Comprehension of underspecified iterative meaning in Japanese

Yao-Ying Lai†, Michiru Makuuchi†, Maria Mercedes Piñango‡ and Sakai Hiromu⊺

†Research Institute, National rehabilitation Center for Persons with Disabilities
1, Namiki 4-chome, Tokorozawa City, Saitama Prefecture 359-8555, Japan
‡Department of Linguistics, Yale University
370 Temple St. New Haven, CT 06520-8366, U.S.A.
⊺Waseda University
3-4-1 Okubo, Shinjuku City, Tokyo, 169-8555, Japan

E-mail: †yaoyinglai@gmail.com

Abstract
We examine the processing of underspecified iterative meaning in sentences like “The athlete jumped for 20 minutes” in Japanese, crossing two verb types (Punctual: jump / Durative: jog) and the length of the intervals denoted by for-adverbials (Short: ten minutes / Long: a year). Additionally, we investigate the impact of individuals’ autistic tendency (AQ) during sentence comprehension. We found that the naturalness of sentences decreased as a function of meaning uncertainty. The self-paced reading experiment showed that sentences with underspecified meaning engendered longer reading times regardless of verb type, and that individuals with higher AQ scores (> low-AQ) stayed longer at the sentence-final region.

Key words sentence processing, aspectual coercion, underspecification, partition measure, self-paced reading, Japanese

1. Introduction
In this study, we investigate the comprehension of sentences like (1), in which an iterative meaning is morpho-syntactically unsupported yet fully understood.

(1) The athlete jumped for 20 minutes at the gym.

Such sentences have been shown to incur additional processing cost and brain activity, as compared to their transparent counterparts like (2) that involve no underspecified iteration (e.g.,

(2) The athlete jogged for 20 minutes at the gym.

Based on these observations, we ask the following questions: (a) how do comprehenders obtain the underspecified iterative meaning, and (b) what is the source of the cost associated with processing sentences like (1)? By addressing these questions, this study aims to uncover the semantic-conceptual components that support meaning comprehension and to investigate the influence of individual difference in social cognitive style on real-time sentence processing. We probe through Japanese, a head-final language typologically different from English.

To capture the underspecified iterative meaning and the associated cost, the ITER-operator hypothesis assumes that durative for-adverbials select for durative verbs and argues that sentences like (1) involve an aspectual mismatch between the punctual verb (e.g., jump) and the durative for-adverbial (e.g., for 20 minutes). This mismatch calls for an ITER-operator to be inserted during comprehension to yield the iterative meaning, increasing the processing cost (e.g., Brennan & Pylkkänan 2008, Paczynski et al. 2014; Piñango et al. 1999, Todorova et al. 2000).

(3) The athlete jumped for 20 minutes at the gym.

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.

Copyright ©2019 by IEICE
(3) The athlete jogged for 2 months at the gym.

The iterative meaning in (3) is captured by the alternative Partition-Measure hypothesis proposed by Deo & Piñango (2011). It assumes no mismatch in (1) and instead argues that for-adverbials require a partition measure over the sub-intervals of the interval they denote (20 minutes) for determining the length of the sub-intervals. An infinitesimal partition measure (with no gap between the sub-intervals) gives rise to a continuous reading, while a non-infinitesimal partition measure (with gaps between the sub-intervals) gives rise to an iterative reading (Deo 2009).

On this hypothesis, the infinitesimal partition measure applies as default; yet when the consequent continuous reading turns out implausible in context, the processor applies the non-infinitesimal partition measure, mining the lexical conceptual structure of the verb and the sentential/discourse context to determine the lengths of the sub-intervals of the interval denoted by for-adverbials. The cost associated with underspecified iteration in (1) and (3) results from this conceptual/contextual search in context for an appropriate partition measure, so as to obtain an appropriate interpretation.

