THE ACQUISITION OF MULTIPLE WH-QUESTIONS
BY HIGH-PROFICIENCY NON-NATIVE SPEAKERS OF ENGLISH

ROBERT BLEY-VROMAN
University of Hawai‘i
NAOKO YOSHINAGA
Hirosaki Gakuin College

This paper investigates the knowledge of multiple wh-questions such as Who ate what? by high-proficiency non-native speakers of English whose first language is Japanese. Japanese grammar is known to license a wider range of such questions than English—who came why, for example—although the precise theoretical account is not yet clear. Acceptability judgments were obtained on six different types of such questions. Acceptability of English examples was rated by native speakers of English, Japanese examples were judged by native speakers of Japanese, and the English examples were judged by high-proficiency Japanese speakers of English. The results for native speakers judging their own language were generally in accord with expectations. The high-level non-native speakers of English were significantly different from native speakers in their ratings of these sentences. However, the ratings were clearly not simply the result of transfer. The consequences of this finding for theories of Universal Grammar in second language acquisition are discussed.

INTRODUCTION

Multiple wh-questions contain two or more wh-phrases, all of which are used to request information. Thus, in response to, Who ate what?, an appropriate answer would be Tom ate an apple, Mary ate a banana, and John ate an orange—this answer supplying information for both who and what. It has long been noted that not all combinations of wh-phrases are equally possible in English. For example, most native speakers of English would feel multiple wh-questions such as (1a) or (1b) to be acceptable but not (1e) or (1f). (Although (1c) and (1d) are less acceptable than (1a) and (1b), their grammaticality status is somewhat more problematic, as will become clear.)

(1) a. Who ate what?
    b. Who sat where?
c. Who sang where?

d. Who went when?

e. Who came how?

f. Who cried why?

In contrast to English, wh-phrases in Japanese are quite free of such restrictions. The questions in (2), which correspond to those of (1), are all completely natural in Japanese.

(2) a. Dare ga nani o tabemashita ka?
who Nom what Acc ate Q
‘Who ate what?’

b. Dare ga doko ni suwatteimashita ka?
who Nom where in sat Q
‘Who sat where?’

c. Dare ga doko de utaimashita ka?
who Nom where at sang Q
‘Who sang where?’

d. Dare gaitsu ikimashita ka?
who Nom when went Q
‘Who went when?’

e. Dare ga donoyooni kimashita ka?
who Nom how came Q
‘Who came how?’

f. Dare ga naze nakimashita ka?
who Nom why cried Q
‘Who cried why?’

LINGUISTIC THEORIES

Standard Account: Complements vs. Adjuncts

It is commonly assumed in the linguistic literature that multiple wh-questions that contain two argument wh-phrases (e.g., Who ate what?) are grammatical, whereas those involving an argument wh-phrase in subject position and adjunct wh-phrases (e.g., Who cried why?) may not be. The “standard” account in the Principles and Parameters framework relies on the Empty Category Principle (ECP) and on the fact that complements, but not adjuncts, are lexically governed.

In this ECP-based approach, it is generally assumed that in English the trace of the
**ACQUISITION OF MULTIPLE WH-QUESTIONS**

subject wh-word (who in the examples above) is antecedent-governed by the WH complex in CP Spec at LF. (In an English multiple wh-question such as those in (1), the subject WH phrase moves to Spec of CP first, where its index percolates to CP Spec.) Therefore, the WH complex is, as a whole, indexed with the subject trace, and the WH complex is unavailable to antecedent-govern other wh-traces. (When other wh-phrases are moved to Spec of CP in LF, they adjoin, but their indices do not percolate.)

The wh-trace of an object, as in *Who saw what?*, as a subcategorized complement of the verb, is in a lexically governed position, so that even if it cannot be antecedent-governed from Spec of CP, it complies with the ECP. Likewise, the where expression in 1b, as a complement of the verb, is lexically governed, so this example obeys the ECP.

The adjunct wh-phrases, in 1c–1f, on the other hand, are not complements of the verbs, so their traces are not in lexically governed positions; hence these multiple wh-questions violate the ECP. (Recall that the WH-complex itself cannot properly govern these traces since its ability to antecedent-govern is “used up,” as it were, by the subject wh-phrase; thus the adjunct wh-phrase traces are not properly governed.) Native speakers should find these sentences less acceptable as a result.

In Japanese, on the other hand, traces in subject position are lexically governed (depending on the analysis, either by nominative *ga* or by Infl). This frees up the WH-complex, which is now available to antecedent-govern phrases in lexically un governed position—specifically adjuncts. Thus, examples like those in (3) will not violate the ECP, corresponding to Japanese speakers’ judgments that such sentences are completely natural and acceptable.

(10)  

a. [CP [IP dare-ga naze kimashita ] ka]?

b. [CP [darei nazej ] [IP ti -ga tj kimashita ] ka]?

