Children’s access to the CAR in processing**

Children who are native speakers of Tokyo Japanese tend to correctly produce appropriate accent combinations as early as age 3. Shirose and Kiritani (2001) reported that children as young as 3 years of age mostly used the correct compound accent in word combinations like the examples in (1), where modifier nouns are either initially accented (i.e., on the initial mora) or unaccented originally (i.e., in the lexicon), and the outcome of the CAR is completely regular. Isobe (2007) further demonstrated that young children (2;10 to 4;2, mean age 3;7) correctly distinguished between a compound and a sequence of two single nouns with omitted case markers (e.g., kirin kukii (tabeta) “giraffe cookie (ate)” meaning “I (ate) a giraffe-shaped cookie” or “giraffe (ate) a cookie”) with over 90% accuracy in an off-line truth value judgment task with spoken stimuli. Although Isobe did not precisely control the accent type of the first noun (three items out of four used initially accented nouns for the compounds’ initial constituent), the children must have relied on the accent change due to the CAR, possibly in addition to durational information. The results of these two studies suggest that young children possess the ability to access phonological knowledge such as the CAR in language production and comprehension. This article investigates whether children are also able to use this knowledge on-line, or further, to actively hypothesize the compound structure prior to encountering the head noun, like the adults in Hirose and Mazuka’s (2015) study. It also focuses on the issue of how children might differ from adults in the timing of their use of this knowledge, as discussed in the next section.

**Children’s on-line sensitivity to linguistic information: The timing issue**

There have been ongoing attempts to investigate children’s on-line sensitivity to prosodic information. Studies on the role of contrastive pitch prominence in on-line referential resolution show that children around age 5 do not have adult-like sensitivity to the contrastive function of pitch accent information (Arnold, 2008; Sekerina & Trueswell, 2012). Ito et al. (2012) demonstrated that at 6 years of age, children start developing the ability to exploit pitch...
accent in contrast resolution in an on-line task (e.g., pinku-no usagi-wa doko? “Where’s the pink rabbit?” -> Jaa, ORENJI-no usagi-wa/saru-wa doko? “Then, where’s the ORANGE rabbit/monkey?”). However, they also found that the children needed extra time during the task to reset their attention before they could use the contrastive prosody in the target item (when they are expected to shift attention from, e.g., the pink rabbit in the first question above to the orange rabbit or monkey in the second question).

With respect to the timing of the effect, a number of Visual World Paradigm studies have reported that children’s eye-movement responses to linguistic information are delayed by several hundred milliseconds compared to adults. First of all, for adults, the time for programming and executing a saccadic eye movement is approximately 150–200 ms (Allopenna, Magnuson, & Tanenhaus, 1988; Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995) while that for 5-year-old children has been reported to be about 300 ms (Arnold, 2008; Trueswell et al., 1999). Children may not be as efficient as adults at processing various types of linguistic and extra-linguistic information. The difference may be manifested to a greater or lesser degree depending on the information type; therefore, the latency of the actual effect of a linguistic manipulation can also vary depending on the type of information being considered. For example, Snedeker and Yuan (2008) reported that a prosodic effect (i.e., on-line PP attachment guided by durational information marking a prosodic boundary) emerged in 4- to 6-year-old children with a delay of roughly 500 ms. Ito, Bibyk, Wagner, and Speer (2014) reported that the ability to use pitch accent to facilitate contrastive interpretation speeds up with age, but even children as old as 11 years exhibit the effect with an approximately 400 ms delay compared to adults. A study on syntactic priming effects in spoken sentence comprehension by Arai and Mazuka (2014) reported that the priming effect reflected in anticipatory eye movement to a possible subject in the primed construction was observed for 5- to 6-year-old children about 600 ms later than for adults.

Our present research question concerns the effect of pitch accent information, which is more tightly linked to lexical-level structures than to the discourse level of processing considered by Ito et al. (2014); therefore, the size of the delay between adults and children may be smaller compared to what has been reported in the previous studies. Nonetheless, there is no a priori way to precisely estimate the latency of an anticipatory effect in children. We therefore used a bottom-up method to identify the periods of time in which the effect of compound predictability is the most prominent for the adult and child groups separately, instead of pre-determining a common time window for our analyses.

**Experiment**

We tested 18 children of 6–7 years of age, using the same Visual World Paradigm procedure and materials as Hirose and Mazuka (2015).

