

# Comprehension of underspecified iterative meaning in Japanese

Yao-Ying Lai<sup>†</sup>, Michiru Makuuchi<sup>†</sup>, Maria Mercedes Piñango<sup>‡</sup> and Sakai Hiromu<sup>†</sup>

<sup>†</sup>Research Institute, National rehabilitation Center for Persons with Disabilities  
1, Namiki 4-chome, Tokorozawa City, Saitama Prefecture 359-8555, Japan

<sup>‡</sup>Department of Linguistics, Yale University  
370 Temple St. New Haven, CT 06520-8366, U.S.A.

<sup>†</sup>Waseda University  
3-4-1 Okubo, Shinjuku City, Tokyo, 169-8555, Japan

E-mail: [†yaoyinglai@gmail.com](mailto:†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., 石井 & 石川 2014, 龍 et al. 2010; Paczynski et al. 2014, Piñango et al. 1999, Todorova et al. 2000).

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

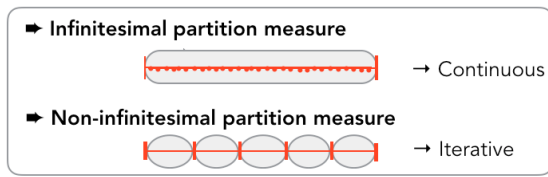
Based on these observations, we ask the following questions: (a) how do comprehenders obtain the underspecified iterative meaning, and (b) what is the source of the cost associated with processing sentences like (1)? By addressing these questions, this study aims to uncover the semantic-conceptual components 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 & Pylkkänen 2008, Paczynski et al. 2014; Piñango & Zurif 2001). However, this account is called into question by cases like (3), in which the composition of a durative verb (e.g., *jog*) and the durative *for*-adverbial involves no mismatch yet an iterative reading arises.

(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 **Interval** (short/long) denoted by *for*-adverbials (~間). Another 80 fillers were introduced, half of them being nonsensical to check

**Table 1:** Conditions & Sample sentences

Condition		<i>for</i> -adverbial		Seg.4 (Verb)	Seg.5 (V +1)	Seg.6 (V +2)	Seg.7 (S-final)
(A) No_Iter Durative verb + Short interval	選手が athlete-NOM	20分間 20 min- <i>for</i>	体育館で Gym-LOC	ジョギングしたと jog-PAST-COMP	学生は student-TOP	コーチに coach-DAT	言った。 say-PAST
	“The student said to the coach that the athlete <b>jogged for 20 minutes</b> at the gym.”						
(B) Punctual_Iter Punctual verb + Short interval	-----	20分間 20 min- <i>for</i>	-----	ジャンプしたと jump-PAST-COMP	-----	-----	-----
	“The student said to the coach that the athlete <b>jumped for 20 minutes</b> at the gym.”						
(C) Durative_Iter Durative verb + Long interval	-----	2ヶ月間 2 month- <i>for</i>	-----	ジョギングしたと jog-PAST-COMP	-----	-----	-----
	“The student said to the coach that the athlete <b>jogged for 2 months</b> at the gym.”						
(D) Gapped_Iter Punctual verb + Long interval	-----	2ヶ月間 2 month- <i>for</i>	-----	ジャンプしたと jump-PAST-COMP	-----	-----	-----
	“The student said to the coach that the athlete <b>jumped for 2 months</b> at the gym.”						

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).

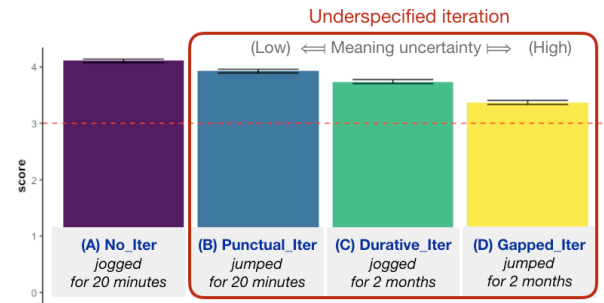
## 2.2. Results

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).

**Table 2:** Results of Naturalness rating

	Condition	Mean	SD	se
A	No_Iter	4.11	1.04	0.03
B	Punctual_Iter	3.93	1.08	0.03
C	Durative_Iter	3.74	1.17	0.04
D	Gapped_Iter	3.37	1.22	0.04



**Figure 1:** Naturalness rating per condition

## 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.

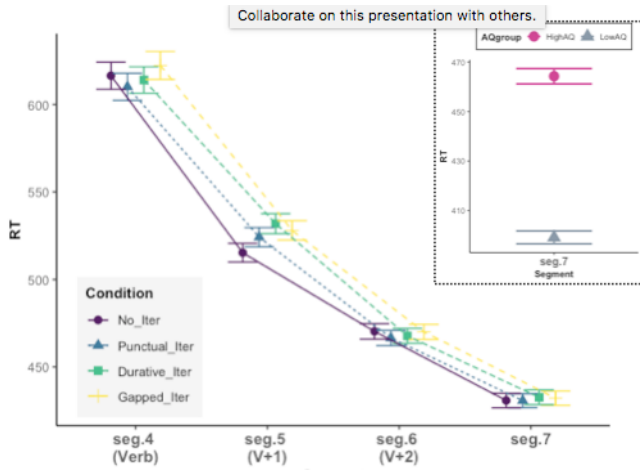
## 3.2 Results

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 3:** Summary of RTs (Exp.2)

