Studia Linguistica (『ことばの科学』), Volume 28, 2014

Task effects on sentence processing using eye-tracking

MANSBRIDGE Michael [1]
HAYAKAWA Kyoko [2]
TAMAOKA Katsuo [3]

Abstract: In sentence processing literature, researchers have generally overlooked how the secondary task they utilize affects obtained data. What has been the yardstick for experimental design up to now is to make the task not too easy and not too difficult, as to ensure participants read and process the sentence without any unnecessary task demands. Currently, in sentence processing studies, there are two general experimental techniques, verification and plausibility, where verification uses comprehension questions to make sure participants understand the sentence, and plausibility uses a sentential decision task to judge if the sentence can exist in the real world. This study compared the processing strategies of Japanese sentences from two experimental tasks, a verification task and a plausibility task, to demonstrate how methodology is a factor in sentence processing. The results indicate that the verification task not only required increased cognitive workload but also had a different processing strategy compared to the plausibility task. The processing in the verification task focused on the disambiguation of thematic roles whereas the plausibility task was sensitive to animacy inconsistencies. However, despite each experimental methodology having a task-dependent effect, it was observed that syntactic effects are task-independent.

Keywords: task-effects, L2 processing, eye-tracking, verification, plausibility

[1] MANSBRIDGE, Michael: Research Student at the Graduate School of Languages and Cultures, Nagoya University, Japan. E-mail: ktamaoka@lang.nagoya-u.ac.jp
[2] HAYAKAWA, Kyoko: Assistant Professor at the Graduate School of Languages and Cultures, Nagoya University, Japan. E-mail: ktamaoka@lang.nagoya-u.ac.jp
[3] TAMAOKA, Katsuo: Professor at the Graduate School of Languages and Cultures, Nagoya University, Japan. E-mail: ktamaoka@lang.nagoya-u.ac.jp

1. Introduction

Understanding how the cognitive mechanisms of experimental methodology interact with language processing is crucial knowledge for linguistic researchers. In psycholinguistic research, there are a countless number of different tasks that can be used to study a vast array of linguistic phenomena. Sentence processing studies overwhelmingly involve more than simply just reading sentences; accordingly, secondary tasks are given which focus on a certain aspect of the sentence that cannot be accomplished without the reading of the sentence. The two secondary tasks that will be discussed in this study are *verification* tasks and *plausibility* tasks which are used to investigate if sentence processing is susceptible to change as a function of secondary task.

Verification tasks employ post-sentence comprehension probes that questions specific information on the previously read sentence. After reading a stimulus sentence, the sentence would disappear and be replaced with the probe which can be regulated by display times or a feedback device. The participant would then proceed onto answering the question using a feedback device such as a gamepad or keyboard. Example: [The guard who the prisoner pushed was punished severely. Positive probe: Was the guard punished? "Yes or No". False probe: Was the prisoner punished? "Yes or No".] In this example, the positive probe is accurate; the guard was punished, so the correct response would be yes. For the false probe, on the other hand, since we cannot extrapolate that the prisoner was punished within the boundary of the sentence, the answer would be no.

Contrary to verification, plausibility judgments are a sentential decision on whether the sentence read can exist in the real world for the given language without a paranormal interpretation; this task can have the decision to be made during or after the display of the sentence. Examples: [Plausible: The thief shot John with the shotgun inside the store. Implausible: The thief shot John with the banana inside the store.] For the first sentence, the sentence is clearly plausible, as we can easily imagine thieves shooting people with guns; accordingly the participant would indicate on a feedback device using buttons that correspond to "Yes" or "True" after finishing reading the sentence. In comparison, the implausible sentence becomes an impossible sentence at the site of "banana" as it not a permissible semantic instrument of the verb shoot. Therefore, the participant would have to select the "No" or "False" button on the feedback device. Plausibility tasks can serve either of two functions. First, it can be used to compare the processing of plausible stimuli against implausible stimuli as in the example above, or it can be used as means to distract participants from the purpose of the experiment by only using implausible items as fillers/distractors.

Prior studies on the use of verification and plausibility answering strategies have shown that the process in which to verify an answer takes longer than to make an inference judgment upon it (Reder, 1982; Reder & Ross, 1983; Reder & Wible, 1984). The reasoning behind this response time discrepancy has been explained by the *fan effect* (Anderson, 1974) which is the activation of a proposition network that concurrently activates linked nodes with related concepts and spreads out across the network field. Consequently, the activation of the network increases as more knowledge is obtained on the original proposition and its linked nodes. The reasoning that this causes a delay in response times is that verification strategies over-activate the network which indicates that more processing resources are being used than what is required. In contrast, plausibility judgments are not as susceptible to activating linked concepts and thus activates only what is required by the task. Another coinciding effect taken from Gordon, Hendrick, and Johnson (2001) shows a *similarity-based interference effect* during comprehension when using multiple nouns of the same noun class type (e.g., general noun, pronoun, or name). Consequently, during the processing of the sentence both nouns are being activated with similar semantics which causes ambiguity for theta-role assignment. This effect taken with the fan effect suggests that during comprehension people use extra cognitive resources for fact-retrieval.

