Integration of Wh-Phrases and Predicates in Japanese Sentence Processing

Hajime ONO† and Kentaro NAKATANI‡
† Faculty of Science & Engineering, Kinki University 3-4-1 Kowakaе, Higashi-Osaka-shi, Osaka, 577-8502 Japan
‡ Faculty of Letters, Konan University  8-9-1 Okamoto, Higashi-Nada-ku, Kobe-shi, Hyogo, 658-8501 Japan
E-mail:  † onohajime@kindai.ac.jp,  ‡ kentaron@konan-u.ac.jp

Abstract In head-final languages like Japanese, one can imagine that a lot of integration processes must occur at the predicate, and that the distances between the grammatically dependent words may increase the memory loads upon integration. Yet, some previous studies failed to observe such a distance / locality effect. For instance, Nakatani & Gibson (2010) observed no distance effect for regular argument-predicate integrations, whereas Nakatani (2009) reported a distance effect for the integration of a negated verb with a Negative Polarity Item, indicating that the distance effect may vary depending on the type of integration. The present study examined the distance effect between a wh-phrase and a predicate / Q-particle, and observed a fine-grained interaction between the effect and the type of the involved dependency.

Keyword Sentence processing, dependency formation, wh-phrase, Japanese, Typing Mismatch Effect, integration cost

1. Introduction
1.1. Locality effects
One of the major tasks for the human sentence processor is to construct a hierarchical structure from a series of words. The task is never easy because the elements that hold a local relationship in the hierarchical structure sometimes appear at a great distance in terms of the linear order; in other words, there is a large discrepancy between the information about the linear order and the representation that the parser aims to build.

The point is clearly illustrated by a well-known observation that it is very difficult to process sentences with multiply embedded structures ([1]-[2]). For instance, a pair of sentences (1) and (2) express the same events, but in general, people have a hard time understanding what (1) means, and usually it takes longer to read (1), compared to (2).

(1) # The rat [which the cat [which the dog bit] chased] ran away.
(2) The dog bit the cat [which chased the rat [which ran away]].

There have been several proposals for the reason why it is difficult to process multiply embedded structures ([3]-[6], etc). According to the Dependency Locality Theory ([4]-[5]), those sentences are hard to understand because readers encounter multiple dependency relations that have not been fully established while they read the sentence. Presumably, the parser has to keep those relations in the working memory, which would render a large processing cost in the system. For instance, the sentence in (1) contains a sequence of NPs whose thematic roles have not been determined, and the parser has to store those NPs in the memory until the appropriate predicate shows up. In addition to the cost associated with storing the elements in memory, when the predicate finally appears, the parser has to retrieve those NPs from the memory in order to integrate them with the predicate and to establish a dependency relation. When the NP is far away, linearly or temporally, from the predicate, some additional processing cost (i.e., locality effect) may be induced upon its retrieval. In (1), the subject in the mid-clause (i.e., the cat) is far away from its predicate chased in that 1 full, non-pronominal NP and 1 verb intervene between them; the matrix subject the rat is even further away from its predicate (ran away) in that 2 full NPs and 2 verbs intervene between them, making the memory retrieval harder. In (2), on the other hand, all the arguments are always adjacent to their predicates, resulting in little retrieval cost.

1.2. Anti-Locality effects
In languages like Japanese, where a predicate appears at the end of a clause, one can imagine that a lot of integration processes must occur at the predicate. However, it has been often observed that in head-final languages, such as Japanese, German, and Hindi, researchers failed to find locality effects, namely a slowdown in the conditions with a longer dependency, at the predicate where various integration processes would occur ([7]-[12], but see some ERP studies in Japanese and German [13]-[14]).

For instance, Vasishth & Lewis [7] compared the reading times of the embedded verb buy, shown schematically in (3) and (4) (the original materials were in Hindi), and found that the condition with an intervening relative clause modifying the NP-ACC was read faster than that without such a relative clause (i.e., Anti-locality effect), which may be due to the pre-verbal structural predictions.
incrementally made by the parser (e.g., [7], [15]-[17]).