**Method**

**Participants**

Eighteen first graders (6;5–7;3, mean age 6;9), without visual or hearing impairment, participated in the study. All children were recruited in Wako City, Saitama, where Tokyo Japanese is spoken. Both parents of all participants spoke dialects of the Tokyo area or its vicinity. The reason for this selection criterion was to avoid possible influences of regional differences in lexical pitch accent.

**Materials**

All visual and audio materials were the same as those used in Hirose and Mazuka (2015). There were 24 experimental items that began with either a lexically accented or a lexically unaccented tri-moraic noun (12 accented (= unambiguous) and 12 unaccented (= ambiguous)).
Each item consisted of a visual display and a recorded sentence using the pattern “~ wa dore?” (“Where is ~?”). Noun type (single vs. compound) was crossed with whether the target noun underwent an accent change when it was part of a compound (Unambiguous vs. Ambiguous), as shown in (2a–d). The 12 compounds were made by combining one of the six different fruits/vegetables with one of the six different animals (repeated between accent conditions) in each C1 accent type.

The acoustic profiles of the critical nouns are reported in Table 1. The speech rate was considerably slower than what would be expected for natural conversation among adults; the adult participants in the original study (Hirose & Mazuka, 2015) performed the task without finding any of the audio stimuli unnatural or strange. The mean and max F0 for the first two morae of single nouns yielded significant differences between Unambiguous (accented in isolation) and Ambiguous (lexically unaccented in isolation) conditions (all $p < .01$ by independent $t$-tests). It was also confirmed that the CAR affected the F0 measures on the initial two morae for accented words in the Unambiguous set, yielding significant differences between compound and single noun conditions ($p < .01$ for mean and max F0), while these measures did not significantly differ for unaccented words in the Ambiguous set. These are the critical premises in our manipulation of items.

In fact, a further inspection of the third mora in our stimuli materials revealed something we did not plan: the mean F0 and the peak F0 of the third mora of the unaccented compound modifiers (e.g., go in ringo koara in the ambiguous set) were higher than those of the unaccented single noun (e.g., go in ringo in the unambiguous set) ($p < .01$ for both by independent $t$-tests). In addition, the single nouns tended to be longer in duration compared to the compound modifiers, which was also the case for the Unambiguous set. (See Table 1.) These differences are presumably phonologically non-meaningful, and we do not have evidence that such differences exist in naturally occurring single noun vs. compound modifier utterances. At least for the adults, the differences in our stimuli had no impact on their on-line interpretations as reflected in their eye-movements (Hirose & Mazuka, 2015). We will come back to this issue in the discussion of our child data.

Each visual display was divided into eight areas, each containing one of the object types listed in (3a–h). The positions of the different object types were varied across items. An example display can be seen in Figure 1.
Procedure
Participants sat in front of a Tobii 1750 eye-tracker. Their eyes were calibrated using Tobii ClearView. The stimuli were presented using E-Prime 1.2. The structure of a trial is illustrated in Figure 2.

For each trial, after successful fixation of 1,000 ms on a fixation cross, the visual display was presented. The sentence was played through a pair of speakers 2,500 ms later. Participants then pointed to the object that they thought was the referent. The experimenter clicked on the designated object for the participant as soon as the response was identified. Participants’ eye movements were recorded from the onset of the auditory stimuli until the mouse click, at a sampling rate of 50 Hz. The experiment took approximately 25 min.

Data analysis and results
Both adults and children performed accurately in selecting (clicking or pointing) the correct target (Table 2).
Once listeners are able to recognize the presence or absence of the subsequent head noun, their gazes will tend to be fixated on the designated target. Our prediction was that if the listeners can exploit the unambiguous status of the compound/single noun accent of the first noun, their looks to the candidate(s) for the correct target would be initiated relatively faster than in the ambiguous conditions.

We expected the divergence in the children’s fixation behavior between conditions to occur at a later point in time than in the adults’, but we had no a priori predictions for any specific period of delay. Therefore, we used permutation-based cluster analysis (also referred to as “nonparametric statistical testing”; Maris & Oosterveld, 2007) as a bottom-up approach to objectively determining the stretch of time (time clusters) in which the two conditions produced a significant difference. To facilitate direct comparison between the data from the current experiment and from the experiment conducted by Hirose and Mazuka (2015), we re-analyzed their data using the same permutation-based analysis procedure, described below.