As a result, the split in grammaticality at the adjunct-complement divide, which is present in English, does not exist in Japanese.¹ We will call this account the “standard” account.

¹ If we accept the idea that Infl is responsible for lexical governement in Japanese, the parametric values relevant here would be whether Infl is a lexical or functional (non-lexical) category (see Cole, Hermon, & Sung 1990). The default setting would be non-lexical Infl. Cole, Hermon, & Sung (1990) discuss the triggering stimulus needed to change the setting (to lexical Infl) as “the occurrence of nongap topics and null topics (topic chains) and the absence of that-trace effects.”
The Status of How and Why as Against When and Where

It has been noted that among adjuncts there seem to be differences in the acceptability of multiple wh-questions containing them. Specifically, adjunct where and adjunct when seem to be more acceptable than adjunct how and why in multiple wh-questions with subject wh-phrases (Huang 1982; Aoun, Hornstein, Lightfoot, & Weinberg 1987). There are several approaches to accounting for this phenomenon. For example, Huang (1982) argues that the trace of where and when can be lexically governed by null prepositions, whereas how and why are not.

Aoun et al. (1987) working in their Generalized Binding Theory, attribute the differences in acceptability to differences in the referentiality of where/when vs. how/why. They claim that “locational and temporal adjuncts [where and when] are referential (as suggested by the existence of pronouns that can be used coreferentially: here, there, then, now), whereas manner and reason adverbs [how and why] are less likely to be referential” (1987, p. 552). Subject traces in English must be bound because of the properties of English AGR. Other wh-words can only appear with a subject wh-word when the traces of these other wh-words need not be bound. This will be the case with the referential wh-phrases: direct object what, where, and when. In contrast, Japanese, according to this account, has no AGR. Therefore, subject traces in Japanese need not be bound, and any wh-words can appear with subject wh-word.²

Kuno and Takami (1993) attribute the special characteristics of how and why in part to functional or discourse-based effects. They propose that syntax requires why and how to be base-generated left-most in questions. (This by itself directly rules out *who came why and *who came how.) The syntax might, in principle, permit structures like *why who came (or *why did who come). However, sentences like these are ruled out by an additional, functionally based Sorting Key Hypothesis, which requires that the surface order of wh-words reflect the way information is categorized, the left-most wh-word functioning as the sorting key of a list. Since we do not usually categorize multiple answers by reasons or means, then why and how will not in general be acceptable as a left-most wh-words in a multiple wh-question. A parallel explanation can be given for the ungrammaticality of any multiple wh-question with why or how preceding what or who

² If we take Aoun et. al’s view, the relevant parametric values would be +Agr language (English) and – Agr language (Japanese or Chinese). At least for multiple wh-questions, we could assume that +Agr is a default and Japanese children can switch +Agr to –Agr by hearing a grammatical multiple wh-questions with adjuncts. This assumes the theory in which AGR is subject to parametric variations. In the early Minimalist program (Chomsky 1992; Chomsky 1995, Chapter 3), languages universally have AGR.
ACQUISITION OF MULTIPLE WH-QUESTIONS

(*Why did you buy what? etc.).

In Japanese, why and how must also be base-generated at the left edge of the sentence, but they can be reordered by scrambling the subject wh-word over why or how (from naze dare-ga kita no? ‘why who-Nom came Q’ to dare-ga naze kita no? ‘who Nom why came Q’). In this way, Japanese scrambling can bring these examples in line with the Sorting Key Hypothesis. (All examples in (1) and (2) satisfy the Sorting Key Hypothesis because all of the left most wh-phrases are who in (1) and dare ‘who’ in (2).)

While we know of no fully developed Minimalist treatment of these matters, the work of Hornstein (1995) is suggestive. English why is there argued to be base-generated in Spec CP, but this syntactic position conflicts with the fact that why cannot function as the “generator” of the list associated with multiple wh-questions (citing Watanabe, 1994 and extending that proposal to English). Although Japanese naze likewise cannot function as a generator, it is not located in Spec CP (adopting the proposal of Rizzi, 1990). (Certain aspects of this proposal are clearly reminiscent of that of Kuno and Takami, but larger differences of framework make them difficult to compare directly.) Hornstein is cautious about whether how should be treated in the same way as why.

Despite the differences in theoretical accounts, there is substantial agreement on the main features of the phenomenon. The facts which we primarily rely on are three. First, all scholars agree that in English direct object (e.g., what) and complement where are grammatical in the relevant multiple wh-questions. Second, some theories also propose that cases with adjunct where and when may be grammatical, at least under certain conditions. Third, Japanese is clearly different: the correspondents of all these examples are completely grammatical in that language.