Recently, Caplan, Chen, and Waters (2008) investigated the task-dependent and task-independent effects of verification, plausibility, and non-word detection using fMRI brain imagining. For both verification and plausibility it was shown that more complex object-extracted relative clauses (ORC) were more difficult to process than subject-extracted relative clauses (SRC). From this, they infer the area responsible for syntactic processing is shared and closely interacts with semantics during theta-assignment at the left posterior inferior frontal gyrus. However, in their verification task it was argued that additional unique brain regions being activated were attributed to other possible cognitive functions such as phonological storage and rehearsal of the more complex sentence type to help disambiguate between similar theta-role arguments. Though their plausibility tasks had increased brain activation in comparison to the non-word detection task, Caplan et al. (2008) do not suspect these activations to be linked with cognitive functions like in the verification task but instead speculate they are related to plausibility of the theta-assignment in relation to the syntactic structure. The findings in Caplan et al. (2008) are intriguing in that the results implicate that verification involves more cognitive work than plausibility judgments. The study, however, does not make a comparison on reading times between experiments.

Another study by Leeser, Brandl, and Weissglass (2011) compared the effects of a verification task and a grammaticality judgment task using self-pace reading (c.f., Just, Carpenter & Wooley, 1982) on second language (L2) leaners of Spanish. In this study, the authors investigated the sensitivity of grammatical violations in Spanish; more specifically, gender agreement and subject-verb inversion. For noun-adjective gender agreement, only the grammaticality task demonstrated effects of gender agreement violations. This was indicated by longer response times at the adjective for the ungrammatical condition and shorter response times in the following spillover regions for the same condition. For subject-verb inversion, though within the grammaticality judgment task there were longer response times for the ungrammatical condition, no main effects of grammatically were found. The verification task did not demonstrate sensitivity to violations for either violation conditions. One interesting result was that the grammaticality judgment task had consistently longer response times than the verification strategy which highlights the fact different reading mechanisms are involved in each task. They contend that the sensitivity to violations found during grammaticality judgments is an artifact of the task, and learners are relying on explicit knowledge of grammar rather than drawing upon the implicit knowledge structure involved during comprehension.

In the present study, we use eye-tracking methodology to investigate the processing of a verification task against a plausibility task in order to investigate the differences of cognitive workload on sentence processing in a second language with a purpose of determining task-dependent and –independent effects involve for each task. The Leeser et al. (2011) study was able to demonstrate behavior responses are longer for grammaticality judgment tasks over verification strategies, but the Caplan et al. (2008) study makes no such claim between plausibility and verification strategies. As a result of using eye-tracking, we are able to measure the differences between the tasks through time course processing stages which elucidates to a general pattern of processing within the sentence (c.f., Clifton, Staub, & Rayner, 2007; Staub & Rayner, 2007). The current study shares more similarities to the Leeser et al. (2011) study over Caplan et al. (2008) as we used second language speakers of Japanese rather than native speakers. The sentence types were also different; instead of using ORCs and SRCs to demonstrate processing difficulty associated with syntactic structure, we used a difference in syntactic voicing with active and passive voiced verbs (c.f., Gordon & Chan, 1995; Gorin, 2005). Additionally, we only analyzed semantically plausible sentences. The choice to not analyze implausible items was due to implausible items not existing in the verification task; the main interest is the difference between real world sentences. If the propositions made by Caplan et al. (2008) are accurate, we not only expect to see extra processing costs (i.e., increased reading time) associated with the verification task, but also a clear difference in processing strategy between the two tasks.

2. Experiment

2.1 Participants

Forty native speakers of Chinese and Korean were recruited for this study (Chinese = 22 & Korean = 18). A Japanese grammar and vocabulary test was given to confirm that the participants had a sufficient level of proficiency. It was determined that the mean average of the grammar test was 89.93% while the vocabulary test had a mean average of 79.48%. No participant was eliminated for poor Japanese proficiency.

2.2 Materials

All experimental items were designed as three word canonical SOV sentences. There were only four comparable conditions between the two tasks. (1) An active voiced sentence with inanimate agent and an animate patient (active-IA): [$_{S}$ R1 *kaze-ga* [$_{VP}$ R2 *Junko-o* R3 *toba-shita*]] ‘The wind pushed Junko’. (2) The passive counterpart (passive-AI): [$_{S}$ R1 *Junko-ga* [$_{VP}$ R2 *kaze-ni* R3 *toba-sare-ta*]] ‘Junko was pushed by the wind’. (3) An active sentence with an animate agent and patient (active-AA): [$_{S}$ R1 *Taro-ga* [$_{VP}$ R2 *Junko-o* R3 *ket-ta*] ‘Taro kicked

Junko’. (4) And the passive counterpart: [$_S$ R1 *Junko-ga* [$_{VP}$ R2 *Taro-ni* R3 *ker-are-ta*]] ‘Junko was kicked by Taro’. For each experimental session there were 24 experimental items used with six of each of the four conditions present for both tasks. Sentences were displayed normally with a mixture of hiragana and kanji script. For the verification task, each sentence was followed by either a positive or false probe. Positive probe: *Taro-ga Junko-o kerimashitaka* ‘Did Taro kick Junko’. False probe: *Junko-ga Taro-o kerimashitaka* ‘Did Junko kick Taro’.