To examine the cognitive mechanisms supporting the comprehension of underspecified meaning and adjudicate the two accounts, we crossed Verb Type (punctual/durative) and the length of the Interval denoted by for-adverbials (short/long) in a naturalness-rating questionnaire (Exp. 1) and a self-paced reading (Exp. 2) experiment. The composition of a durative verb with a durative for-adverbial of a short interval (e.g., jogged for 20 minutes) gives rise to a continuous reading, serving as the baseline. In addition, we investigated the impact of individuals’ cognitive processing style indexed by Autism Spectrum Quotient (AQ, Baron-Cohen et al. 2001; 若林 et al. 2006), which has been shown to factor into language processing (e.g., Caruana & Brock 2014; Yu 2010; Zhang 2018). The ITER-operator account predicts a main effect of verb type, such that sentences with punctual verbs will be more taxing than those with durative verbs due to an aspectual mismatch in composition with durative for-adverbials. On the other hand, the Partition-Measure hypothesis predicts that all sentences with underspecified iterative meaning will engender more cost than their transparent counterparts regardless of verb type (i.e. a main effect of Interval and an interaction showing an effect of verb type within short-interval). Besides, we expect that individuals with higher autistic traits will require more effort in processing underspecified meaning due to a greater difficulty in conceptual-contextual integration (e.g., Minshew & Goldstein 1998) or a tendency towards deliberative reasoning (Brosnan et al. 2016) during comprehension.

2. Experiment 1: naturalness-rating questionnaire

2.1. Method

2.1.1. Materials

We created fifty sets of 4 conditions in Japanese (Table 1), crossing Verb Type (punctual/durative) of the critical verbs and the length of the Interval (short/long) denoted by for-adverbials (~間). Another 80 fillers were introduced, half of them being nonsensical to check

<table>
<thead>
<tr>
<th>Condition</th>
<th>for-adverbial</th>
<th>Seg.4 (Verb)</th>
<th>Seg.5 (V +1)</th>
<th>Seg.6 (V +2)</th>
<th>Seg.7 (S-final)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) No_Iter Durative verb + Short interval</td>
<td>選手が athlete-NOM</td>
<td>20分間 20min-for</td>
<td>体育館で Gym-LOC</td>
<td>ジョギングしたと jog-PAST-COMP</td>
<td>学生は student-TOP</td>
</tr>
<tr>
<td>(B) Punctual_Iter Punctual verb + Short interval</td>
<td>------------</td>
<td>20分間 20min-for</td>
<td>ジャンプしたと jump-PAST-COMP</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>(C) Durative_Iter Durative verb + Long interval</td>
<td>2ヶ月間 2month-for</td>
<td>ジョギングしたと jog-PAST-COMP</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>(D) Gapped_Iter Punctual verb + Long interval</td>
<td>2ヶ月間 2month-for</td>
<td>ジャンプしたと jump-PAST-COMP</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
</tbody>
</table>
if the participants attended to the reading task. An item-analysis with ANOVAs on verb frequency showed no significant difference in either their lemma forms ($F_{(1,98)}=0.15$, $p=.699$) or the past-tense forms ($F_{(1,98)}=0.258$, $p=.613$) between the two verb types. Verb lengths did not differ neither ($F_{(1,98)}=0.141$, $p=.708$).

2.1.2. Participants

Twenty-five participants were recruited from Waseda University in Japan (18 female); all aged 18–30 (mean age = 20.64) and without reading disability by self-report. Written consent was obtained from each participant prior to the experiment.

2.1.3. Procedures

The questionnaire was implemented via the Qualtrics Survey Tools (Qualtrics, Provo, UT). The participants were asked to rate the naturalness of each sentence from a scale 1–5 (5=completely natural) after reading, and then choose the interpretation for the sentence regarding the iterativity of the event denoted by the predicate (once, more than twice, more than 10 times, more than 100 times, or nonsensical).

2.1.4. Data analysis

Three participants were identified as outliers and their data were excluded prior to the statistical analyses. For the remaining 22 participants, the rating scores were transformed to $z$-scores to avoid potential scale bias (Shütze & Sprouse 2014). We performed a mixed-effects model analysis in the R statistical package environment (Bates et al. 2015; R Core Team 2017). Effects were assessed by model comparisons via Likelihood Ratio Tests (Winter 2013). All models included random intercepts for participant and item-set as well as random slopes for verb type plus interval over participants and item-sets (simplified from the maximal random-effect structure that failed).