CONSIDERATIONS FOR SECOND LANGUAGE ACQUISITION

Regardless of the linguistic analysis, it is clear that Japanese learners of English, if they are to be completely “successful,” must possess the knowledge that the target language allows only a subset of multiple wh-questions that are possible in their native language. The focus of this study is to investigate the extent to which high-proficiency non-native speakers of English have acquired knowledge of these differences. Since multiple wh-questions are virtually never explicitly presented in the classroom, to our knowledge, and since the grammaticality differences among them certainly are not introduced, the predominant source of information must presumably be naturalistic input. Yet, such
positive input does not include any obvious indication of the ungrammaticality of the sentences. Thus, as is often pointed out in the literature on language acquisition, if knowledge of the grammaticality distinctions among such types is attained, this knowledge must derive from the innate language acquisition device of Universal Grammar. Universal Grammar is clearly designed to develop exactly the properties on which the possibilities for multiple wh-questions depend, and it must be the case that it does so on the basis of readily available input.

One view of SLA holds that adult second language acquisition and child language development are the same in being guided by Universal Grammar. This is sometimes called the "full-access" hypothesis. Under this view, one would predict that naturalistic L2 positive evidence must also be able to result in successful learning by adults in these cases (see e.g., Epstein, Flynn, & Martohardjono, 1996). The "full access" view thus leads to the expectation of successful learning of these UG-governed aspects of the target language by adult second language learners, possibly from quite early stages of acquisition, and native-like performance by high-proficiency non-natives is certainly to be anticipated.

Other views of second language acquisition would suggest that the mechanism that guides child language acquisition is not available to adult language acquisition or interfered with by other factors. If UG does not guide adult language learning as it does child language development, then, broadly speaking, success in these areas should not be achieved.

While there are many subtle differences among the theories of non-access/partial access, for our purposes we initially consider only the contrast between such theories, taken as a group, and the full-access view. We return to the various differences in the discussion. While it cannot be expected that the results will provide unequivocal answers to these complex issues, they will illuminate the debate.

RESEARCH QUESTIONS

The purpose of the experiment is to investigate the extent to which high-proficiency Japanese second-language speakers of English approximate English native speakers in judging the acceptability of the six different multiple wh-questions in English. Native speaker judgments on the acceptability of the six types of multiple wh-questions are also examined. Native English speakers’ performance is compared to nonnatives, which is in turn compared with nonnatives’ own judgments in their own native language. Research
questions and hypotheses for the study are presented as follows:

1. How do native speakers of English distinguish among these six different types of English multiple wh-questions in their acceptability judgments?

   Native speakers of English should distinguish among these types of English multiple wh-questions. More precisely, if the standard account is correct, those with subject wh-phrases and subcategorized complement wh-phrases should be rated as acceptable, whereas others are not acceptable. If Huang (1982), Aoun et. al (1978) or Kuno and Takami (1993) is correct, then only those with subject who and how or why would be ungrammatical.

2. Do native speakers of Japanese accept all of these six different types of multiple wh-questions in Japanese?

   If linguistic-theoretical accounts are correct, native speakers of Japanese should accept all six different types of Japanese multiple wh-questions.

3. Do high-proficiency Japanese learners of English behave like English native speakers in their acceptability judgments?

   If adult learners are guided by UG, they should resemble English speakers in their acceptability judgments.

THE STUDY

Subjects

A total of 45 subjects participated in the study: a group of 18 native speakers of English and a group of 27 Japanese speakers of English. The native speakers of English were graduate students or faculty members at the University of Hawai‘i. During the time of the study, the Japanese speakers of English were either graduate students (81.5%) or had finished the University of Hawai‘i (typically, working toward a Ph.D.) (18.5%). Their advanced-level English proficiency is commensurate with their status at an English-speaking university in which instruction and academic work is in English and where they must interact with and compete with native speakers. The native-speaker group

---

3 In most cases, TOEFL scores were available: mean = 610.19, SD = 26.6 (total possible = 677; 600 = 84th percentile; this means that those people are in the top 16% on average). These TOEFL scores underestimate the proficiency of the students, perhaps substantially. Often, the scores are several years old, with TOEFL taken before arrival in the U.S. and before the extensive exposure to English which American university life entails. Among subjects with current TOEFL scores (less than two years old), there were no scores under 600.
consisted of 13 men and five women; the Japanese speaker group consisted of 13 men and 14 women.

Materials

There were two translation-equivalent versions of the questionnaires: one in English and one in Japanese. The questionnaires consisted of a page of instructions followed by three pages of test material. At the end was a page for voluntary comments or opinions. The test materials consisted of 24 pictures, each followed by a multiple wh-question of which the subjects were asked to rate the acceptability. There were four tokens of each of six types of multiple wh-questions (a total of 24 items). Examples of six types of multiple wh-questions used in the study are shown in (25).

(25)  
b. Who is sitting where? : where-complement type (whereC)  
c. Who is singing where? : where-adjunct type (whereA)  
d. Who is walking when? : when type (when)  
e. Who is getting down how? : how type (how)  
f. Who is crying why? : why type (why)

A sample item from the questionnaire is presented in Figure 1.

