Understanding Wh-Questions in Context

Edson T. Miyamoto  
University of Tsukuba

1 Introduction

The question in (1a) is ambiguous, it allows the two alternative readings in (1bc).

(1)

a. Ambiguous wh-question

Motoki-ga donna-ryori-o tukutta-ka Taku-wa kotaeta-no?
Motoki-Nom wh-dish-Acc cooked-QM Taku-Top answered-QM

b. Embedded-scope wh-phrase: “Did Taku answer what kind of dish Motoki cooked?”

c. Matrix-scope wh-phrase: “What kind of dish did Taku answer Motoki cooked?”

The embedded interpretation in (1b) is preferred over the matrix interpretation in (1c). One possible explanation is that working-memory constraints favor short dependencies between words in the sentence (e.g., between wh and QM; Miyamoto and Takahashi, 2002a; MT, hereafter). But other factors may also play a role (e.g., implicit prosodic contours imposed during silent reading, pragmatics-related requirements; for a recent summary, see Kitagawa and Hirose, 2012; KH).

A problem with previous experimental work is that questions such as (1a) were shown without surrounding context, making the task artificial and not providing information about the interpretation favored after the question was read. Two reading-time experiments are reported confirming the preference for the embedded interpretation. The preference remains with word orders that have been claimed to force the matrix interpretation (Takahashi, 1993).

2 Experiment 1

As a baseline for (1a), the question marker ka (QM) was replaced with the declarative to (that, hereafter) resulting in the unambiguous question in (2).

(2) Unambiguous wh-question with matrix-interpretation

Motoki-ga donna-ryori-o tukutta-to Taku-wa kotaeta-no?
Motoki-Nom wh-dish-Acc cooked-that Taku-Top answered-QM

“What kind of dish did Taku answer Motoki cooked?”
Reading times to each corresponding region in the two questions were compared. For example, the first region (the embedded subject) in (1a) was compared to the same region in (2). There should be no reading-time difference at this point since the same words are read in the two sentences. Similarly, no difference is predicted for the second region (the wh-phrase). The crucial comparison is for the third region (the embedded verb and marker). The declarative marker in (2) should be read more slowly than the QM in (1a) if the matrix interpretation is dispreferred (MT).

2.1 Method

2.1.1 Participants
A total of 42 native Japanese speakers, students at the University of Tsukuba, were paid to participate in the experiment. Results for one participant are not reported (see the analysis section).

2.1.2 Materials
Sentences (1a) and (2) are schematically shown in (3ab).

The questions varied according to a 2 by 2 design with two factors manipulating the embedded clause, namely, its word order (SOV or OSV) and marker (i.e., the marker on the embedded verb: that or QM). (There was a third factor discussed later, in a total of eight conditions.)

Bigrams (i.e., the frequency of co-occurring pairs of segments) can affect reading times (McDonald and Shillcock, 2003), and may be a factor in the comparison between (3ab) because accusative-marked wh-phrases are likely to occur with high frequency immediately followed by a QM-marked verb (e.g., in single-clause questions). The non-canonical OSV conditions in (3cd) avoid such adjacency effects as the embedded subject intervenes between the wh-phrase and the verb. If bigram-based expectation is responsible for the predicted difference between (3ab), then no such a difference should be observed between (3cd).

Contexts like the following were included before each question.

A third factor manipulated the type of answer given by Y. In (4), it is an NP (“hamburger”), therefore it assumes a matrix-interpretation for X’s question. For the embedded interpretation, the answer was always affirmative (e.g., kotaetayo “answered” or “he, did”; hereafter, “yes”).
The contexts did not include wh-questions to avoid priming effects and were not intended to force the matrix or embedded interpretation for the ambiguous questions in (3ac). Nevertheless, the contexts may have favored the embedded interpretation. According to recent work, the matrix interpretation of (1a) is pragmatically marked and can only be supported by complex contexts where alternative scenarios are entertained justifying the need for the embedded QM (KH). This is clear given the type of meaning that KH assume for the matrix interpretation of (1), which is indistinguishable from the same question replacing ka with kadoka “whether”.

But even if the the embedded interpretation is favored, it may nevertheless be felicitous to answer it with an NP. This is because even for unambiguous questions (e.g., “did Mary say what John cooked?”) the matrix answer (e.g., “hamburger”) is informative (Grice, 1975).

Twenty-four sets of test items were created with the wh-phrase as an accusative-marked object of the embedded predicate (following the schemes in (3)). Twenty-four distractor filler items were created including wh-phrases with various types of markers. All items followed the structure of (4), that is, a dialogue, a warning (highlighted with asterisks), X’s question and Y’s answer.