(3) Nothing intervening
   … NP-ACC buy …

(4) Relative clause intervening
   … NP-ACC [RC that was on the table] buy …

Also, Nakatani [10] and Nakatani & Gibson [12] investigated to what extent the integration cost at the verb region increases, varying the types of subject NPs and the distance between the subject NP and the verb. For instance, Nakatani & Gibson [12] examined the locality effect of the integration cost, manipulating the distance between the subject NP syoki-ga ‘secretary-NOM’ and the verb hookokusita ‘reported’ that assigns a thematic role to the subject, ensuring that the integration takes place. In the doubly nested condition (5), there were two nested complement clauses intervening between the subject NP and the verb, while in the non-nested condition (6), two clauses were placed at the beginning of the sentence so that the subject NP and the verb were adjacent to each other. They compared the reading times of the verb hookokusita ‘reported’ in (5) and (6), and found no significant reading time difference.

   [syoki-ga daigisi-ga syusyoo-ga]
   [secretary-NOM congressman-NOM p.minister-NOM]
   [utatanesita-to] koogisita-to hookokusita]
   dozed C] protested C] reported

“The secretary reported that the congressman had protested that the prime minister had dozed.”

(6) Non-nested: [NP3 V3-C] [NP2 V2-C] [NP1 V1]
   [syusyoo-ga utatanesita-to] daigisi-ga
   [p.minister-NOM dozed C] [congressman-NOM]
   [koogisita-to] [syoki-ga hookokusita]
   protested C] [secretary-NOM reported]

On the other hand, Nakatani [10] used an NPI (Negative Polarity Item) in the subject position, and found a locality effect for integrating an NPI with a negated verb (i.e., the negated verb was read faster when the distance between the NPI and the verb was shorter), indicating that the locality effect may vary depending on the type of integration.

2. The current study

An issue that calls for further scrutiny is why some dependency formation is sensitive to distance / locality while others are not. Focusing on the previous findings in Japanese for a moment, a dependency with an NPI is distance-sensitive while that with a regular NP is not. A difference between the two dependencies is that the dependency between a regular NP and a verb is mainly concerned about thematic integration, but the dependency between an NPI with a negated verb obviously involves more than just a thematic integration. An NPI has a grammatical feature that must be licensed by a negation morpheme. Assuming that there are distance-sensitive dependency and distance-insensitive dependency, and if the thematic integration is the former kind and the NPI-negation integration is the latter kind, the patterns observed in the previous studies in Japanese would follow.

In the current study, we would like to examine another kind of dependency to test such a classification is on the right track. Specifically, we test whether a wh-dependency between a wh-phrase and a verb with a Q-particle ka (question-licensing particle) is sensitive to distance between the two. Along the similar line of reasoning, we also examine a case where a wh-phrase is integrated with a verb without a Q-particle. Of course, there should be no integration between the wh-phrase and a Q-particle in such a case, but still integration should occur in terms of a thematic role assignment. Such a ‘partial’ integration is possible since the dependency between the wh-phrase and the Q-particle is unbounded; a Q-particle in a higher clause can license the wh-phrase as far as it c-commands the wh-phrase.

Processing mechanism for a Japanese wh-dependency has been extensively studied ([16], [18]-[19]). Miyamoto & Takahashi [19] examined sentences such as (7) and (8), and found that the parser had a strong bias to encounter a Q-particle as soon as possible. For instance, a Q-particle appeared with the embedded verb tukatteiru ‘use’ in (7), and the Q-particle appeared in the grammatically available earliest position to license the wh-phrase dono pasokon ‘which computer’. They found that the reading time of the embedded verb in (8) was slower than that in (7), which they called Typing Mismatch Effects (TME). They suggested that the slowdown occurred in (8) because the parser had to reanalyze the scope of the wh-phrase to a higher clause when it encountered the embedded verb with a declarative complementizer.
Using sentences like (8), where the wh-phrase is partially integrated with the verb, we can examine whether the magnitude of TME changes as the distance between the wh-phrase and the verb becomes larger.

3. Experiment

A self-paced reading experiment was conducted in order to investigate whether the distance effect is observed in sentences with wh-dependency.

53 native speakers of Japanese (mostly undergraduate students at Konan University) participated in the study. They were paid ¥800 for their participation in the experiment, which took about 30 minutes.