Figure 1. An example of an item used in the questionnaire

Who is eating what?  
-3 -2 -1 0 1 2 3
ACQUISITION OF MULTIPLE WH-QUESTIONS

Procedures
The questionnaire was distributed to the subjects during the Fall semester of 1995. The subjects were asked to rate the acceptability of the sentences. A seven-point rating scale was used, ranging from -3 to +3, "completely impossible" to "completely possible". English speakers were administered only the English questionnaire; all Japanese speakers completed both the English and Japanese questionnaires. Following standard practice, the Japanese subjects took the English questionnaire first, then the Japanese questionnaire, to avoid any direct carryover from Japanese judgments to English. There were no time limits.

Analyses
All the hypotheses essentially concern the way judgments of different example-types will cluster, and with the relative sizes of the differences among those clusters. In order to test these hypotheses, cluster analyses (Ward's method) were performed to investigate the way the different sentence types grouped with respect to acceptability. The cluster analyses were performed on each set of native language data and on the non-native data. In addition, to address hypotheses 1 and 2, a one-factor repeated-measures analysis of variance (ANOVA) was performed for each set of native language data. The one within-subjects factor was sentence type (what, where-complement, where-adjunct, when, how, and why). For testing hypothesis 3, a two-factor repeated-measures ANOVA was performed to determine the significance of the overall differences between English native speakers' performance and Japanese subjects' performance in judgments of English multiple wh-questions. The between-subjects factor was language (English vs. Japanese) and the within-subjects factor was sentence type. Planned comparisons were conducted to determined which pairs of means were significantly different. The alpha decision level was set at .05 for all inferential statistics, with adjustments made for multiple comparisons as appropriate.

RESULTS

Native Data
The results of the ANOVA indicated that the effect of sentence type was statistically significant for English, $F(5, 17) = 26.496, p = .0001$, reflecting the fact that there was a clear decline of the acceptability from what type to why type as shown in Table 1 and
Figure 2. The effect of sentence type was also statistically significant for Japanese, $F(5, 26) = 3.672, p = .0038$, reflecting that fact that a similar tendency in acceptability of sentences was seen as in case of English, although its slope is very much flatter than the one in English, and all sentence types are quite close to "completely acceptable" ratings, and even the lowest was 2.36 as also shown in Table 1 and Figure 2.

Table 1
Descriptive statistics for English ratings (EE) and Japanese ratings (JJ)

<table>
<thead>
<tr>
<th></th>
<th>English (EE)</th>
<th></th>
<th>Japanese (JJ)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Mean</td>
<td>SD</td>
<td>Count</td>
</tr>
<tr>
<td>what</td>
<td>18</td>
<td>2.889</td>
<td>.323</td>
<td>what</td>
</tr>
<tr>
<td>whereC</td>
<td>18</td>
<td>2.681</td>
<td>.468</td>
<td>whereC</td>
</tr>
<tr>
<td>whereA</td>
<td>18</td>
<td>1.944</td>
<td>1.376</td>
<td>whereA</td>
</tr>
<tr>
<td>when</td>
<td>18</td>
<td>.928</td>
<td>1.964</td>
<td>how</td>
</tr>
<tr>
<td>how</td>
<td>18</td>
<td>.339</td>
<td>2.199</td>
<td>why</td>
</tr>
<tr>
<td>why</td>
<td>18</td>
<td>-1.464</td>
<td>2.059</td>
<td>when</td>
</tr>
</tbody>
</table>

Figure 2. Mean differences among six types of multiple wh-questions in English and Japanese by native speakers.
The results of the planned comparisons between pairs of adjacent means are summarized in Table 2.

Table 2
*Summary of the Planned Comparisons for Native Data*

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th></th>
<th>Japanese</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>F</em>-Value</td>
<td><em>p</em>-value</td>
<td><em>F</em>-Value</td>
<td><em>p</em>-value</td>
</tr>
<tr>
<td>what vs. whereC</td>
<td>0.213</td>
<td>0.6457</td>
<td>0.014</td>
<td>0.9067</td>
</tr>
<tr>
<td>whereC vs. whereA</td>
<td>2.657</td>
<td>0.1068</td>
<td>0.996</td>
<td>0.3202</td>
</tr>
<tr>
<td>whereA vs. when</td>
<td>5.068</td>
<td>0.0269*</td>
<td>0.417</td>
<td>0.5197</td>
</tr>
<tr>
<td>when vs. how</td>
<td>1.701</td>
<td>0.1957</td>
<td>0.345</td>
<td>0.5583</td>
</tr>
<tr>
<td>how vs. why</td>
<td>15.937</td>
<td>0.0001*</td>
<td>1.378</td>
<td>0.2426</td>
</tr>
</tbody>
</table>

As indicated in Table 2, there seems to be a major break in acceptability in English between the *why* types and the rest of sentence types, and another noticeable break between the *what*, *where-complement* and *where-adjunct* types on the one hand, and the *when* type, *how*, and *why* types on the other. As for Japanese, there were no significant differences between any pairs of adjacent means.

