Ueno, M. & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language*, 86, 243-271.

Event-Related Brain Indices of Japanese Scrambling

Mieko Ueno and Robert Kluender

Department of Linguistics, University of California, San Diego

Running head: ERP INDICES OF JAPANESE SCRAMBLING

March 2002

Please send all correspondence to Mieko Ueno, Department of Linguistics, University of California, San Diego, La Jolla, CA 92093-0108. Email: ueno@ling.ucsd.edu.

This research was supported by NIHDC02503-01A1; the first author was also supported in part by NIH training grant NIHDC00041-08/09 through the Center for Research in Language, and by a Joseph Naimam graduate fellowship in Japanese Studies at UCSD. The authors would like to thank Laurie Stowe, Maria Polinsky, John Moore, S.-Y. Kuroda, Matthew Walenski, Kara Federmeier, Marta Kutas, Seana Coulson, Susan Garnsey, Janet Dean Fodor, Edson Miyamoto, Hiroko Yamashita, Reiko Mazuka, Atsu Inoue, Hiroko Hagiwara, and an anonymous reviewer for helpful comments and discussion, Naoyuki Osaka and Mariko Osaka for providing their Japanese version of the reading span test, Yasu-Hiko Tohsaku for help with materials, and Ronald Ohst, H. Wind Cowles, and Sean Perry for technical support.

Abstract

This study examined the processing of Japanese *wh*-questions with preposed (scrambled) vs. *in-situ* (canonical SOV word order) *wh*-objects, and of yes/no-questions with scrambled vs. *in-situ* demonstrative objects. Questions with scrambled objects elicited bilateral slow anterior negative potentials between filler and gap. Scrambled demonstratives elicited P600 effects following the filler and (L)AN /P600 effects at the gap, while scrambled *wh*-words elicited primarily (L)AN effects at the gap. This replicated effects in response to filler-gap dependencies created by *wh*-movement in other languages, supporting the existence of universal parsing operations for all types of filler-gap dependencies. We suggest that these results are most generally compatible with notions of canonicity in sentence processing.

Key Words:

event-related brain potential, ERP, anterior negativity, LAN, P600, Japanese sentence processing, scrambling, filler-gap dependency, head-driven parser, canonical word order

Event-Related Brain Indices of Japanese Scrambling

Introduction

Japanese is a language with SOV (Subject-Object-Verb) word order: canonically, the subject and then the object(s) precede the verb, which always appears clause-finally. However, subjects and objects can occur in a different order relative to each other. The word order in (1a) is canonical, and that in (1b) is non-canonical.

(1) a. Calvin-ga pizza-o mottekita.
Calvin-NOM(INATIVE) pizza-ACC(USATIVE) brought
'Calvin brought pizza'

b. Pizza-o Calvin-ga mottekita.
pizza-ACC Calvin-NOM brought
'Calvin brought pizza'

Saito (1985) proposed that sentences like (1b) are created by moving the direct object NP *pizza-o* 'pizza-ACC' from its underlying SOV position to the beginning of the clause (specifically, to the specifier of the adjoined S position in the phrase structure configuration in vogue at the time), as shown in Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

This preposing operation in Japanese is generally referred to as scrambling. Assuming the analysis shown in Figure 1, scrambling is one way of creating a so-called "filler-gap" dependency (Fodor, 1978), in which 'pizza-ACC' is the displaced element referred to as the "filler", with the canonical object position (marked by t_i in Figure 1) representing the site of origin, or "gap", of this displacement.[1]

There is already an abundant literature of behavioral (Holmes & O'Regan, 1981; MacWhinney & Pleh, 1988; King and Just 1991), event-related brain potential (ERP: Kluender & Kutas, 1993a, b; King & Kutas, 1995; Müller, King & Kutas, 1997), positron emission tomography (PET: Stromswold, Caplan, Alpert, & Rauch, 1996; Caplan, Alpert, & Waters, 1998; Caplan, Alpert, & Waters, 1999; Caplan, Alpert, Waters, & Olivieri 2000), and functional magnetic resonance imaging (fMRI: Just, Carpenter, Keller, Eddy, & Thulborn, 1996), studies that have investigated the role of verbal working memory in the processing of filler-gap dependencies in English. It has long been postulated that associating displaced fillers with their gaps as in (2a) involves additional verbal working memory load, inducing an extra processing cost relative to sentences like (2b) that do not involve such dependencies.