2.3 Procedure

Each participant was seated in front of a computer monitor in a quiet room. Prior to the beginning of each experimental session, the participant’s right eye was calibrated to the eye-tracking camera using a 9-point calibration and validation method (desktop Eyelink 1000 SR Research Ltd., Ontario, Canada). Before the display of an experimental trial, a drift correcting mask was presented on the far left center of the screen indicated by the circle “◎”. Once a participant accurately fixated on the mask, the experimenter then accepted the fixation to allow the presentation of a trial item that replaced the mask. Each stimulus item had a maximum display of eight seconds. However, each task differed in experimental procedure during the presentation of the stimulus item. For the verification task, participants were instructed to read the sentence at their natural pace, and once finished reading and comprehending the sentence, they were instructed to press any button on the gamepad to replace the sentence with a verification question that targeted the assignment of thematic roles. For the plausibility task, participants were instructed to read the sentence at their natural pace, and decide whether or not the sentence as a whole was semantically plausible by pressing a corresponding button on the gamepad. In both tasks, feedback indicated by ‘Correct’ or ‘Incorrect’ in Japanese was given after they selected their response in the form of a message that was displayed in the middle of a new screen.

2.4 Analyses

Prior to the analyses, trials with tracker loss were removed from the data set. In addition, trials that lacked participant feedback were also removed from the data set. Subsequently, all fixations below 80 ms were merged into a neighboring fixation within a one character distance.

This was then followed by the removal of all remaining fixations under 80 ms and also fixations that were over 1000 ms. To analyze each part of the sentence, the sentence was divided into three aforementioned interest regions to capture reading times and eye-movements at each part of the sentence. Region 1 is always the subject of the sentence, region 2 is the object for active sentences and the agent for passive sentences, and region 3 is always the verb.

A series of linear mixed effect (LME) modeling (Baayen, Davidson, & Bates, 2008) analyses were then conducted on the collected reading times and eye-movements using SPSS; subject, items, and L1 groups as random factors. Via eye-tracking, we are able to report the reading times (RT) for each task within each part of the sentence. Accordingly, we are able to compare the tasks through different stages of processing. The earliest processing measure we will report is the first-fixation duration which refers to the first-fixation RT at a specified interest region within the sentence. This is followed by the first-pass reading which is composed of all fixations made within an interest region from when it is first entered from the left until it is exited in either direction. This measure is also an indication of early processing but is nevertheless a later processing measure compared with the first-fixation duration data as a given word or phrase may have multiple fixations during its first reading. The late processing measure we will report in this study is dwell-time which refers to the summation of all fixations that fall within a region. Additionally, the regression-out percentages will be reported which refers to the likelihood of making an eye-movement out of an interest region into a previous part of the sentence. However, only the second and third interest regions will be reported as it is not possible to regress-out of the first region. The total reading time of the sentence and the accuracy percentages will also be reported to analyze sentence as a whole. In the reporting of the results for this experiment, analyses done within tasks will only be reported if there is first an interaction of the tasks and conditions. Multiple comparison analyses were done using Bonferroni adjustments.

2.5 Results

Please see Tables 1-3 for the means and standard errors of the RTs, regressions, and accuracies for the task and task conditions. Reading times are in milliseconds.