2.2. Results

The rating scores are summarized in Table 2. All critical conditions were rated above 3 in average. The mixed-effect model analysis revealed a significant main effect of Interval ($\chi^2(1)=21.236, p<.001$), a significant main effect of Verb Type ($\chi^2(1)=14.221, p<.001$), as well as a significant interaction between these two factors ($\chi^2(1)=13.047, p<.001$). All pairwise comparisons (corrected) among conditions showed significant differences except for the comparison between No_Iter vs. Punctual_Iter ($p=.098$) and between Punctual_Iter vs. Durative_Iter ($p=.362$). The pattern {No_Iter > Punctual_Iter > Durative_Iter > Gapped_Iter} suggests that the scores decreased as a function of meaning uncertainty in the absence of larger context: the harder the length of the subintervals can be determined, the lower the perceived naturalness of the sentence (Figure 1).

3. Experiment 2: self-paced reading

To examine the real-time comprehension of underspecified meaning, we conducted a self-paced reading with the moving-window paradigm. In addition, we recorded the participants’ Autism Spectrum Quotient (AQ) to investigate the impact of individual processing styles on sentence comprehension.

3.1 Method

3.1.1. Materials

We adopted the best-rated 40 sets of the four critical conditions, along with the 80 fillers, examined in the rating questionnaire, amounting to 240 sentences in total. Of the critical sets, half of the critical verbs were intransitive (e.g., ジムで跳んだ “jumped at the gym”), while the other half were transitive, for which the locative adverbial phrases in between the for-adverbial (~間) and the critical verb in the intransitive configuration were replaced by object noun phrases (e.g., 椅子を蹴った “kicked the chair”). Note that the availability of underspecified iterative meaning is
independent of verb transitivity. We pseudo-randomized the stimuli by creating four master scripts, splitting the sentences into 40 blocks; each block contained items of distinct conditions and distinct item-set index. We then used these four master scripts to generate distinctive scripts, randomizing both the block order and the item order within each block for each participant.

3.1.2. Participants

Forty-nine participants were recruited from Waseda University in Japan, aged 18–30 (mean age = 23.38) and without reading disability by self-report. Written consent was obtained from the participants prior to the experiment.

3.1.3. Procedures

The experiment was implemented using the E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The participants were instructed to read each sentence segment-by-segment at their own pace on a computer screen. Every trial began with a dot appearing on the left-edge of the sentence to signal the starting point of the sentence, along with a series of dashes corresponding to the number of the characters of the sentence. The participants pressed the spacebar to start, causing the first segment to show up; with subsequent pressing, the next segment appeared and the previous segment returned to dashes. For 45% of the trials, the participants received a Yes/No comprehension question probing the content of the sentence (unrelated to event iterativity) after reading. The experiment consisted of four sessions (60 trials per session) with breaks in between. A practice session was given prior to the real sessions. After the self-paced reading task, the participants took the AQ.

3.1.4. Data analysis

One participant was identified as an outlier, whose data was excluded; the data from the remaining 48 participants (24 females) were analyzed. The accuracy of the comprehension questions achieved 89.95% in average. Following previous studies, items with comprehension questions that were incorrectly answered were removed, and the reading times larger than 2.5 SDs from the mean per region were excluded from analysis (6.81% data points in total) (cf. 石井 & 石川 2014; 龍 et al. 2010, Ryo 2015; Zarcone 2017). We analyzed the data at four regions of interest, Verb, V+1, V+2, sentence-final (S-final), using a linear mixed-effect model analysis with the lme4 package (Bates 2015), carried out in R (R Core Team 2017). Also, the participants were categorized into two groups according to their AQ scores (high AQ group mean= 29.29 vs. low AQ group mean = 17.46). The models on the RTs incorporated Verb Type, Interval, AQ group, as well as verb transitivity as the fixed factors. All models incorporated random intercepts for subject and item-set, given that the models with more complicated random-effect structures failed to converge. Effects were assessed by model comparisons, contrasting the model with the effect in question against a model without it.