We broke our data down into 100 ms time windows, starting from the onset of the second mora of the first noun, where the difference in the accent pattern between the conditions first becomes evident. We initially considered 12 time windows from the onset up to 1,200 ms, because the clicking responses started at around 1,300 ms in some of the trials. For each of the visual targets, we summed the number of gazes for each window. The test statistics (i.e., \( t \)-statistic assessing the linear model fit) on the difference between the two experimental conditions were calculated for each time window. Consecutive windows in which the test statistics reached significance at the 0.05 level in the same direction were clustered together. Next, for each such cluster, the cluster-level test statistic was obtained, which is the sum of the test statistics for the individual time windows within the cluster. A re-sampling procedure was then applied within each time window, whereby the original data sets were re-assigned randomly to the two conditions 1,000 times to create a permutation distribution of the resulting test statistics. The obtained permutation distributions for all 100 ms windows within a given cluster were then combined into a cluster permutation distribution. By calculating the proportion of \( t \)-values in the distribution that were larger than the cluster \( t \)-value, we obtained the cluster \( p \)-value; for example, if the observed cluster \( t \)-value is greater than 97.5% (= .05 level of significance in a two-tailed test) of the cluster permutation distribution, the reliability of the identified cluster is confirmed.

Our primary analyses focused on the looks to the two compound candidates. Figures 3 and 4 plot the proportion of looks to designated visual objects (Target Compound and Competitor Compounds\(^3\)), as proportion can be quite intuitively interpreted. For the statistical analyses, we used the logit (log of looks to these critical visual object(s) against all others) as the dependent variable as it is less distorted by cases of tracking loss. The 12 time windows (of 100 ms) from 0–1,200 ms from the offset of the second mora were considered in the analysis, because clicking occurred as fast as 1,200 ms from the onset in some of the adult trials. The 1,200 ms roughly covered the entirety of the compound utterances; none of the children made pointing responses within this time period.

\(^3\)Including a Competitor Compound object was considered crucial in the original Hirose and Mazuka (2015) experiment to test whether the processing advantage of the Unambiguous prosody found in adults was anticipatory in nature, which could take place before the segmental information of the head noun. In the present study, direct comparisons between Target Compounds and Competitor Compounds are not reported: such comparisons would not be very informative, given the size of the latency for the prosodic unambiguity effect in children.

### Table 2. Clicking/pointing accuracy in adults and children.

<table>
<thead>
<tr>
<th>Clicking/pointing accuracy (%)</th>
<th>Unambiguous compound</th>
<th>Ambiguous compound</th>
<th>Unambiguous single noun</th>
<th>Ambiguous single noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adults</td>
<td>98.7</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Children</td>
<td>95.3</td>
<td>94.3</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Anticipating a compound

The first set of analyses address the question of whether pitch accent change can result in early anticipation of the compound representation. When a listener is presented with a scene such as Figure 1 and has heard a spoken question asking about a compound object (e.g., *mikan ri'su* "tangerine squirrel"), there is no question that the listener will eventually look at the Target Compound to click on/point to it. Here, we are interested in the eye-movement pattern when the listeners have only heard the first constituent of the compound (e.g., *mikan*). If the absence of the HL accent, which would be present if the noun were a single head, initiates a prediction of the compound interpretation, the listener will start looking at compound-looking objects consistent with the input so far, namely either a Target Compound or a Competitor Compound, rather than at a Target Singleton that is only segmentally consistent with the input. This is what we predict for the Unambiguous set. In contrast, if there is no obvious accentual difference before and after the CAR is applied, as in the Ambiguous set, a concentration of looks to the relevant compound targets will not start until the presence of the head noun becomes evident. We also conducted a direct comparison between the Unambiguous and Ambiguous conditions asking for the Target Compound, which was not considered in Hirose and Mazuka (2015). This should allow us to evaluate the comprehenders’ sensitivity to the prosodic information regarding whether the pitch accent changes from the lexically determined original pattern.

Figure 3 (adults) and Figure 4 (children) plot the proportion of looks to the two target objects combined, starting from the onset of the second mora of the first noun, depending on whether the linguistic stimulus was asking for the compound target (solid lines) or the single noun target (dotted lines).

---

**Figure 3.** Proportion of looks to the Target Compounds + Competitor Compounds combined for the reanalyzed adult data from Hirose and Mazuka (2015), showing the time clusters for the effect identified by the permutation-based cluster analysis. For comparisons between Compound and Single N linguistic stimuli, the black bar (bottom) indicates the time span for the significant effect for the Unambiguous set, and the gray bar (bottom), the Ambiguous set. The significant time cluster for the comparison between Unambiguous vs. Ambiguous Compound stimuli is indicated by the striped bar (top). The x-axis is aligned to the offset of the second mora of the first noun. Relevant word boundaries are indicated by vertical lines.