2.5.1 Total reading times of sentences and accuracies

In the total reading time of the sentence, a significant main effect of task was found, [$F(1, 43) = 115.31$, $p < .001$], which showed that the RTs for the verification task ($M = 3{,}481$ ms, $SE = 56$ ms) were significantly longer than the plausibility task ($M = 2{,}256$ ms, $SE = 32$ ms). In addition to the main effect, an interaction of task:condition was found [$F(7, 399) = 30.68$, $p < .001$]. Thus, each task was then separately analyzed. Within the verification task, a robust effect of sentence condition was found at the sentence level [$F(3, 736) = 15.89$, $p < .001$]. This effect revealed through multiple comparisons that the passive-AA condition (4) had significantly longer RTs compared to the other three conditions (4 > 1, 2, 3); additionally, the active-AA condition (3) had significantly longer RTs in comparison to the active-IA (1) condition (3 > 1). Similarly, the plausibility task had a significant effect of condition type, [$F(3, 725) = 20.20$, $p < .001$]. The effect, however, was seemingly reversed; the active-IA (1) condition had significantly longer RTs than the other three conditions (1 > 2, 3, 4). The passive-AI condition (2) also had longer RTs than the active-AA (3) condition (2 > 3). For accuracy percentages, the main effect of task was not significant, [$F(1, 1813) = 0.34$, $p = .561$, ns], showing that the overall mean accuracies of the two tasks were analogous (V: $M = 82.94\%$, $SE = 1.23$; P: $M = 83.99\%$, $SE = 1.21$). However, since a significant interaction of task:condition was revealed, [$F(7, 1786) = 11.89$, $p < .001$], each task was then separately analyzed. The verification task produced an effect of condition, [$F(3, 922) = 7.87$, $p < .001$], which through multiple comparisons demonstrated that the passive-AA (4) was significantly less accurate than all other conditions (4 < 1, 2, 3). Likewise, the plausibility task had an even more robust effect, [$F(3, 891) = 19.81$, $p < .001$]. In contrast to the verification task, the plausibility task produced a different pattern of results in which the active-IA condition exhibited lower accuracies in comparison to both active-AA and passive-AA conditions (1 < 3, 4). Additionally, the passive-AI had significantly lower accuracy than the active-AA and passive-AA conditions (2 < 3, 4).

2.5.2 Subject (R1)

At the subject, there was an immediate main effect of task for first-fixation durations, [$F(1, 42) = 54.86$, $p < .001$], revealing that at the earliest possible processing measure the verification task was reliably longer than the plausibility task. Subsequently, we then tested for an interaction

Table 1. Total reading times and accuracy ratings for task conditions

	Total Reading Time		Accuracy	
	Means	SE	Means	SE
Verification				
1	3,187	97	86.86	2.20
2	3,381	112	83.12	2.44
3	3,520	109	87.98	2.13
4	3,900	127	73.71	2.90
Plausibility				
1	2,573	83	71.49	3.00
2	2,312	64	79.91	2.68
3	2,063	54	89.13	2.06
4	2,156	53	95.22	1.41

between the tasks and conditions which the analysis revealed to be significant, [$F(7, 1549) = 8.35$, $p < .001$]. However, it is unlikely that voice effects would be apparent already as the first true indication of syntactic-voice appears at the verb morphology in the third region. Analyses were then conducted on both tasks, but neither the verification task, [$F(3, 737) = 1.03$, $p = .378$, ns] nor the plausibility task [$F(3, 727) = 0.62$, $p = .605$, *ns*], produced a significant effect demonstrating that there were differences between animacy and voice for the four conditions. Given these points, it is apparent that for first-fixation durations a pattern of results cannot to be seen; as such, there is only a reading time difference between the two tasks at the earliest processing measure.

Moving onto first-pass RT, a main effect was still observed, [$F(1, 41) = 23.45$, $p < .01$]. Analogous to the previous measure, the verification task produced longer reading times than the plausibility task. Again, the follow-up analysis revealed that there was an interaction, [$F (7, 430) = 8.44$, $p < .001$]; thus, further analyses were then conducted on each task. In the verification task, there were still no effects found, [$F(3, 775) = 0.07$, $p = .974$, *ns*]. The plausibility task, on the other hand, did have a robust effect, [$F(3, 724) = 16.05$, $p < .001$]. This analysis revealed that the RTs for the active-IA condition were longer than all other conditions (1 > 2, 3, 4).

Table 2. Reading times and regressions for tasks

	First-Fixation		First-Pass		Dwell-Time		Regression-Out	
	Means	SE	Means	SE	Means	SE	Means	SE
Region 1								
Verification	239	3	519	12	1,254	26	-	-
Plausibility	209	2	426	9	775	16	-	-
Region 2								
Verification	244	4	367	7	1,086	25	22.81	1.51
Plausibility	243	3	394	8	721	16	10.48	1.11
Region 3								
Verification	301	5	804	20	1,161	27	95.44	0.75
Plausibility	312	5	630	13	782	16	85.73	1.28

Altogether, at the completion of the reading of the subject, the plausibility task, unlike the verification task, developed a processing pattern.

At later processing measures, i.e. dwell-time, it was shown that there was no attenuation of the main effect for task [$F(1, 42) = 85.75$, $p < .001$]; the difference in reading time had only increased through processing stages. The succeeding analysis for interaction of task:condition was still significant, [$F(7, 428) = 20.01$, $p < .001$]. However, at this processing stage the verification task was able to produce a significant effect of condition type, [$F(3, 736) = 5.77$, $p < .01$], revealing that the RTs for the active-IA condition were shorter than the active-AA and passive-AA conditions (1 < 3, 4). Comparable to the verification task, the plausibility task also had a significant effect, [$F(3, 724) = 16.77$, $p < .001$]; yet, akin to the first-pass reading measure, the active-IA continued to be significantly longer than other conditions (1 > 2, 3, 4), which is in direct contrast to the result found in the verification task.

Overall, it is shown that at the subject the verification task induced reading times throughout all measures with task-dependent results showing up earlier in the plausibility task.