If we consider absolute ratings, with positive values as being acceptable and negative values as being unacceptable, the results indicate that for English, the *what* type and the *where-complement* type are nearly perfect; the *where-adjunct* type is clearly on the acceptable side; and the *when* type, though lower, is also on acceptable side. The *how* type is close to the border line, and the *why* type is on the unacceptable side. For Japanese, every type is very close to "completely possible," with means never lower than just a bit below 2.5.

*Native English Data vs. Non-Native Data*

The results of a two-factor repeated measure ANOVA indicated that there was a significant interaction of language with sentence type, $F(5, 215) = 5.186, p = .0002$. This
reflects the fact that the non-natives are rating six types of multiple wh-questions differently from native speakers of English. Table 3 and Figure 3 show the comparison between the two groups (English data are repeated here for comparison).

Table 3
Descriptive Statistics for Learner's Ratings (JE) in Comparison to English Natives' Ratings

<table>
<thead>
<tr>
<th></th>
<th>English (EE)</th>
<th></th>
<th>Japanese (JJ)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count  Mean  SD</td>
<td></td>
<td>Count  Mean  SD</td>
<td></td>
</tr>
<tr>
<td>what</td>
<td>18  2.889 .323</td>
<td></td>
<td>27  2.185 1.030</td>
<td></td>
</tr>
<tr>
<td>whereC</td>
<td>18  2.681 .468</td>
<td></td>
<td>27  1.306 1.987</td>
<td></td>
</tr>
<tr>
<td>whereA</td>
<td>18  1.944 1.376</td>
<td></td>
<td>27  .894 2.138</td>
<td></td>
</tr>
<tr>
<td>when</td>
<td>18  .928 1.964</td>
<td></td>
<td>27  .559 2.209</td>
<td></td>
</tr>
<tr>
<td>how</td>
<td>18  .339 2.199</td>
<td></td>
<td>27  -.128 2.349</td>
<td></td>
</tr>
<tr>
<td>why</td>
<td>18  -1.464 2.059</td>
<td></td>
<td>27  -.328 2.404</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Mean differences among six types of multiple wh-questions in English by learners in comparison to native speakers.
ACQUISITION OF MULTIPLE WH-QUESTIONS

In order to determine which pairs of means are significantly different, adjacent pairs of means for Japanese were compared. The results of these planned comparisons are summarized in Table 4 (English results are repeated here for comparison.)

Table 4
Summary of the Planned Comparisons for Learners’ Data in Comparison to Native English Data

<table>
<thead>
<tr>
<th></th>
<th>English F-Value</th>
<th>English p-value</th>
<th>Japanese F-Value</th>
<th>Japanese p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>what vs. whereS</td>
<td>.213</td>
<td>.6457</td>
<td>7.368</td>
<td>.0075*</td>
</tr>
<tr>
<td>whereS vs. where-Ad</td>
<td>2.657</td>
<td>.1068</td>
<td>1.609</td>
<td>.2068</td>
</tr>
<tr>
<td>where-Ad vs. when</td>
<td>5.068</td>
<td>.0269*</td>
<td>1.070</td>
<td>.3029</td>
</tr>
<tr>
<td>when vs. how</td>
<td>1.701</td>
<td>.1957</td>
<td>4.495</td>
<td>.0359*</td>
</tr>
<tr>
<td>how vs. why</td>
<td>15.937</td>
<td>.0001*</td>
<td>.381</td>
<td>.5382</td>
</tr>
</tbody>
</table>

Figure 4. Another view of the mean ratings for six types of multiple wh-questions in English by learners in comparison to native speakers.
As Table 4 shows, the major break lies between the what type and the other sentence types, reflecting that fact that Japanese non-native speakers rated the what type as being of high acceptability, but they rated rest as of intermediate acceptability. Figure 4 shows this rather clearly: all the ratings other than the what type cluster around zero, whereas the ratings by English native speakers are quite spread out on the scale.

Although we saw that the mean ratings for the what type are quite high, it is worth noting that they are still not as high as those of native speakers nor as high as the subjects’ own mean ratings for the Japanese equivalents. In fact, all their mean ratings are lower than those of native speakers of English (except for the why type), and their ratings are much lower than their own almost-perfect ratings for Japanese sentences.

The results of a cluster analysis for each native and non-native group even more clearly reveal the different rating patterns between native and non-native speakers. See Figures 5 and 6. Indeed, cluster analysis is particularly appropriate for this sort of case: what matters is not so much significant differences between adjacent means as the overall grouping pattern. (The corresponding dendrogram for native Japanese ratings is in Figure 7; since there is so little difference among the sentence types in the native Japanese data, there is little cluster structure of interest. One can see this in the very “flat” structure of the figure in Figure 7, which is drawn to the same scale as Figures 5 and 6.)