Experimental materials consisted of 24 sets of sentences prepared in a $2 \times 2$ factorial design distributed to 4 lists, manipulating Distance between the wh-phrase and the verb that assigns a thematic role to the wh-phrase (Distant vs. Local) and the Position of Q-Particle (Mid-Q vs. High-Q). In the Distant conditions, there was a complement clause intervening between the subject wh-phrase and the verb, and in the Local conditions, the complement clause was pre-posed so that the subject wh-phrase and the verb were adjacent to each other. The Mid-Q conditions included a Q-particle ka at the mid-verb (the verb in the mid-clause), which was the thematic role assignor for the wh-phrase, whereas in the High-Q conditions, a Q-particle appeared at the verb in a higher clause, which did not hold a thematic relation to the wh-phrase. Sample materials from the Distant conditions are shown in (9) and (10); the Local conditions were identical to them except for the position of the wh-phrase, which was placed right before its theta-assigner (mid-verb). Each list was combined with 54 filler sentences, so the total of 78 sentences was presented in a pseudo-random order.

The experiment was conducted with Linger written by Douglas Rohde, using Apple MacMini computers on Mac OS X. Each sentence used in the experiment was presented word-by-word in a self-paced non-cumulative moving-window reading fashion. Most of the sentences appeared on a single line, and the sentences were segmented as shown in Table 1. Participants were instructed to read the sentences at the speed as they do normally. Each of the sentences, both targets and fillers, was followed by a comprehension question about the content which can be answered by yes or no, which encourages the participants to pay attention to the content of the sentence.

4. Results

Data from the participants whose mean accuracy rates for the comprehension questions were less than 70% (which amounted to 21 participants) were eliminated, in order to make sure that only the data from serious participants were included. For each of the rest of the participants (n=32), a regression equation predicting reading times (RTs) from word length (the number of characters) was calculated using all filler and target items ([20]). In each region, the RT predicted by the participant’s regression equation was subtracted from the actual measured raw RT to obtain a residual RT.

Table 1
Segmentation of the sentence for each condition

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Distant, Mid-Q</td>
<td>Wh-nom</td>
<td>[ NP-nom Mod1</td>
<td>Mod2</td>
<td>NP-acc</td>
<td>Adv</td>
<td>Verb]-C</td>
<td>Verb-Q</td>
<td>LocPP</td>
<td>NP-top</td>
<td>Verb</td>
</tr>
<tr>
<td>b. Local, Mid-Q</td>
<td>[ NP-nom Mod1</td>
<td>Mod2</td>
<td>NP-acc</td>
<td>Adv</td>
<td>Verb]-C</td>
<td>Wh-nom</td>
<td>Verb-Q</td>
<td>LocPP</td>
<td>NP-top</td>
<td>Verb</td>
</tr>
<tr>
<td>c. Distant, High-Q</td>
<td>Wh-nom</td>
<td>[ NP-nom Mod1</td>
<td>Mod2</td>
<td>NP-acc</td>
<td>Adv</td>
<td>Verb]-C</td>
<td>Verb-C</td>
<td>LocPP</td>
<td>NP-top</td>
<td>Verb-Q</td>
</tr>
<tr>
<td>d. Local, High-Q</td>
<td>[ NP-nom Mod1</td>
<td>Mod2</td>
<td>NP-acc</td>
<td>Adv</td>
<td>Verb]-C</td>
<td>Wh-nom</td>
<td>Verb-C</td>
<td>LocPP</td>
<td>NP-top</td>
<td>Verb-Q</td>
</tr>
</tbody>
</table>
This adjusts for word-length differences (in Region 11) as well as overall differences in participants’ reading rates. Analyses of raw RTs revealed similar patterns of statistical significance.

The residual reading times were trimmed so that data points beyond 3 SD from the relevant condition × region cell mean were discarded.

At Region 8 (the mid-verb region), the Distant conditions were read slower than the Local conditions (F1(1,31)=48.5, p<.001; F2(1,23)=40.5, p<.001), indicating a strong locality effect. This effect was found between the Mid-Q conditions (F1(1,31)=27.2, p<.001; F2(1,23)=45.2, p<.001) as well as between the High-Q conditions (F1(1,31)=35.2, p<.001; F2(1,23)=9.5, p<.01). Also, there was a main effect of the position of Q-particle (F1(1,31)=6.1, p<.05; F2(1,23)=8.7, p<.01), in such a direction that the Mid-Q conditions, which contained a Q-particle in this region, were read slower than the High-Q conditions. Finally, there was an interaction between the two factors (F1(1,31)=4.1, p=.051; F2(1,23)=5.0, p<.05), and the further pairwise comparisons indicate that the effect was due to the slower reading time of the Distant/Mid-Q condition, compared to the Distant/High-Q condition (F1(1,31)=6.1, p<.05; F2(1, 23)=10.7, p<.005). There was no difference between the two Local conditions (Fs<1).