2.1.3 Procedure and analyses

A non-cumulative self-paced reading presentation was used. Segments were initially shown masked with underscores and were revealed one at a time at each button press. Contexts were shown one sentence at a time. The crucial question was shown in five segments as in (3). Participants rated each dialogue on a 7-point scale (1 totemo husizen ‘very strange’, 7 totemo sizen ‘very natural’). The 24 test items were distributed into eight lists according to a Latin Square design. Each participant saw one list interspersed with filler items in random order.

Because participants tend to speed up as the experiment progresses, reading times (in milliseconds) were residualized over trial number. Reading times beyond 2.5 standard deviations from the participant’s mean for each segment were removed (affecting less than 3.6% of the data). The experiment had a total of eight conditions in a 2 by 2 by 2 design. For the reading times to the question only the first two factors (word order and marker) were included in the analyses because the type of answer was not relevant at that point. For the analyses of the reading times to the answer, the third factor (i.e., the type of answer presented, NP or “yes”) was also included.

The unambiguous questions in (3bd) require an NP answer, and a “yes” answer should be rated as less natural. That was the case for 41 participants. One participant rated “yes” answers higher than NP answers in the OSV conditions, therefore this participant’s results are not reported.

Trends similar to those reported were observed with all data included.

2.2 Results and discussion

2.2.1 Reading times to the contexts and crucial questions

There were no reading-time differences in the context regions (analyses of variance over participants’ and items’ means: all Ps > .2). For region 1 of the crucial questions, there was an interaction between type of complementizer and word order but this is a spurious result since those factors were not manipulated at that point (sometimes participants rest at the first segment of the sentences leading to such unexpected differences). There was no difference in region 2 (all Ps > .3). Reading times for regions 3 to 5 in Figure 1 are residualized over trial number (negative/positive numbers indicate reading times faster/slower than the overall means).
The crucial region was number 3 (the embedded verb and marker) but the effect persisted in region 4 (the matrix subject), therefore the two regions were analyzed together (there was no interaction between region and the two factors of interest, \(P_s > .3\)). There was a main effect of marker as the \textit{that} conditions were read more slowly than the QM conditions (\(F_1(1,40) = 7.73, P = .008; F_2(1,23) = 6.58, P = .017\)), replicating previous results (e.g., MT). (The main effect of word order with region 3 faster than region 4, is not of interest here.) There was no interaction between marker and word order (\(Fs < 1\)), therefore bigrams are unlikely to explain the result.

Nevertheless, word order did lead to some marginal effects. In region 3 alone, there was an interaction between marker and word order in the participant analysis (\(F_1(1,40) = 5.02, P = .031; F_2(1,23) = 1.48, P = .24\)). This interaction was absent in region 4 indicating that the OSV order delayed the marker-related effect but did not eliminate it.

In region 5 (the matrix verb and the question marker \textit{no}), there was a main effect of marker as the QM conditions were slower than the \textit{that} conditions (\(F_1(1,40) = 5.14, P = .029; F_2(1,23) = 12.9, P = 0.002\)). This slowdown may be related to implicit prosodic contours for the matrix interpretation (see KH for related discussion; but see the correlation analyses later on).

### 2.2.2 Reading times to the answers and overall ratings

The following are the reading times to the answers and medians for the 7-point naturalness ratings.

<table>
<thead>
<tr>
<th>Marker:</th>
<th>Mean reading times (in ms) to the answers</th>
<th>Median naturalness ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answer:</td>
<td>QM</td>
<td>\textit{that}</td>
</tr>
<tr>
<td>“yes”</td>
<td>“yes”</td>
<td>“yes”</td>
</tr>
<tr>
<td>SOV</td>
<td>-274</td>
<td>240</td>
</tr>
<tr>
<td>OSV</td>
<td>-255</td>
<td>182</td>
</tr>
</tbody>
</table>

As expected the \textit{that} conditions were interpreted with matrix scope. The “yes” answers had lower ratings (2.0) than the NP answers (6.0) regardless of word order (for SOV and OSV separately: \(Ps < .001\), Wilcoxon paired tests on medians over participants and over items). The matrix preference is also clear in the reading times (“yes” slower than NP answers, \(Ps < .01\)).
For the QM conditions, the embedded-interpretation preference observed in the reading times to the questions (regions 3 and 4) remains as “yes” answers were faster than NP answers ($P_s < .001$; there was no effect of word order or interaction: all $F_s < 1$). Although informative, NP answers were slow to read, perhaps because accommodation is needed to recompute the NP answer (e.g., “hamburger” as “yes, hamburger”).