(2) a. What did Calvin bring __ ?

b. Did Calvin bring pizza?

In the ERP literature, this processing cost has been argued to be reflected in a negative-going voltage deflection with a left frontal maximum between 300 and 500 ms post-word-onset, known as left anterior negativity or "LAN", which typically first appears immediately after the introduction of a filler in sentence structure. Kluender and Kutas (1993 a, b) examined ERPs at the positions immediately following the filler and the gap of various English *wh*-questions, and concluded that both storing a filler in working memory and its retrieval for filler-gap assignment are indexed by LAN. King and Kutas (1995) reported related effects in the processing of object and subject relative clauses as in (3).

		Region 1	Region 2	Region 3
(3) a.	The reporter [who __	harshly attacked	the senator]	admitted the error.
b.	The reporter [who	the senator	harshly attacked __]	admitted the error.

In averages across the entire sentence, King and Kutas found a relatively frontal, bilateral slow negative potential in response to object relative sentences (3b) when compared to subject relative sentences (3a) in Regions 1 and 3. In addition, as in Kluender and Kutas (1993a), there was a phasic LAN effect immediately following the gap in the object relative clause condition (3b) (i.e. at the main verb *admitted* in Region 3). This effect was superimposed on and independent of the slow bilateral negative potential ranging across Region 3.[2]

To date there have been two studies of the processing of object vs. subject *wh*-questions in German. Kluender and Münte (1998) showed slow left-lateralized anterior negative potentials associated with the verbal working memory load in bi-clausal object *wh*-questions relative to bi-clausal subject *wh*-questions. Fiebach, Schlesewsky, and Friederici (2001) likewise reported slow negative potentials lateralized to left anterior regions of scalp in response to embedded object *wh*-questions vs. embedded subject *wh*-questions, in addition to a late positivity (a so-called P600 effect, Osterhout & Holcomb, 1992) at the pre-gap position of *wh*-object questions. This latter effect was reminiscent of P600 effects reported for pre-gap positions in English *wh*-

object questions, interpreted as indexing costs of syntactic integration related to gap-filling (Kaan, Harris, Gibson, & Holcomb, 2000).

In the present study, sentences with and without scrambled objects were compared to each other, in order to determine if there would be indices of increased working memory load in scrambled sentences similar to those seen in cases of *wh*-movement in English and German. German is in fact a language that has both *wh*-movement and scrambling. Rösler, Pechmann, Streb, Röder, and Henninghausen (1998) examined the processing of ditransitive sentences in which subject, indirect object, and direct object (S-IO-DO) were presented in various scrambled orders within verb-final clauses. Subjects responded faster to comprehension questions based on canonical word order S-IO-DO sentences than to questions based on scrambled S-DO-IO sentences. These response time data corroborated findings from an earlier acceptability judgment study using the same stimuli (Pechmann, Uszkoreit, Engelkamp, & Zerbst, 1994), and both measurements corresponded with ERP responses to the same stimuli: the determiner of the scrambled direct object in the S-DO-IO sentences elicited greater anterior negativity, especially over the left side of the head, when compared to the determiner of the indirect object in canonical S-IO-DO sentences.

As for studies of Japanese scrambling, Yamashita (1997) reported that it did not take subjects any longer to read scrambled sentences than it did to read canonically ordered sentences. However, when Mazuka, Itoh, and Kondo (in press) compared canonical order SOV sentences with scrambled OSV sentences, subjects rated the scrambled sentences more difficult and misleading. This difference in acceptability manifested as significant processing costs for scrambled OSV sentences in both self-paced reading and eye-movement studies. Crucially, the subject NP in scrambled sentences like 'husband-ACC wife-NOM __ awaited' took longer to read than the direct object NP in canonical word order sentences like 'husband-NOM wife-ACC awaited.' Mazuka, Itoh, and Kondo (2001) used the same type of stimuli in an ERP study, and reported late positivities (P600 effects) in response to both NPs in scrambled OSV ('husband-ACC wife-NOM __ awaited') sentences.[3,4]