2.5.3 Object/Agent (R2)

For the earliest processing measure (i.e., first-fixation duration), there was no main effect of task, [*F*(1,

42) = 0.51, *p* = .480, *ns*], yet the subsequent analysis on interaction between tasks revealed that a task:condition interaction was nonetheless present, [*F*(7, 417) = 2.34, *p* < .05]. When analyzing the tasks separately it was observed that the verification task did not produced an effect, [*F*(3, 565) = 0.37, *p* = .772, *ns*], but the plausibility task did have a significant effect of sentence condition, [*F*(3, 724) = 5.51, *p* < .01]. The analysis from multiple comparisons revealed that the first fixation times on the agent within the passive-AA condition were significantly shorter than the passive-AI agent and the active-IA object (4 < 1, 2). Additionally, the active-AA object was read faster than the passive-AI agent (3 < 2). Despite the tasks not having a difference between overall means, the plausibility task was able to produce effects between conditions.

At the first-pass reading measure, there was no main effect, [*F*(1, 46) = 1.91, *p* = .174, *ns*]; and like before, an interaction effect was observed, [*F*(7, 446) = 4.38, *p* < .001]. Within the verification task a main effect was not seen, [*F*(3, 736) = 1.33, *p* = .265, *ns*]. The plausibility task, on the other hand, had a main effect, [*F*(3, 724) = 8.11, *p* < .001], demonstrating that the passive-AA condition was faster than the active-IA and passive-AI conditions (4 < 1, 2).

At the late processing measures, there was an overall main effect of task found for dwell-times, [*F* (1, 41) = 61.85, *p* < .001], which indicated that the verification task produced reliably longer reading times than the plausibility task. Moreover, an interaction was observed, [*F*(7, 431) = 19.78, *p* < .001], which allowed for further analyses on each task. The analysis for the verification task revealed that there was an effect of sentence type, [*F*(3, 734) = 13.78, *p* < .001]. This revealed that the RTs for the passive-AI condition were reliably faster than the other three conditions (2 < 1, 3, 4). Additionally, the active-IA condition was faster than the passive-AA condition (1 < 4). Incidentally, the plausibility task also had a significant effect, [*F*(3, 723) = 11.78, *p* < .001]. But in contrast to the verification task, it was shown that the RTs for the active-IA condition were significantly longer than all other conditions (1 > 2, 3, 4).

At the second region, R2, regression-out proportion data was available for analysis. Congruent with the RT measures, there was a significant main effect of task, [*F*(1, 1502) = 46.53, *p* < .001], which revealed that the verification task was approximately twice as likely to regress out from the direct object, or agent, into the subject in comparison to the plausibility task. In addition to the main effect, there was an interaction of task:condition, [*F* (7, 1456) = 4.90, *p* < .001], which allowed for further analyses within each task. Within the verification task, a main effect was found, [*F*(3, 737) = 4.05, *p* < .01], demonstrating that the passive-AI condition regressed out less than the active-AA condition (2 < 3).

Table 3. Reading times and regression percentage for the task conditions

Verification:		First-Fixation		First-Pass		Dwell-Time		Regression-Out		Plausibility:	First-Fixation		First-Pass		Dwell-Time		Regression-Out	
		Means	SE	Means	SE	Means	SE	Means	SE		Means	SE	Means	SE	Means	SE	Means	SE
Region 1																		
	1	249	7	533	27	1,134	48	-	-		205	4	523	28	952	41	-	-
	2	232	6	517	23	1,239	49	-	-		214	5	424	18	704	30	-	-
	3	242	8	519	23	1,352	52	-	-		209	4	396	12	753	27	-	-
	4	234	6	506	27	1,300	57	-	-		209	4	383	13	721	27	-	-
Region 2																		
	1	247	7	378	13	1,065	44	24.39	3.01		256	7	418	19	848	39	18.52	3.06
	2	250	8	353	13	907	44	16.92	2.69		256	7	441	19	704	29	6.78	1.89
	3	243	6	382	16	1,163	56	28.78	3.17		235	5	392	15	729	30	12.20	2.29
	4	237	7	351	16	1,223	55	20.47	3.09		231	6	341	13	632	28	5.94	1.60
Region 3																		
	1	316	10	688	29	1,008	46	96.52	1.30		313	12	560	30	798	39	91.19	2.25
	2	298	10	910	42	1,257	60	94.36	1.66		295	9	742	26	912	35	79.33	3.04
	3	295	10	718	38	1,037	47	96.00	1.39		332	10	535	24	617	27	87.63	2.37
	4	296	9	918	45	1,376	63	94.74	1.71		307	9	674	24	810	28	85.32	2.40

Within the plausibility task there was a significant effect of sentence condition, [*F*(3, 759) = 6.53, *p* < .001], demonstrating that the regression-out proportion ratio for the active-IA condition was significantly greater than both the passive-AA and passive-AI conditions (1 > 2, 4).