The dendrograms are read from bottom to top, and at the first step, the two closest cases (most similar cases, or least distant cases) are combined into a single cluster. Thus, the earlier stages of the linkage shows the strongest similarity between the clusters. If we examine the dendrogram in Figure 5 for native speakers’ ratings, the what type and the where-complement type go together (in fact, there is no clustering which separates all the what examples from the where-complement examples). The what/where-complement cluster in turn goes together with the where-adjunct type. This group of the what/where-complement types and the where-adjunct types further go together with the when and how types, which independently are more closely clustered together. At the final stage, the group of those types is combined with the why types, indicating that the why types are most distant from other types (the squared Euclidean distance between them was 292.65).

In case of non-native speakers’ ratings in Figure 6, the why type and the how type go together, and the where-adjunct type and the when type go together, which further go together with the where-complement type. These groups and the why/how types then merge together. At the final stage, the group of those types is merged with the what type, indicating that the what type is most distant from other types (the squared Euclidean distance between them was 220.75).
Figure 5. Clustering of native English speakers' ratings of English
Figure 6. Clustering of Japanese speakers' ratings of English
Figure 7. Clustering of Japanese speakers' ratings of Japanese

Hence, and in summary, for native-speakers, the *what* and *where*-complement types form a compact cluster, and the major cluster-break is between the *why*-type and all other types; whereas, for the non-natives, the major cluster break is between the *what* type and all other types.

**DISCUSSION**

We posed three research questions, to which we now relate our results.

1. _How do native speakers of English distinguish among these six different types of English multiple wh-questions in their acceptability judgments?_

Native speakers of English should distinguish among these types of English multiple wh-questions. More precisely, if the standard account is correct, those with subject wh-phrases and subcategorized complement wh-phrases should be rated as acceptable, whereas others are not acceptable. If Huang (1983), Aoun et al (1978) or Kuno and Takami (1993) is correct, then only those with subject _who_ and _how_ or _why_ would be ungrammatical.

The results were generally in conformity with the existing syntactic accounts. The English *what*-type questions and those with subcategorized *where*-phrases (*where*-complements) were rated about 2.5 or higher by native speakers of English, and all theoretical approaches consider these grammatical. _When_ and _where*-adjunct fall lower on the scale, but still in the positive range. Let us call the range 0.5 to 2.0 “intermediate
acceptability." The positive ratings for these examples is more in accord with Huang, Aoun et al.'s and Kuno and Takami's account than the standard account, but the fact that ratings are clearly lower than that of the what and where-complement case is also compatible with the standard view. Why and how are both below (or just barely above) the zero line—a finding compatible with all theories.

One might possibly have expected the ratings for the why and how cases to be lower, given their theoretical ungrammaticality under all views. One possible reason for this slightly better than expected acceptability may relate to the functionally based List Requirement Hypothesis, proposed by Kuno and Takami (1993). They argue that the acceptability of multiple wh-questions must be accounted for by interactions of functional factors (such as the List Requirement Hypothesis) and syntactic factors (such as those discussed earlier).

The List Requirement Hypothesis is that a clear list of items that correspond to each of the multiple wh-expressions of a question must be predetermined for such a question to be acceptable. Examples (23) and (24), from Kuno and Takami (p. 121) show how who came why and who died how become somewhat more acceptable, given the appropriate listing context:

(23) The left column gives you a list of people who came, and the right column gives you a list of reasons why they came. The lines connecting entries of the left column to those of the right column tell you who came why.

(24) The police are looking into the unexpected deaths of patients which have happened in the past ten years at this hospital to find out who died how.

Who came why? and Who died how? are ungrammatical and unacceptable in isolation, but become somewhat more acceptable if they are given in appropriate contexts which make clear that there is a list of items that corresponds to each of the multiple wh-expressions such as in (23) and (24), satisfying the List Requirement Hypothesis.

The relative acceptability of such example sentences would thus depend in part on how straightforward it is to imagine the list of items that would correspond to each of the multiple wh-questions in a given context. In our questionnaire, sentences to be rated were accompanied by a picture, which might have served as a predetermined list of items. This may have raised ratings of how and why types.
ACQUISITION OF MULTIPLE WH-QUESTIONS

Some such overlay of functional considerations on grammaticality might also account for the fact that other aspects of the ratings are not a perfect reflection of theoretical syntax. For example, although *where-adjunct* and *when* are both of intermediate acceptability (and are treated in a unitary fashion by all the grammatical theories under review), *where-adjunct* is still somewhat more acceptable than *when*. Also, it is possible that *what* is slightly more acceptable than *where-complement*, though the difference here is not statistically significant in this sample, and the examples are grouped closely in our cluster analysis. Perhaps, in general, lists of items that link persons with things are easier to imagine than lists which link persons with places; and these are easier to imagine than lists that link persons with times, manners, or reasons.