---

*A* value reported here is the sum of the t-statistics for the individual time windows within the entire cluster in consideration. A *p* value reported here refers to the proportion of the overall permutation distribution within the cluster that has a greater test statistic than the observed value.
The first vertical lines indicate the onset of the head noun (for the compound utterances). The horizontal bars represent the clusters of time windows where looks to the compound objects reliably deviated between the single noun and the compound conditions (the black bars are for the Unambiguous set, and the gray bars for the Ambiguous set). The significant time cluster for the comparison between Unambiguous vs. Ambiguous Compound stimuli is indicated by the striped bar (top). The x-axis is aligned to the offset of the second mora of the first noun. Relevant word boundaries are indicated by vertical lines.

We first consider the Unambiguous set, where the first noun, whether compound modifier or single head, was originally accented in the lexicon. For the adults, as can be seen by comparing the black solid and dotted lines in Figure 3, there were significantly more looks to the compound targets in the compound condition compared to the single noun condition from 0–1,200 ms (indicated by the black bar in Figure 3, cluster \( t = 82.69, \) Cluster \( p < .001 \)). This effect started immediately following the onset of the second mora of the compound modifier, reconfirming adults’ pre-head prediction of the compound representation. For the children, the same effect was identified from 600–1,200 ms (indicated by the black bar in Figure 3, cluster \( t = 30.82, \) Cluster \( p < .001 \)).

To evaluate whether the children had to wait until the presence or absence of the head noun became evident, or used the accent change information to initiate the decision-making process, let us compare these results with those from the Ambiguous set.

The Ambiguous sets (temporarily ambiguous between single noun and compound accent) produced different response patterns. For adults, the two conditions produced a significant difference during the 700–1,200 ms cluster (indicated by the gray bar in Figure 3, cluster \( t = 28.62, \) Cluster \( p < .001 \)) (compare the gray solid vs. dotted lines in Figure 3). For children (compare the gray solid vs. dotted lines in Figure 4), there were two clusters where the two conditions produced a reliable difference, but in opposite directions. Children looked significantly more at the correct single noun target during the 900–1,200 ms cluster (indicated by the second gray bar in Figure 4, cluster \( t = 13.46, \) Cluster \( p < .001 \)). This was an effect in the same direction as that for adults with some delay, which was reasonably expected. For both groups, the onset of the divergence was well after the
time the presence or absence of the subsequent head noun would be fully recognized. Compared to
the adults, the children generally took longer to identify that the auditory stimuli referred to a
compound, but the children were relatively faster when the first noun unambiguously carried the
compound accent. The children generally tended to look at the compound targets more often
compared to the adults, regardless of the prosodic information of the input prior to the disambigua-
tion by the presence or absence of the second noun as a compound head. This may reflect a possible
overall tendency of children to pay more attention to more salient and animate entities (e.g.,
imaginary creatures) than to simple inanimate objects (e.g., fruit). What was rather unexpected
was a transient effect going in the opposite direction, found at 500–700 ms (indicated by the first
gray bar in Figure 4, cluster $t = 5.14$, Cluster $p < .005$). This finding will be discussed in the general
discussion section together with the findings reported in the next section.

Finally, the looks to the two candidate compound targets were directly compared between the
Unambiguous and the Ambiguous linguistic conditions (shown by the black and gray solid lines in
Figure 3 and 4). For adults, the two conditions produced a significant difference, demonstrating the
advantage of the Unambiguous input during the 100–900 ms cluster (indicated by the striped bar in
Figure 3, cluster $t = 29.93$, Cluster $p < .001$). For children, the same effect appeared during the
500–800 ms cluster (indicated by the striped bar in Figure 4, cluster $t = 7.57$, Cluster $p < .001$). Both
adults and children reacted to the audio instructions more quickly when the stimuli unambiguously
carried the compound accent. The overall reaction speed was slower in children by about 400 ms.