2.5.4 Verb (R3)

The early processing at the verb revealed that there was neither a main effect, [*F*(1, 57) = 0.83, *p* = .368, *ns*], nor was there an interaction, [*F*(7, 505) = 1.68, *p* = .113, *ns*] for the first-fixation durations. Accordingly, no further analyses were conducted.

During first-pass RT there was a significant main effect of task, [*F*(1, 85) = 74.26, *p* < .001], demonstrating that the RTs for the verification task were reliably longer than the plausibility task at this processing stage. Furthermore, the analysis showed a significant interaction, [*F*(7, 408) = 16.78, *p* < .001], which allowed for further analyses on each task. In the verification task, a significant effect of condition was observed, [*F*(3, 726) = 14.49, *p* < .001], which indicated that the RTs were significantly longer for both the passive-AI and passive-AA conditions in comparison to the active-IA and active-AA conditions (2, 4 > 1, 3). Within the plausibility task, a significant effect was also seen, [*F*(3, 711) = 17.54, *p* < .001]. Congruent with the verification task, the exact same effect was found; the passive conditions were significantly longer than the active conditions (2, 4 > 1, 3).

For the dwell-time measures (i.e., late processing) there was also a significant main effect of task, [*F*(1, 40) = 50.06, *p* < .001], indicating that the plausibility task was significantly faster than the verification task. Once again, an interaction of task:condition was found, [*F*(7, 418) = 21.59, *p* < .001]. Further analyses showed that within the verification task a main effect of condition was present, [*F*(3, 726) = 17.65, *p* < .001]. The effects found here were the same as the first-pass reading; the passive conditions were significantly longer than the active conditions (2, 4 > 1, 3). Within the plausibility task, a main effect was also found, [*F*(3, 729) = 17.75, *p* < .001]; for this task, however, it was indicated that the RTs for the active-AA condition were significantly faster than other three conditions (3 < 1, 2, 4). Additionally, it was shown that passive-AI condition had significantly longer RTs than the active-IA condition (2 > 1); it was suggested through a marginal value that the passive-AI condition was longer than the passive-AA condition (*p* = .51).

Congruent with the RT data, a significant main effect of task was found for regression-out proportion, [*F*(1, 76) = 18.05, *p* < .001], which as before revealed that the verification task was significantly more reliable to make a regression out of the verb into previous parts of the sentence. Moreover, there was an interaction found for task:condition, [$F(7, 556) = 4.79$, $p < .001$]; however, in further analysis on the verification task, [$F(3, 691) = 0.29$, $p = .831$, *ns*] revealed no effects. The plausibility task, [$F(3, 712) = 3.92$, $p < .01$], demonstrated that the active-IA condition was more likely to make a regression out of the verb than the passive-AI condition (1 > 2).

3. Discussion

The results of this study conclusively demonstrate that when participants have to use a verification answering strategy (i.e., fact retrieval) reading times will increase during the processing of the sentence. One unassuming explanation as to why reading time is a function of task is that participants are preparing for later fact retrieval. In other words, as they are reading along they are storing the argument structure of the sentence in working memory (c.f., Baddeley, 1996, 2003, MacDonald, Just, & Carpenter, 1992); and as a result, the extra memory work adds to the overall processing associated with reading the sentence. Since the reading times associated with a plausibility judgment are much quicker, it becomes evident that plausibility judgments do not require intricate short term memory input during sentence processing. This is congruent with past cognitive studies on verification versus plausibility (c.f., Reder, 1982; Reder & Ross, 1983; Reder & Wible, 1984) which report on the difference of the behavioral response times between answering strategies rather than sentence processing time; accordingly, these results provide evidence that fan effects (Anderson, 1974) are first present in the reading of the sentence that supersede the use of the answering strategy.

This study's use of plausibility judgments was dissimilar from the type of questions used in the studies conducted by Reder and colleagues (Reder, 1982; Reder & Ross, 1983; Reder & Wible, 1984), which targeted plausibility of thematically related items of studied items and items that did not allow use of fact retrieval strategies. In contrast, in this study judgments were made within the sentence on its plausibility of being a sentence in the language (c.f., Caplan et al., 2008; Tamaoka, et al., 2005; Koizumi & Tamaoka, 2010). Admittedly, the tasks between these studies are not equivalent, but they are nevertheless comparable as neither relies on direct retrieval from previous knowledge derived from the study.

But even despite the differences in the methods of the studies, it has been a major argument in this study that the difference in cognitive mechanisms of verification and plausibility also occurs during the processing of the sentence. The results clearly show that the differences between the two tasks are not only representative of a difference in overall reading time. When comparing the pattern of results of each task against each other it becomes evident that they are indeed different from one another at the majority of the reported reading time measures.