The “yes” answers were rated higher than the NP answers in the QM-SOV condition ($P_s < .001$) but not in the QM-OSV condition ($P_s > .8$) raising the possibility that wh-fronting (as in the OSV conditions) facilitates matrix scope. This is tested in Experiment 2.

### 2.2.3 Correlation analyses

For each of the eight conditions, mixed-effects analyses (Baayen et al., 2008, and references therein) were conducted to examine how the reading times to regions 3, 4 and 5 (as fixed effects, without interaction terms) were related to the reading times to the answer (intercepts for participants and for items were included as random effects). The table below reports results for which $P_s < .00625$ (with Bonferroni correction, a 5% level corresponds to $0.05/8 = 0.00625$).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Region</th>
<th>Estimate</th>
<th>SE</th>
<th>$t$</th>
<th>$P_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSV-that-yes</td>
<td>3</td>
<td>2.34</td>
<td>.56</td>
<td>4.20</td>
<td>.0001</td>
</tr>
<tr>
<td>SOV-QM-NP</td>
<td>4</td>
<td>-1.14</td>
<td>.41</td>
<td>-2.80</td>
<td>.0060</td>
</tr>
</tbody>
</table>

In the OSV-that-yes condition, there was a positive correlation between the reading times to region 3 and the reading times to the answer, potentially suggesting that slow reading times at *that* were associated with stronger expectations for an NP answer, hence the slow reading times to the “yes” answer (the correlation was negative in the OSV-that-NP, albeit not reliable, $P > .6$).

In the SOV-QM-NP condition, there was a negative correlation between region 4 and the answer, indicating that relatively-long reading times at the QM were associated with expectations for an NP answer. Note that the matrix interpretation was probably seldom expected, but in those rare instances when it was expected, it may have led the QM to be surprising and slow to read.

Recall that the QM conditions were read more slowly than the *that* conditions in region 5 (the matrix verb) of the crucial questions. This could be related to matrix-scope computations (see KH for listeners relying on overt prosodic contours at that point). In that case, a negative correlation should be observed between region 5 and NP answers, since computing the matrix scope should facilitate the matrix answer. However, correlations between region 5 and the answer were positive. An alternative explanation is that readers slowed down because they noticed the ambiguity.

### 3 Experiment 2

In Experiment 1, the embedded-interpretation preference tended to be weaker in the OSV order (see the reading times to region 3 and the naturalness ratings). But in preliminary results with a new group of 42 participants, the preference remained in (5) (contra Takahashi, 1993).

(5) Donna-ryori-o Taku-wa Motoki-ga tukutta-ka kotaeta-no?
    wh-dish-Acc Taku-Top Motoki-Nom cooked-QM answered-QM

The OSSV conditions ((5) represented as (6a) and its baseline (6b)) replaced the SOV conditions of Experiment 1. Otherwise the same 2 by 2 by 2 design and procedure were used.
In the QM conditions, “yes” answers were favored over NPs: they were read faster ($P_s < .01$; no interaction with word order, OSSV or OSV: $P_s > .15$) and were rated higher ($P_s < .05$).

<table>
<thead>
<tr>
<th>Marker</th>
<th>Mean reading times (in ms) to the answers</th>
<th>Median naturalness ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QM</td>
<td>that</td>
</tr>
<tr>
<td>Answer:</td>
<td>“yes” NP</td>
<td>“yes” NP</td>
</tr>
<tr>
<td>OSSV</td>
<td>-249 448 -97 -324</td>
<td>4.5 3.0 2.0 6.0</td>
</tr>
<tr>
<td>OSV</td>
<td>-265 116 -38 -221</td>
<td>5.0 3.0 2.0 6.0</td>
</tr>
</tbody>
</table>

Nevertheless, comprehension may have unfolded differently in OSSV-QM questions (partially supporting Takahashi, 1993). When reading times to the embedded-verb region were compared to those of the SOV condition of Experiment 1, there was an interaction between marker and word order (SOV, OSSV) in the participant analysis ($P_1 = .01$, $P_2 = .165$) as OSSV-QM questions were numerically slower than OSSV-*that* questions (the opposite trend of the SOV and OSV conditions).

## 4 General discussion

A number of issues need to be addressed in the future. There were too few filler items in the experiments (24 fillers to 24 test items) and all accusative-marked wh-phrases were in the test items, therefore experiment-specific strategies cannot be ruled out. Data on “S wh-Acc S V V” questions would provide a more complete picture (Miyamoto and Takahashi, 2002b, for a related discussion). But the results are promising and suggest that reading times to answers can provide useful information on ambiguous wh-phrases. If the results of Experiment 2 hold, they raise intriguing possibilities about the processing of scope-freezing constructions (Takahashi, 1993).

## References