**Falsifying the single noun interpretation**
The second set of analyses was conducted to confirm that the pitch accent change due to the CAR
accelerates the process of ruling out the single noun interpretation, which is otherwise the default
analysis. Hearing a lexically accented word like *mi’kan* “tangerine” where it is clear that the CAR
does not apply should instantly rule out the possibility that it is part of a compound, without the
need to confirm the absence of a head noun (Unambiguous set). In such cases, the listeners were
expected to look at the Target Singleton more than the compound candidates with which the input
was segmentally consistent. If it were *ringo* “apple”, the apparent absence of the HL accent nucleus
would not be evidence either for or against single noun status (Ambiguous conditions). An increase
of looks to the Target Singleton over the compound targets may not be observed until after the offset
of the first noun. Again, we are interested in whether the effect of the Unambiguous conditions can
be observed in children, and if so, what the time course of the effect is, as compared to that for
adults.

Let us again consider the Unambiguous conditions first. Figure 5 (adults) and Figure 6
(children) plot the proportion of looks to the single noun target, depending on whether the
first noun in the linguistic stimulus was indeed a single head noun (dotted lines) or a modifier in
a compound (solid lines). The pattern is analogous to the results reported in the last section. For
adults, the proportion of looks to the single noun target was significantly greater in the single
noun condition compared to the compound condition during the cluster of 0–1,200 ms (indi-
cated by the black bar in Figure 5, cluster $t = 104.517$, Cluster $p < .001$) (compare the black
dotted vs. solid lines in Figure 5). For children, the same effect was identified from 400–1,200 ms
(indicated by the black bar in Figure 6, cluster $t = 56.942$, Cluster $p < .001$) (compare the black
dotted vs. solid lines in Figure 6). Adults were able to rule out the compound analysis prior to
the onset of the head noun, while the children were able to do so after a delay of approximately
400 ms, at which point the presence or absence of the head noun was evident in the input.

The data from the Ambiguous conditions also exhibited the mirror image of the data reported in
the last section. For adults, the two conditions significantly deviated from each other during the
700–1,200 ms cluster (indicated by the gray bar in Figure 5, cluster $t = 50.502$, Cluster $p < .001$)
(compare the gray dotted vs. solid lines in Figure 5). For children, there were again two clusters
producing opposite effects, during the 900–1,200 ms (indicated by the second gray bar in Figure 6,
cluster $t = 17.967$, Cluster $p < .001$), where the single noun stimuli led to more looks to the correct
Figure 5. Proportion of looks to the Target Singleton for the reanalyzed adult data from Hirose and Mazuka (2015), showing the time clusters for the effect identified by the permutation-based cluster analysis. For comparisons between Compound and Single N linguistic stimuli, the black bar (top) indicates the time span for the significant effect for the Unambiguous set, and the gray bars (top), the Ambiguous set. The significant time cluster for the comparison between Unambiguous vs. Ambiguous Single N stimuli is indicated by the striped bar (top). The x-axis is aligned to the offset of the second mora of the first noun. Relevant word boundaries are indicated by vertical lines.

Figure 6. Proportion of looks to the Target Singleton for the child data, showing the time clusters for the effect identified by the permutation-based cluster analysis. For comparisons between Compound and Single N linguistic stimuli, the black bar (top) indicates the time span for the significant effect for the Unambiguous set, and the gray bars (top), the Ambiguous set. The significant time cluster for the comparison between Unambiguous vs. Ambiguous single N stimuli is indicated by the striped bar (top). The x-axis is aligned to the offset of the second mora of the first noun. Relevant word boundaries are indicated by vertical lines.
single noun targets, and also at an earlier point, when the compound candidates attracted more looks for a short time between 400 ms and 700 ms (indicated by the first gray bar for Figure 6, cluster $t = 8.08$, Cluster $p < .001$) (compare the gray dotted vs. solid lines in Figure 6).

Again, for both groups, the identification of the correct single noun target occurred well after the time the presence or absence of the subsequent head noun would be fully recognized. The eye-movement patterns also suggested that children were initially somehow misled by the prosodic information towards the opposite of the correct interpretation, whereas adults showed no such pattern.