3.1 Task-independent effects

Before digressing into the differences found within in each task, it is important to first note the similarities as it demonstrates what effects are task-independent. It has been reported (Koizumi & Tamaoka, 2010; Tamaoka, Asano, Miyaoka, & Yokosawa, 2013; Caplan et al., 2008) that the syntactic parser is activated during plausibility judgments such that processing difficulty can be observed in syntactically more complex sentence types (e.g., passive vs. active or ORC vs. SRC). Accordingly, this proposition is validated by our finding that syntactic processing at the verb had processing costs associated with the more complex passive voice conditions. This supports the finding proposed by Caplan et al., (2008) that the syntactic parser is activated in both verification and plausibility tasks and is sensitive to syntactic complexities regardless of the experimental task. Granted, there were some differences associated to the voice effect that interacted with animacy in our plausibility task. But even so, both passive voice conditions resulted in longer reading times when compared to their active counterparts.

3.2 Plausibility

At the sentence level, it was shown that the verification task's most difficult condition was the passive sentence with two animate nouns (4), while in the plausibility task it was the active sentence with an inanimate agent and animate object (1). So, even though the processing during at the verb was the same, the overall processing is not directly associated with syntactic complexity in both tasks.

In the literature, one major argument of plausibility during sentence processing is that it occurs rapidly at early processing measures for implausible or inconsistent semantics as a function of incremental processing (Staub, Rayner, Pollatsek, Hyönä, & Majewski, 2007; Murray, 2006; Kamide & Mitchell, 1999). In the case of this study, all experimental items were plausible sentences. However, the items with an inanimate agent may be considered as an inconsistent animacy value for agency preferences, especially in the active condition where the inanimate agent sits in subject position (Traxler, Williams, Blozis, & Morris, 2005; Ferreira, 1994). With the assumption that an inanimate agent is inconsistent with agency expectations, our results are in line with previous studies. For the active-voice condition with an inanimate agent, there was an effect shown during first-pass reading at the position of the inanimate agent in subject position. In addition to this, for the passive counterpart with an inanimate agent, there was an effect shown at the agent during first-fixation durations. In both cases, early processing measures reveal an effect on the agent for inconsistent animacy. During late processing, presumably after the grammatical function of the verb and thematic roles were established, it is shown that participants also take in consideration of the plausibility of the sentence as well since the processing difficulty associated with the passive-AI condition was attenuated within its inconsistent domain and its active counterpart became significantly more difficult at spec of V, region 2. This effect can be explained by past research on subject animacy effects where participants will front an animate object to subject position and passivize the sentence to avoid having an inanimate subject (Bock, Loebell, & Morey, 1992). Consequently, since the conditions with both an animate agent and patient lacked any presupposed violation in animacy, they were free from effects associated with their animacy during plausibility.

3.3 Verification

Moving into comprehension, it has been argued that there is a similarity based interference of thematic arguments when processing syntactically complex sentences, like ORCs, if the sentence is unbiased to an assumed assignment (Gordon et al., 2001; Gordon, Hendrick, & Levine, 2002; Caplan et al., 2008; Traxler et al., 2005; MacDonald et al., 1992). In line with these studies, the more complex passive voice condition was more difficult to process when both nouns were animate which is likely attributed to thematic disambiguation of competing nouns in working memory. Albeit, unlike the plausibility task, these effects were not able to be observed until the processing of the verb which suggests that the participants begin to have processing difficulty after grammatical argument is established. Another justification as to why the other passive condition with an inanimate agent failed to induce processing cost has been argued by Mak, Vonk, and Schriefers (2002); thematic roles contrasting in animacy attenuates the processing costs associated with ambiguity of thematic roles. This result is in contrast with the finding of Rayner, Warren, Juhasz, and Liversedge (2004) which revealed an early effect of an impossible thematic role assignment using a verification strategy. However, in their study it was impossible to appropriately assign the theta-role to the direct object, and in their condition where it was more or less inappropriate but possible, effects were not seen immediately. Considering this, our results are consistent with theirs as all of our items were possible utterances.

In addition to working memory ambiguity, the cognitive differences that drive the fan effects may be attributed to the storage of orthographic and phonetic representations into short term memory (Caplan et al., 2008). Despite the fact that phonological processing can occur during a semantic judgment when using tongue-twisters (McCutchen & Perfetti, 1982; Hanson, Goodell, & Perfetti, 1991), work done by Caplan et al. (2008), Caplan and Waters (1999), Daneman, Reingold, and Davidson (1995), and Chen and Shu (2001) argue that phonological representations are secondary and even optional to lexical and orthographic activation. As such, it can serve to help resolve ambiguity, phonetic interferences, and as well be used a strategy to rehearse the sentence in memory for later fact retrieval and overall comprehension (Caplan et al., 2008; Keller, Carpenter, & Just, 2003; McCutchen, Bell, France, & Perfetti, 1991). Following this assumption, phonological processing during silent reading may be an ancillary cognitive function that is used in comprehension which contributes to an increase of reading times.