Region 9 was a spill-over region that contained a locative PP across conditions. As in the previous region, the Distant conditions were read slower than the Local conditions (F1(1,31)=8.2, p<.01; F2(1,23)=4.5, p<.5). However, unlike Region 8, the High-Q conditions were read slower than the Mid-Q conditions (F1(1,31)=4.3, p<.05; F2(1,23)=3.2, p=.089). This is likely to be a Typing Mismatch Effect ([19]), though it seems somewhat weak. There was no interaction at this region (Fs<1).

In Region 11, the final verb region, the High-Q conditions, which contained a Q-particle in this region, were read slower than the Mid-Q conditions (F1(1,31)=20.3, p<.001; F2(1,23)=18.7, p<.001). There was no Distance effect or interaction of the two factors (Fs<1). No distance effect was found even if only the High-Q conditions were compared (Fs<1).

5. General Discussion

One of the major findings in the current experiment is that there was a locality effect on the dependency between the wh-phrase and the verb, as revealed in the strong locality effect found between the High-Q conditions in Region 8, where no Q-particles were involved. The reading times of the verb became longer when the subject wh-phrase was placed at a distance. This clearly contrasts with the observation from the literature that no slowdown was observed at the verb even when there were some intervening materials between the non-wh subject and the verb ([12]).

We argue that the distance effect at the verb indicates that there is some extra integration cost strongly related to wh-phrases, but not to “regular” referential NPs. More specifically, we argue that the integration of the wh-phrase is costly since it involves operators. In some framework on semantics ([21][22]), the way of semantic calculation is different depending on the semantic type of the argument NPs. For instance, in a sentence The boy slept where the subject is referential, the semantic type of the subject is an individual. On the other hand, in sentences like Who slept? and Every boy slept, the semantic type of these quantificational subjects is a function that takes another function (in this case, slept) as its argument. Since the meaning involving an operator requires some additional computation, it seems plausible to assume that the integration between operators and predicates may involve a different type of processing that is not found in the integration between referential NPs and predicates. In terms of the on-line processing, it could be that the integration of thematic relations is unmarked and thus is kept activated because it constitutes the baseline eventive semantics, while the integration of other types of grammatical relations such as scope-related relations could be temporarily deactivated in the course of processing because they constitute secondary semantic relations.

Another finding is the main effect of the position of Q-particle,
but it is quite likely that this effect was due to the difference between the two Distant conditions, as we also found a significant interaction of the two factors. Comparing the two Distant conditions in Region 8, the participants took longer to read the verb with the Q-particle ka (Mid-Q) than the verb with a declarative complementizer to (High-Q). This result seemed to be puzzling at first, because in previous studies (e.g., [19]), it was the verb with a declarative complementizer to that was read slower when preceded by a wh-phrase (TME), whereas the contrast we found here was in the opposite direction (anti-TME). Why did we find an anti-TME? We argue that the verb with a Q-particle in this region was read slower because the Mid-Q conditions involved the Wh-Q integration process that was not present in the High-Q condition, which should lead to the anti-TME we found. Note that this Wh-Q integration cost must be distinguished from the operator-predicate integration mentioned earlier. Note that the operator-predicate integration should occur regardless of the presence of the Q-particle; in the Mid-Q conditions, the Wh-Q integration takes place in addition to the operator-predicate integration. The overall pattern with respect to the integration processes and their distance sensitivity was summarized in Table 2.

Then why did we not find a TME, which is reported in the literature? We can only speculate on this, but one possibility is that the magnitude of the anti-TME (which mainly reflects the cost of Wh-Q integration) was so large in the Distant conditions that the TME was completely overwhelmed. Assuming that the magnitude of the distance effects in the two Distant conditions was the same, this approach suggests that the Wh-Q integration cost and the reanalysis cost (this is what TME reflects, according to Miyamoto & Takahashi [19]) were differently influenced by distance. Because those two costs are based on the different processes anyway (i.e., one about integration, the other about reanalysis), it is not at all surprising to find such a difference between the two.

Note that we did not find an anti-TME between the Local conditions. We can conjecture that the anti-TME and the TME cancelled out each other in the Local conditions, because the anti-TME was relatively small. Note further that we found a slowdown of the High-Q conditions in the spillover region (Region 9), and if this is really an effect originating from the previous region, this slowdown might support the view that the TMEs were already present in Region 8.