A direct comparison of looks to the Target Singleton object between the Unambiguous and the Ambiguous linguistic conditions was also conducted. For adults, the target objects corresponding to lexically accented nouns (i.e., Unambiguous, e.g., mi’kan) started to attract more looks compared with those corresponding to lexically unaccented nouns (i.e., Ambiguous, e.g., ringo) during the 300–800 ms cluster (indicated by the striped bar in Figure 5, cluster $t = 16.76$, Cluster $p < .001$) (compare the black and gray dotted lines in Figure 5). This is not in line with the notion that the single head noun interpretation is always the default analysis as long as the prosodic information explicitly contradicts with that analysis; it seems that adult listeners are somewhat indecisive to commit to the simplest single head analysis without a positive signal that CAR has applied. For children, the same effect appeared during the 500–800 ms cluster (indicated by the striped bar in Figure 6, cluster $t = 29.22$, Cluster $p < .001$) (compare the black and gray dotted lines in Figure 6). Adults and children both identified the correct single-noun object faster when the pitch accent eliminated the possibility of it being a compound modifier. It is important that this effect is arguably independent from the children’s unexpected pattern of reaction to the compound-denoting audio stimuli reported earlier. The overall reaction speed was faster in adults by approximately 200 ms.

**Estimating the delay in response to the segmental input**

To better evaluate the relative delay between children and adults in accelerated compound processing when the prosody provides a pre-head cue (i.e., in the Unambiguous condition), it would be desirable to have some basis to examine the cause of the delay. For example, if the experimental task, with its eight visual objects, is so complex that it interferes with children’s natural on-line interpretation of the auditory input, the nature of the delay might be irrelevant to predictive processing exploiting the change in accent. In addition, we encountered a difficulty in interpreting the children’s data reported so far: the design of the experiment relied upon the assumption that the Ambiguous conditions would provide a baseline to evaluate the processing advantage due to the accent change on the pre-head input. As it turned out, however, the children temporarily reacted to the input lacking the accent change in an unpredicted direction. This makes it difficult to use the Ambiguous condition results as a baseline for children. We need an independent basis to argue that the size of the delay in children’s responses compared to adults’ in the Unambiguous conditions was not solely due to slowness in executing the task, but has something to do with the on-line process of referring to phonological knowledge about the compound accent. We conducted an additional post-hoc analysis to address these issues. We calculated the looks to the Target Compound + Competitor Compound + Target Singleton combined (e.g., (a: mikan ri’su “tangerine squirrel”) + (c: mikan ta’nuki “tangerine raccoon”) + (b: mi’kan “tangerine”) in Figure 1) in the Unambiguous compound condition. We then compared that to the looks to the two Irrelevant compounds + Irrelevant simplex combined (e.g., (g: meron ri’su “melon squirrel”) + (h: meron ta’nuki “melon raccoon”) + (f: me’ron “melon”) in Figure 1) from the very onset of the target word. The two groups of visual targets are supposed to be comparable in visual complexity (two imaginary complex creatures and a simple fruit). The intention is to capture the speed of reaction to the segmental information alone: the first mora (e.g., mi’ as in mikan) would rule out the latter group because it does not match the first mora of the first noun that the three members have in common. We call this the “Mismatch” group (e.g., for mi’kan, the Mismatch group contains meron ta’nuki “melon raccoon”, meron ri’su “melon squirrel”, and meron “melon”). The listeners would further need to wait for the accent information and the semantic content of the head noun to choose among the former group (e.g.,
mikan ri’su “tangerine squirrel”, mikan ta’nuki “tangerine raccoon”, and mi’kan “tangerine”). This is the “Match” group, which involves the process that the data presented so far directly addresses. At the point of the first process (eliminating the Mismatch group), the apparent overall bias in children towards compound objects should not interfere with the comparison as we look at the collective gazes in each group where any background preference towards compound objects should affect both groups.

Figure 7 presents the proportion of looks to any of the members in each group. In this analysis, the dependent variable was the sum of gazes for each 100 ms window collected from the 20 ms bin within each window. For the adults, the number of looks to the Match group and the Mismatch group showed a significant difference at 500 ms from the word onset (cluster $t = 77.72$, Cluster $p < .001$, as indicated by the black bar in Figure 7); for the children, the significant difference appeared at 600 ms (cluster $t = 66.58$, Cluster $p < .001$, as indicated by the gray bar in Figure 7). That is, children stop paying attention to the visual objects that are not consistent with the very first mora of the input almost as quickly as adults, with only a 100 ms delay. This allows us to say that the 600 ms delay reported earlier in the section Anticipating a Compound for children compared to adults (in choosing one of the compound objects consistent with the input in the Unambiguous condition) was not entirely due to a slowdown caused by dealing with multiple visual options.