4. Conclusion

The results of this study support the argument made by Caplan et al. (2008) that participants are using more cognitive processes than that are required to process the sentence in order to complete a verification task. These extra cognitive functions in turn contribute to an induced reading time of the sentence which is similar to a fan effect (i.e., over activation). More importantly, each task relies on a different set of thematic heuristics to complete the task which in turn alters processing. Verification is sensitive to similar entities during thematic disambiguation, and plausibility is sensitive to the semantic requirements assumed by a possible thematic role. Though, both tasks share similarities in syntactic processing with difficulty being found with passive sentence types at the verb. Yet both tasks are purposeful as verification tasks may be useful in studies to demonstrate overall comprehension difficulty, and plausibility tasks, can be effective to demonstrate processing difficulty associated with syntactic difficulty free of semantic effects if experimental items are controlled to be semantically neutral. As a final point, it is important for a researcher to understand that language use in a laboratory setting is a function of task, and experimental methodology needs to be appropriately selected to target a particular language or cognitive phenomena.

［References］

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive psychology*, *6 (*4), 451-474.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93 (*24), 13468-13472.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of communication disorders*, *36 (*3), 189-20.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological review*, *99 (*1), 150.

Caplan, D., Chen, E., & Waters, G. (2008). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, *44 (*3), 257-275.

Chen, H. C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review*, *8 (*3), 511-518.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye movements: A window on mind and brain*, 341-372.

Daneman, M., Reingold, E. M., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21 (*4), 884.

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language*, *33 (*6), 715-736.

Gordon, P. C., & Chan, D. (1995). Pronouns, passives, and discourse coherence. *Journal of Memory and Language*, *34* (2), 216-231.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27* (6), 1411.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13 (*5), 425-430.

Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, 42 (4), 351-373.

Hanson, V. L., Goodell, E. W., & Perfetti, C. A. (1991). Tongue-twister effects in the silent reading of hearing and deaf college students. *Journal of Memory and Language*, *30* (3), 319-330.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111* (2), 228.

Kamide, Y., & Mitchell, D. C. (1999). Incremental pre-head attachment in Japanese parsing. *Language and cognitive processes, 14 (*5-6), 631-662.

Keller, T. A., Carpenter, P. A., & Just, M. A. (2003). Brain imaging of tongue-twister sentence comprehension: Twisting the tongue and the brain. *Brain and Language*, *84 (*2), 189-203.

Koizumi, M., & Tamaoka, K. (2010). Psycholinguistic evidence for the VP-internal subject position in Japanese. *Linguistic inquiry*, *41 (*4), 663-680.

Leeser, M. J., Brandl, A., & Weissglass, C. (2011). Task effects in second language sentence processing research. *Applying priming methods to L2 learning, teaching and research: insights from psycholinguistics*, 179-198.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive psychology*, *24 (*1), 56-98.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47* (1), 50-68.

McCutchen, D., & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, *21 (*6), 672-687

McCutchen, D., Bell, L. C., France, I. M., & Perfetti, C. A. (1991). Phoneme-specific interference in reading: The tongue-twister effect revisited. *Reading Research Quarterly*, 87-103.

Murray, W. S. (2006). The nature and time course of pragmatic plausibility effects. *Journal of psycholinguistic research*, *35* (1), 79-99.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30* (6), 1290.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review*, *89 (*3), 250.

Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9 (*1), 55.

Reder, L. M., & Wible, C. (1984). Strategy use in question-answering: Memory strength and task constraints on fan effects. *Memory & Cognition*, *12 (*4), 411-419.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. Gaskell (ed.), *The Oxford handbook of psycholinguistics*, *327*-342. New York: Oxford University Press.

Staub, A., Rayner, K., Pollatsek, A., Hyönä, J., & Majewski, H. (2007). The time course of plausibility effects on eye movements in reading: Evidence from noun-noun compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33 (*6), 1162.

Tamaoka, K., Asano, M., Miyaoka, Y., & Yokosawa, K. (2013). Pre-and Post-head Processing for Single-and Double-Scrambled Sentences of a Head-Final Language as Measured by the Eye Tracking Method. *Journal of psycholinguistic research*, 1-19.

Tamaoka, K., Sakai, H., Kawahara, J. I., Miyaoka, Y., Lim, H., & Koizumi, M. (2005). Priority information used for the processing of Japanese sentences: Thematic roles, case particles or grammatical functions?. *Journal of Psycholinguistic Research*, *34 (*3), 281-332.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53 (*2), 204-224.

MANSBRIDGE Michael (MANSBRIDGE, Michael)
名古屋大学大学院 国際言語文化研究科・研究生
michaelp.mansbridge@gmail.com

早川 杏子(HAYAKAWA, Kyoko)
名古屋大学大学院 国際言語文化研究科 日本言語文化専攻・助教
hayakawa@lang.nagoya-u.ac.jp

玉岡 賀津雄（TAMAOKA, Katsuo）
名古屋大学大学院 国際言語文化研究科・教授
ktamaoka@lang.nagoya-u.ac.jp