2. Do native speakers of Japanese accept all of these six different types of multiple wh-questions in Japanese?

If linguistic-theoretical accounts are correct, native speakers of Japanese should accept all six different types of Japanese multiple wh-questions. In the results, native speakers of Japanese accepted all six different types of Japanese multiple wh-questions, conforming to theoretical grammaticality. Even the lowest mean ratings of the six types was very close to the maximum ratings of 3. There was a slight (but statistically significant) effect for type of question. This, too, is plausibly the result of some subordinate functional factor—perhaps something like the List Requirement Hypothesis.

3. Do high-proficiency Japanese learners of English behave like English native speakers in their acceptability judgments?

If adult learners are guided by UG, they should resemble English speakers in their acceptability judgments. The results indicated that high-proficiency Japanese learners of English rated the acceptability of six multiple wh-questions significantly differently from native speakers of English.

The ratings for the *what* type were quite high, though not as high as those of native speakers nor their own ratings for Japanese equivalents, whereas the rest were clustered around the “intermediately acceptable” area. In addition, the cluster analyses revealed important differences between native and non-native acceptability.
IMPLICATIONS FOR THE SLA THEORIES

The Japanese knowledge of multiple wh-questions is clearly not the same as that of native speakers. Native speaker judgments largely follow the fault lines given by grammatical theory. Most notably, native speakers group what and complement where together in a single cluster, and the other cases fall off in acceptability, with why being worst. The non-natives, in contrast, made the major distinction between what and all other cases; the complement where examples are not grouped with direct objects.

The non-native speakers tested in this study are quite advanced. They have survived in graduate-level courses in the English-speaking university, and most of them have lived in the United States for many years. Therefore, it seems implausible that the acquisition device would not yet have been exposed to that simple and readily available data which, under the assumptions of UG-based acquisition theory, should suffice for development of the relevant knowledge. (In fact, the notions of lexical and antecedent government on which the standard theory bases the complement-adjunct split certainly need not be learned at all.)

Nonetheless, the precise triggering data for establishing the properties of multiple wh-questions is not yet known, so that one cannot say with complete confidence that no UG-based account could explain why differences between natives and very high-proficiency non-natives should remain. Further, it is at least conceivable that there is some way in which a default parameter setting, together with input and native-language factors, might result the judgment patterns observed in our non-native subjects. In the absence of any concrete proposal, one is left with mere speculation. (But see footnotes 1 and 2 for some ideas of initial parameter settings.) Much theoretical work will need be done in the area of UG-based SLA research before firm conclusions can be reached.

Some scholars of second language acquisition have taken the position that the Universal Grammar that guides child language acquisition is fully available (and thus parameters can be reset in the course of the second language acquisition), but that second-language competence can deviate from native-language competence as a result of transfer of parameter settings from the native language (see e.g., Vainikka & Young-Scholten 1996; Schwartz & Sprouse 1996; Eubank 1996). This view is often considered a variant of the “full-access hypothesis.” While the results at hand do show differences between non-natives and natives, these differences cannot be attributed to transfer in any obvious way. Furthermore, the differences exist at a much more advanced stage than would probably be anticipated under such a view.
Other theories of second language acquisition (sometimes considered varieties of the "partial access" view) have emphasized the differing role played by positive evidence and negative evidence in child language development and second language acquisition. For second language acquisition, research results have suggested that exposure to positive input is not always sufficient for second language acquirers to learn the ungrammaticality of target-like structures if the native-language correspondents are grammatical. In such cases, learners will assume that the target language is like the native language and will remain stuck, perhaps forever. (Although explicit negative evidence may in theory be useful in these circumstances, the effects of such negative evidence appear to be only short-term.) See White (1991; 1990/1991; 1992) for discussion. However, from this perspective, the expected failure for Japanese learners of English should be that of accepting all types of English multiple wh-questions, as in Japanese, but this was not the case. The question would remain unanswered of why the *what* type was quite high in the non-native ratings whereas other types cluster round the zero-range.

Non-parameter-settings accounts, such as the Fundamental Difference Hypothesis, hold that second language acquisition depends on individual learners' noticing specific structural patterns in the input and adding them to their developing structure-stores (Bley-Vroman 1990, pp. 42-43; 1996). Learning of structural patterns will in general be conservative—a learner only incorporates a pattern into the grammar if it is encountered (and noticed) in the input. A learner who has not encountered (and noticed) target-language multiple wh-questions will not be inclined to accept them as grammatical. The system also is assumed to possess some sort of strengthening mechanism: the more a pattern is encountered and noticed, the more it will be accepted. High-proficiency learners may then come to accept multiple wh-questions, to the extent they have been encountered.