Alternative analysis on the timing of selecting the single noun interpretation

Part of the analysis presented in the section Falsifying the Single Noun Interpretation attempted to demonstrate that the single noun interpretation was selected faster for lexically accented nouns (Unambiguous) than for lexically unaccented nouns (Ambiguous). Our intention was to compare the ratio of looks to the single noun target and to the relevant compound candidates. The children’s unexpected reaction to the audio stimuli asking for the compound target complicated the comparison, however. To follow up on our argument, we compared the relative timing of looks to the single noun target between adults and children. This is also a between-item analysis, but the timing at which the difference between the conditions (Unambiguous vs. Ambiguous) occurs can be directly compared...
between adults and children, to provide an estimate of the effect of non-ambiguous accent information on the first-encountered noun.  

Figure 8 presents the looks to the Target Singleton when that is the correct target object referred to by the audio stimuli. The onset of the x-axis is aligned to the offset of the second mora. Both adults and children were faster at identifying lexically accented nouns (Unambiguously single head nouns) as the target than at identifying lexically unaccented nouns (Ambiguous between single head nouns and modifiers of compounds) as the target. The onset of that effect was faster in adults (starting at 600 ms, as indicated by the black bar in Figure 8, cluster $t = 25.43$, Cluster $p < .001$) than in children (starting at 1,100 ms, as indicated by the gray bar in Figure 8, cluster $t = 20.54$, Cluster $p < .001$) with a difference of 500 ms, which is about the same as the difference in the timing of the effect of falsifying the compound interpretation reported earlier.

Discussion

*When there is a pre-head cue on the modifier (Unambiguous conditions), can children use the cue? if so, how much longer do they take compared to adults?*

There was a larger delay (400–600 ms) in the eye-movement response of this study’s 6- to 7-year-old children compared to adults in processing the pitch accent information to deduce the application (or non-application) of the CAR. For the children, the fact that increases in fixations to compound targets took place 300 ms earlier in the Unambiguous condition than in the Ambiguous condition suggests that the pitch accent change facilitates the processing of a compound representation. Analyses of the looks to the visual objects denoting the single noun interpretation were also consistent with this interpretation; looks to the correct single noun target increased for unambiguous (accented) nouns 500 ms earlier than for ambiguous (unaccented) nouns in the child data. It should be mentioned that, because of the delay of effect compared to adults, the difference in eye movements occurred only after the onset of the difference in the acoustic stimuli (even if that difference had not yet been processed), which makes it difficult to claim that the effects found in children really reflect pre-head anticipation instead of the overall facilitation of the
comprehension process. It should be noted that the Ambiguous conditions (using the lexically unaccented nouns), which were originally meant to provide a baseline to evaluate evidence for anticipation in the Unambiguous conditions, might not be appropriate as a baseline, due to the possible extra processing cost of the contrastive information, as discussed above. We will provide arguments in detail below.

When there is no early cue on the modifier (Ambiguous conditions)

When the first noun they encountered was lexically unaccented (and thus temporarily ambiguous at both segmental and prosodic levels), neither adults nor children distinguished between a compound modifier and a single head noun until after they recognized the subsequent disambiguating segmental input: for compounds, that would be the compound head, while for single nouns, that would be the particle (wa “TOP”) plus the wh-expression (do’re “where”). These results were expected, based on the previous research (Hirose & Mazuka, 2015). The children’s latency in identification, as seen in an increase of looks to the relevant compound candidates, was only 100 ms, compared to adults.

There was an unexpected finding. At a point before the children finally correctly identified the target, they appeared to temporarily interpret the stimuli incorrectly (e.g., when hearing the ringo of ringo ko’ara, their looks to the single noun target temporarily increased, and vice versa). What causes such responses must be suprasegmental information at the first nouns (as they were segmentally identical, and the following word should lead to complete disambiguation). As we discussed in the materials section, the mean F0 and the peak F0 of the third mora of the unaccented compound modifiers (e.g., go as in ringo ko’ara) were higher than those of the unaccented single nouns (e.g., go as in ringo) in our stimuli. Even though this acoustic difference did not affect adults’ eye-movement patterns, we can imagine the possibility that children interpret such minor acoustic differences in their own way. We do not know exactly what acoustic correlate (e.g., the elevated pitch peak, the enlarged pitch range, or both) is responsible for this seemingly strange effect. Either do we have a definitive account as to why such pitch prominence on the compound modifier led the children towards the single head noun analysis of that noun (against the speaker’s intention). This might have been some automatic process reacting to the f0 acoustic phenomena, or it might reflect a particular way in which children at this age give such acoustic phenomena certain phonological interpretations (e.g. contrastive information). These are questions for future investigation.