Alternatively, as for the slowdown in Region 9, it could be that the TME was simply not induced until the spillover region in our experiment for some reason – maybe because our materials were somewhat more complex than those in the previous studies. Assuming that the TME showed up in Region 9 for the first time, the reading time data showed that the magnitude of the TME in the Distant conditions was not different from the Local conditions, which is evident from the fact that there was no interaction between the two factors. The observation that the TME was not sensitive to distance is fully compatible with the suggestion that the TME reflects the reanalysis cost ([19]). It has been argued that the reanalysis cost is distance-sensitive only when the left edge of the constituent to which the parser has to return was fairly easy to identify ([6]), and such a theory in fact predicts there is no distance effect in the current reanalysis process triggered by the lack of Q-particle.

Finally, in Region 11, which is the sentence-final verb region, the participants took more time to read the conditions with a Q-particle (High-Q) than the conditions without a Q-particle (Mid-Q). The interpretation of this finding requires some care, because the above-mentioned difference corresponds to the different verbs used in the experiment. The verbs that appeared without a Q-particle were the verbs that take an interrogative clause as its complement, such as tazuneru 'ask'. On the other hand, the verbs that appeared with a Q-particle were the verbs that allow a declarative clause as its complement, such as syutyoosuru 'claim'. Thus, the difference in the reading time data may simply indicate that interrogative-taking verbs were somehow read faster than non-interrogative-taking verbs. However, considering that the length-adjusted residual reading times were used in the analysis, and that the effect was very large, it could be that the presence of the Q-particle was really causing the found effect. Note that the main effect of the Q-particle was independently found in Region 8, where the same verb was used across conditions.

But then, we need to account for the lack of locality effect in Region 11. As we have claimed that Wh-Q integration is one kind of dependency that is sensitive to distance (see also Table 2), it seems natural to expect to find a locality effect between the two High-Q conditions. However, we did not find such an effect. The lack of locality effect can be accounted for by following a re-activation process suggested by Vasishth & Lewis [7]. According to them, the locality effect in integration arises because the element stored in memory will decay and it requires efforts to recall the element whose activation level is not high. This memory-decay account has a direct relevance to the current

Table 2
Integration type and distance sensitivity

<table>
<thead>
<tr>
<th>distance sensitivity</th>
<th>distance insensitive</th>
<th>distance sensitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>integration type</td>
<td>thematic integration</td>
<td>operator predicate integration</td>
</tr>
<tr>
<td>Wh-phrase + Verb-Q</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Wh-phrase + Verb-Comp</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Regular NP + Verb</td>
<td>○</td>
<td></td>
</tr>
</tbody>
</table>

- 103 -
observation since it makes a specific prediction that the locality effect would disappear if elements were re-activated as the incoming words are processed. We argue that although wh-phrases appeared in different positions in the Distant/High-Q condition and the Local/High-Q condition, wh-phrases in both conditions were re-activated at the same point in the embedded clause (at the verb in Region 8) in order to establish dependencies with respect to the thematic integration and operator-predicate integration. Then, such a re-activation nullifies the difference between the two conditions. This is in fact another support for the existence of the integration between the operator and the predicate even though one of the integration processes, namely the Wh-Q integration, is postponed.

Finally, let us touch upon some previous findings of locality effects involving wh-phrases in a different experimental setting. Felser, et al. [23], using German, found a P600 effect at the verb when a wh-phrase was placed at a distance (cf. [24]). It would be interesting to see whether different types of integration processes would indeed show different electrophysiological indices if the current paradigm is tested in an ERP experiment.

6. Conclusion
In this paper, we have shown that the dependency between the wh-phrases and the verb exhibits a locality effect that has not been often found in the dependency with referential NPs in head-final languages. The locality effect was found even when a verb appeared with or without a Q-particle. We have argued that the integration of a verb with a referential NP does not show a locality effect because the thematic integration is not sensitive to distance, while the integrations between a wh-phrase and its thematic role-assigner and between a wh-phrase and a Q-particle are both distance-sensitive. It was further suggested that the integration with a wh-phrase triggers an extra process with a verb because processing operators requires a marked, secondary semantic calculation. Taken together, studies on dependency integration provide a great testing ground for a model of the real-time sentence comprehension.

References