Region	Condition	Mean RT	SD	se
Seg.4 (Verb)	No_Iter	616.42	327.34	7.75
	Punctual_Iter	610.08	330.55	7.76
	Durative_Iter	613.96	316.60	7.56
	Gapped_Iter	622.31	335.16	7.98
Seg.5 (V+1)	No_Iter	515.29	225.77	5.34
	Punctual_Iter	524.17	234.66	5.49
	Durative_Iter	531.86	237.91	5.66
	Gapped_Iter	528.04	231.81	5.50
Seg.6 (V+2)	No_Iter	470.32	185.98	4.39
	Punctual_Iter	466.55	186.52	4.37
	Durative_Iter	467.74	179.50	4.28
	Gapped_Iter	470.05	182.21	4.32
Seg.7 (S-final)	No_Iter	430.82	176.40	4.15
	Punctual_Iter	430.73	171.17	4.00
	Durative_Iter	432.63	176.80	4.22
	Gapped_Iter	432.13	172.71	4.09

**Figure 2: Results of Exp.2**



#### 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 aspectual

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

- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders*, 31(1), 5-17.
- Bates, D. et al. (2015). Package ‘lme4’. *Convergence*, 12(1).
- Brennan, J., & Pykkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and language*, 106(2), 132-143.
- Brosnan, M., Lewton, M., & Ashwin, C. (2016). Reasoning on the autism spectrum: a dual process theory account. *Journal of autism and developmental disorders*, 46(6), 2115-2125.
- Caruana, N., & Brock, J. (2014). No association between autistic traits and contextual influences on eye-movements during reading. *PeerJ*, 2, e466.
- De Martino, Benedetto, et al. (2008). Explaining enhanced logical consistency during decision making in autism. *Journal of Neuroscience* 28(42), 10746-10750.
- Deo, A. (2009). Unifying the imperfective and the progressive: partitions as quantificational domains. *Linguistics and philosophy*, 32(5), 475-521.
- Deo, A. S., & Piñango, M. M. (2011). Quantification and context in measure adverbs. In *Semantics and Linguistic Theory* (Vol. 21, pp. 295-312).
- Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, 25(3), 348-368.
- 石井創 & 石川潔. (2014). 意味カテゴリーは離散か連続か—アスペクトの実時間処理に基づく検討. *JCSS Japanese Cognitive Science Society*, 318-324.
- Jackendoff, R. (1997). *The architecture of the language faculty* (No. 28). MIT Press.
- Lai, Y.-Y., Piñango, M. M., (2017). Comprehending underspecified meaning: resolving ambiguity by contextual and conceptual search. In: *Conference handbook of the 19th Annual International conference of the Japanese Society for Language Sciences*. pp. 134-137.
- 龍盛艶, 小野創, & 酒井弘. (2010). 文理解の過程における事象タイプの認知—日本語のアスペクト情報処理を手がかりに—. *認知科学*, 17(2), 313-331.
- Minshew, N. J., & Goldstein, G. (1998). Autism as a disorder of complex information processing. *Mental Retardation and Developmental Disabilities Research Reviews*, 4(2), 129-136.
- Ryo, O. B. A. (2015). A Delayed Effect in Aspectual Mismatches and Processing Cost of Aspectual Adverbs.
- Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience*, 26(9), 1905-1917.
- Piñango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax- semantics interface. *Journal of Psycholinguistic Research*, 28(4), 395-414.
- Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from <https://www.pstnet.com>.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schütze, C. T., & Sprouse, J. (2014). Judgment data. *Research methods in linguistics*, 27.
- Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 22, No. 22).
- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, 53(1), 1-25.
- 若林明雄, & 東條吉邦. (2004). 自閉症スペクトラム指数 (AQ) 日本語版の標準化. *心理学研究*, 75(1), 78-84.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.
- Yu, ACL (2010). Perceptual Compensation Is Correlated with Individuals' Autistic Traits: Implications for Models of Sound Change. *PLoS ONE*, 5(8). doi: 10.1371/journal.pone.0011
- Zarcone, A., McRae, K., Lenci, A., & Padó, S. (2017). Complement coercion: The joint effects of type and typicality. *Frontiers in psychology*, 8, 1987.
- Zhang, M., Piñango, M. M., & Deo, A. (2018). Real-time roots of meaning change: Electrophysiology reveals the contextual-modulation processing basis of synchronic variation in the location-possession domain. In *CogSci*.