Do children process compounds with an Unambiguous accent and an Ambiguous accent differently?

Let us turn to the direct comparisons within the compound cases and within the single noun cases. Both adults and children focused on the correct (type of) targets faster when the noun unambiguously carried the compound accent or the single head accent, compared to when the accent of the noun was compatible with either analysis. This demonstrates that both adults and children do use the pitch accent information to expedite the compound or single noun processing whenever possible.

For adults, we can conclude this is an anticipatory process because the effect emerges at the point before the presence or absence of the head noun becomes evident. For children, the same effect was observed about 400 ms later than for adults, by which time the presence or the absence of the following noun has already been revealed. This latency makes it difficult for the pitch accent information to contribute to children’s pre-head predictive processing; however, we can conclude that children do exploit the pitch accent information to facilitate distinguishing between the compound and single head noun during processing.

Compound stress change in English vs. compound accent change in Japanese

The CAR is a mandatory rule, and no other phonological rule can produce the same output in Japanese. We have argued above that Japanese-speaking children are sensitive to the change in accent unambiguously signaling application or non-application of the CAR in the on-line
comprehension process, although they did not respond fast enough to demonstrate the advantage prior to encountering the bottom-up information as to whether the head noun follows. This finding appears to contrast with the situation in English: even though a noun can exhibit distinctive stress status whether it is the modifier part of a compound or an independent head, children, including young teenagers, are not generally successful in discriminating between the two alternative representations (Atkinson-King, 1973; Good, 2008; Vogel & Raimy, 2002). Because our results rule out the possibility that young children in general are insensitive to prosodic cues to word-internal structure, the apparent poor performance of children in compound comprehension in English may be due to the fact that the prosodic correlate associated with compound stress in English is more ambiguous or less reliable than in Japanese. For one thing, as discussed already, the primary/non-primary status of word stress cannot be determined in isolation (Kunter, 2011: p. 15); therefore, stress potentially carries a temporary ambiguity. For another, stress in English can be used to mark new information, “and thus could cause the general pattern to be overridden” (Vogel & Raimy, 2002). Vogel and Raimy further pointed out that confusion arises because new information is typically provided by the first constituent of a compound, which may be another source of ambiguity regarding stress status in English.

**Conclusion**

Based on the results from Japanese, where accent change can provide a dichotomous cue between a single noun and a compound, we argue that young speakers of any language might possibly have the same on-line sensitivity to a change in prosodic prominence if it unambiguously signals a particular structural representation. The lack of evidence for such a claim to date may be ascribed to two facts. First, in many cases prosodic prominence is ambiguously correlated with more than one possible process (e.g., compound stress and contrastive focus in English), unlike the present case in Japanese. Second, the effect in children is difficult to catch on the fly due to the processing delay. Our study suggests that there are at least two components to the approximately 400–600 ms of delay found in the children’s responses, as compared to the adults’ responses. One component is a delay in object identification based on the segmental information alone. The relevant finding in our study would be the 100 ms delay reported in the post-hoc analysis illustrated in Figure 7. The other component is the delay associated with the detection of evidence for the application or non-application of the CAR and preparation for the appropriate nominal structure. An example of relevant evidence for this type of delay is the finding that the processing advantage of the pitch accent change for processing a compound analysis was detected in children with an approximately 400 ms delay compared to adults (the section *Anticipating a Compound*).

This study has demonstrated that both children and adults process compounds and single nouns faster when the accent pattern of the noun unambiguously signals whether or not it is a compound accent. A question that remains open is whether the children’s sensitivity to the pitch accent information reported in this study actually reflects a predictive process that is the same as the adults’, in which case, children are able to make the decision on the compound status of the input only on the basis of the first-encountered noun alone, prior to the subsequent input. Given the delay of effect in children, we are unable to provide a clear answer in this study. Future research, possibly adopting a simpler visual scene (i.e., fewer visual objects to choose from) which could enable children to respond faster with less cognitive burden might allow us to pin down the question whether children can indeed make a pre-head anticipatory decision.

**Acknowledgments**

We are greatly indebted to Franklin Chang, Douglas Roland and Sho Tsuji for their various comments and advice. We deeply appreciate various valuable suggestions from the anonymous reviewers and the editor of the journal.
**Funding**

This research was supported in part by MEXT/JSPS KAKENHI Grant Number 25580086 to the first author.

**References**