3.2 Results

A summary of RTs by condition was presented in Table 3, visualized in Figure 2. At the critical verb region, a main effect of verb transitivity was found (intransitive > transitive; \( \chi^2(1) = 4.638, p< .032 \)). The V+1 region showed a significant main effect of Interval (long > short; \( \chi^2(1) = 5.67, p=0.017 \)). Also, a significant main effect of AQ group (high AQ > low AQ) appeared at the S-final region (\( \chi^2(1) = 7.78, p=.005 \)). In addition, a significant interaction between Verb Type and Interval emerged with intransitive verbs in the high-AQ group (\( \chi^2(1) = 4.60, p<.032 \)). Furthermore, contrast analyses between No_Iter vs. {Punctual_Iter, Durative_Iter, Gapped_Iter} revealed that No_Iter engendered significantly shorter RTs than the latter three conditions which involve underspecified iterative meaning (\( t = -2.43, p=.015 \)). No effect of Verb Type was found at any region.

<table>
<thead>
<tr>
<th>Region</th>
<th>Condition</th>
<th>Mean RT</th>
<th>SD</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seg.4</td>
<td>No_Iter</td>
<td>616.42</td>
<td>327.34</td>
<td>7.75</td>
</tr>
<tr>
<td></td>
<td>Punctual_Iter</td>
<td>610.08</td>
<td>330.55</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>Durative_Iter</td>
<td>613.96</td>
<td>316.60</td>
<td>7.56</td>
</tr>
<tr>
<td></td>
<td>Gapped_Iter</td>
<td>622.31</td>
<td>335.16</td>
<td>7.98</td>
</tr>
<tr>
<td>Seg.5</td>
<td>No_Iter</td>
<td>515.29</td>
<td>225.77</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>Punctual_Iter</td>
<td>524.17</td>
<td>234.66</td>
<td>5.49</td>
</tr>
<tr>
<td></td>
<td>Durative_Iter</td>
<td>531.86</td>
<td>237.91</td>
<td>5.66</td>
</tr>
<tr>
<td></td>
<td>Gapped_Iter</td>
<td>528.04</td>
<td>231.81</td>
<td>5.50</td>
</tr>
<tr>
<td>Seg.6</td>
<td>No_Iter</td>
<td>470.32</td>
<td>185.98</td>
<td>4.39</td>
</tr>
<tr>
<td></td>
<td>Punctual_Iter</td>
<td>466.55</td>
<td>186.52</td>
<td>4.37</td>
</tr>
<tr>
<td></td>
<td>Durative_Iter</td>
<td>467.74</td>
<td>179.50</td>
<td>4.28</td>
</tr>
<tr>
<td></td>
<td>Gapped_Iter</td>
<td>470.05</td>
<td>182.21</td>
<td>4.32</td>
</tr>
<tr>
<td>Seg.7</td>
<td>No_Iter</td>
<td>430.82</td>
<td>176.40</td>
<td>4.15</td>
</tr>
<tr>
<td></td>
<td>Punctual_Iter</td>
<td>430.73</td>
<td>171.17</td>
<td>4.00</td>
</tr>
<tr>
<td></td>
<td>Durative_Iter</td>
<td>432.03</td>
<td>176.80</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td>Gapped_Iter</td>
<td>432.13</td>
<td>172.71</td>
<td>4.09</td>
</tr>
</tbody>
</table>
4. Discussion