To acquire the grammaticality distinctions among these multiple wh-questions, the learner must not only notice wh-phrases, the learner must distinguish among kinds of wh-phrases—must distinguish, say, between *who*-phrases and *what*-phrases, or between subject phrases and non-subject phrases, or between subject phrases and complement phrases and adjunct phrases. Not all learners will be expected to make the same distinctions; however, at least some will make approximately the right distinctions and notice the descriptively accurate facts about the occurring types of multiple wh-questions. With time, one would expect some learners to come to distinctions in acceptability (pattern strength, actually) which would reflect, though dimly, the distinctions made by native speakers. One predicts considerable inter-learner variability, though acceptability
judgments of a body of learners, taken collectively, might in time begin to resemble the distinctions of the native speaker grammar. On the other hand, one does not expect to find sharp distinctions in grammaticality.

If multiple wh-questions are indeed infrequently occurring structures and thus are not often encountered and noticed by learners, this might explain the general lower ratings of learners in English, even if all the patterns are possible in their native language. Such an account would not attribute this acquisition difficulty to language transfer interacting with the parameter-setting mechanisms of UG.

The infrequency of multiple wh-questions is generally undisputed; as Givón notes, "While multiple WH-questions are a possible pattern in English, they are clearly not a very common one" (1993, p. 259). Indeed, the rarity of these structures, which is hardly surprising when one considers the special contexts that they require, is one of the things which makes the grammaticality distinctions particularly interesting from the point of view of language learnability theory. On the other hand, detailed studies of the frequency of multiple wh-questions have, to our knowledge, never been undertaken. Still less do we have investigations of the comparative frequencies of different subtypes.

Of course, what really would matter, under the pattern accumulation theory, is not just input frequency, but whether aspects of the input are noticed. Although informal and anecdotal, the Japanese subjects' own comments about the questionnaire are suggestive and seem to relate to the rarity of the structural patterns and to the importance of "noticing." One non-native speaker, who rated all of what type and subcategorized complement where type as +3 and the rest as −3 (except for one token of nonsubcategorized where rated 0), said that he had heard, "Let's see who was standing where" in the movie Twelve Angry Men. He said that his judgments were based on whether he had heard a similar pattern before. Another non-native speaker also noted that all sentences should be unacceptable, because he had never heard of them; but, he said that verbs such as eat and hold "need an object"; therefore, he said, it might not be impossible to have sentences like who is eating what? or who is holding what? He rated all of what types as 0, and all the rest as −3. Another noted that she rated all of who is eating what? and who is holding what? as +3 because she had "very often heard native English speakers making such [what-type] questions." She also noted that in certain cases, she recalled having been corrected in her English by native speakers and having been told that when there are two wh-words, one of them should be placed after 'and'. She apparently remembered cases with why and when. Based on these corrections, she rated those with why and when as −3 and, as she said, generalized this unacceptability to other cases of
where and how (but not to the what type, presumably since she had "often heard" those). This subjective data must be interpreted with great caution; future research must test these ideas in a systematic way. (The last case, incidentally, casts some doubt on the view that negative evidence is ineffective.)

While it is true that our results showed that Japanese speakers' acceptability ratings of the English examples were significantly different from those of native speakers, it is also true that the lines in Figure 3 do seem to have a similar slope—there are some similarities in acceptability judgments, at least to degree. In part, this similarity might be a result of the input-sensitive, noticing-based pattern accumulation mechanisms discussed above. Alternatively (or additionally) the similarities might derive in part from Kuno and Takami's List Requirement Hypothesis and our extended interpretation of that hypothesis, which was also invoked in explaining the graded acceptability of English native results and Japanese native results above. That is, the Japanese speakers' ratings for English might reflect the probably universal pragmatic tendency of List Requirement Hypothesis, in that it is easier to presuppose lists of items that correspond to who and what than to who and where, which might be easier than who and why.

In our view, the clarification of the issues raised in this study will require three main research thrusts. First, the interaction of functional/discourse and syntactic factors in sentences of this sort will need to be investigated. Second, the possibility that learners are basing their judgments in part on the patterns which they notice in the input will require both a more elaborated theory of patterns and, more extensive empirical studies of frequency and other factors which may affect noticing. Corpus investigation is obviously indicated. Third, the possibility that UG is guiding acquisition here can only be tested with a more fully-developed theory of the parameter-setting mechanisms which could result in the failure of high-proficiency non-natives to demonstrate native-like judgment clusters. We are skeptical about the prospects here, especially as regards the need to predict a major split between what and the other types and the need to develop an account which will result in children's developing the right syntax and in adults' failing to do so, even after many years of input exposure.
REFERENCES


ACQUISITION OF MULTIPLE WH-QUESTIONS


Robert Bley-Vroman
Department of ESL
University of Hawai‘i
1890 East-West Road
Honolulu, HI 96822

vroman@hawaii.edu

Naoko Yoshinaga
Hirosaki Gakuin College
Eibeibungaku-ka
13-1 Minori-cho
Hirosaki-shi, Aomor-ken 036
Japan

naoks@infoaomori.ne.jp