Our results show that sentences with underspecified meaning, while being acceptable, received lower rating scores than the transparent counterparts (Exp.1). We interpret the decrease in rating scores as reflecting the degree of meaning uncertainty and the effort of determining the appropriate partition measure in the absence of context. In particular, sentences of the Gapped_Iter condition such as "The athlete jumped for 2 months at the gym" were rated the lowest among the four conditions. While such sentences are deemed sensible, it is harder to pin down the exact length of the subintervals of the interval denoted by "for 2 months" based on the conventionalized duration of the event denoted by the verb (jump), relative to other iterative counterparts, namely "jumped for 20 minutes" (Punctual_Iter) or "jogged for 2 months" (Durative_Iter). The pattern (No_Iter > Punctual_Iter > Durative_Iter > Gapped_Iter) in rating suggests that the deeper the meaning search is required, the less natural it appears to comprehenders (Lai & Piñango 2017). It should be noted that naturalness rating is not associated with online reading times directly, as the above pattern did not emerge in Exp.2.

Results of the self-paced reading experiment (Exp.2) show that sentences with underspecified iterative meaning, irrespective of verb type, required greater processing effort than those without during real-time comprehension. The absence of the Verb Type effect undermines the ITER-operator account, which claims that the composition of a punctual verb and a durative for-adverbial (irrespective of the interval length) constitutes an aspecual mismatch that is repaired by a meaning-shift strategy during comprehension. On the other hand, the findings are more consistent with the Partition-Measure account, in the sense that the cost was modulated by the lengths of the intervals denoted by for-adverbials and not by verb type. When the interval is too long such that the infinitesimal partition measure (continuous reading) is inapplicable according to the lexicalized duration of the event denoted by the predicate, the non-infinitesimal partition measure is applied. The latter requires a costly search in the lexical conceptual structure of the predicate and the sentential/discourse context to determining the lengths of the subintervals of the interval denoted by for-adverbials. This is further supported by the result that No_Iter on one hand engendered shorter RTs than the three conditions involving underspecified meaning on the other hand (Punctual_Iter, Durative_Iter, Gapped_Iter) in the contrast analysis.

Nevertheless, the Partition-Measure hypothesis does not capture our findings in full. In addition to the main effect of Interval, it also predicts an interaction between Interval and Verb Type, which was obtained only with intransitive verbs in the high AQ group at the V+1 region. We reason that this results from the less frequent configuration with intransitive verbs and the weaker capacity of high-AQ individuals to determine a proper meaning in real-time; under these circumstances the processing source is more constrained for computing the underspecified meaning, thereby highlighting the effect. The effect of verb transitivity at the Verb region in Exp.2 was not predicted by any account. We suspect that the longer RTs induced by intransitive verbs as compared to transitive verbs might result from the configuration frequency that [subjectNP + locativeAdv + Verb] is less frequent than [subjectNP + objectNP + Verb] in Japanese. A finer examination is needed to disentangle this issue. Crucially, the processing profiles varied with individuals’ autistic traits: individuals with higher AQ scores spent significantly longer times than those with lower AQ scores at the sentence-final region. This is taken to reflect a propensity for deliberation (Brosnan et al. 2016; De Martino et al. 2008) or a difficulty of information integration (Minshew & Goldstein 1998) in individuals with higher autistic traits. Taken together, verb transitivity that is perhaps associated with language-specific...
components and individual processing styles play a role in meaning comprehension as well, though these have not been taken into consideration in previous studies on the processing of underspecified meaning.

5. Conclusion

We investigate the cognitive mechanisms and individual factors that underlie the comprehension of underspecified iterative meaning. Results show that such meaning processing requires more effort to mine the lexical conceptual structure (i.e. the conventionalized duration denoted by the predicate) and sentential/discourse context to determine an appropriate partition measure for the interval denoted by for-adverbials. Also, individuals with higher autistic traits require additional effort during comprehension, likely reflecting circumspect deliberation in decision-making or effortful information integration.

6. Acknowledgement

This research was supported by JSPS Kakenhi Kiban (A) (15H01881; PI: Hiromu Sakai).

7. Reference


Fukumoto, S. et al. (2014). Implicit and explicit categorization of disyllabic and trisyllabic acoustic Along with the prediction accuracy and rotation angle, we also present the number of steps for the group of participants who showed higher accuracy for the visual and auditory conditions. *Journal of CogSci* 28(42), 313-331.