To sum up, convergent evidence from several ERP studies of filler-gap dependencies suggests that the following ERP components might be relevant in the context of clause-internal scrambling in Japanese: slow negative potentials across the course of a sentence related to holding a filler in working memory, phasic LAN effects between 300 and 500 ms related to retrieving a filler from working memory for purposes of assigning it to a gap, and P600 effects related to integrating a gapped position in the syntactic parse. In the study reported here, our main question of interest was whether sentence constituents displaced via scrambling in a strictly head-final language with no syntactic process of *wh*-movement like Japanese, and thereby deviating from canonical SOV word order, would nevertheless show evidence of being processed by the brain like a filler-gap dependency. In particular, we were interested in seeing if there would be any evidence in the ERP record of filler maintenance between a scrambled sentence constituent and its canonical word order position, and if there would be any ERP evidence of gap-filling once that canonical position was reached, or whether there would instead be ERP evidence that the parser waited until the sentence-final subcategorizing verb before attempting filler-gap association. The former pattern of results would be predicted by an active filler strategy (Frazier & Clifton, 1989), in which the parser attempts to assign a filler to a gap as soon as possible, while the latter pattern of results would be predicted by a head-driven parsing model (Pritchett, 1992), in which arguments are attached into the syntactic parse tree and assigned thematic roles only when the subcategorizing verbal head is processed at sentence end. Insofar as there was already evidence from behavioral studies of head-final languages like German (Bader & Lasser, 1994) and Japanese (Mazuka et al., in press) falsifying the predictions of a head-driven parser, we suspected that if there were any evidence at all for gap-filling in the ERP record of scrambled sentences, it would be seen early, i.e. at the canonical sentence position of scrambled constituents and before the end of the sentence.

Methods

Subjects

20 (11 female) native speakers of Japanese between 19-29 years of age (mean: 25) who had been outside Japan for less than 2 years were included in the study. Subjects had normal or corrected-to-normal vision, were right-handed, and had no neurological or reading disorders. Subjects were reimbursed for their time.

Materials

Stimuli consisted of four conditions of mono-clausal questions, namely, (a) yes/no-questions with demonstrative objects *in-situ*, (b) object *wh-in-situ* questions, and (c) yes/no-questions with scrambled demonstrative objects, and (d) scrambled object *wh*-questions, as shown in (4).

(4) Sample Stimuli

a. *In-situ* demonstratives

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

sono inochishirazuno bokenka-ga toto *sore-o* mitsuketa-ndesu-ka.
the/that reckless adventurer-NOM finally that-ACC discovered- POL- Q

'According to the local newspaper, did that reckless adventurer finally discover that?'

b. *Wh-in-situ*

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

sono inochishirazuno bokenka-ga toto *nani-o* mitsuketa-ndesu-ka.
the/that reckless adventurer-NOM finally what-ACC discovered- POL- Q

'According to the local newspaper, what did that reckless adventurer finally discover?'

c. Scrambled demonstratives

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

sore-o sono inochishirazuno bokenka-ga toto__ mitsuketa-ndesu-ka.
that-ACC the/that reckless adventurer-NOM finally discovered- POL- Q

'According to the local newspaper, did that reckless adventurer finally discover that?'

d. Scrambled *wh*

Ano jimotono shinbun-ni yoruto
the local newspaper-to according

nani-o sono inochishirazuno bokenka-ga toto__ mitsuketa-ndesu-ka.
what-ACC the/that reckless adventurer-NOM finally discovered- POL- Q

'According to the local newspaper, what did that reckless adventurer finally discover?'

200 sets of sentences containing these four conditions were constructed. Fillers consisted of four conditions of interrogative sentences, namely, (a) ditransitives in canonical word order, (b) ditransitives with accusative-marked demonstrative objects scrambled within the VP, preceding

the dative objects, (c) ditransitives with accusative-marked demonstrative objects scrambled within S (IP), preceding the nominative subjects, and (d) embedded clause *wh*-questions with *wh*-subjects. The *in-situ* subject *wh dare-ga* 'who-NOM' was used to prevent subjects from always expecting *wh*-words with object-marking (*nani-o* 'what-ACC' or *dare-o* 'who-ACC'). Ditransitive sentences were used to lead subjects to expect dative objects as well. Different types of scrambling constructions were used so that subjects would not expect scrambled constituents in any particular sentence position.

The 200 sets of experimental sentences were placed in a Latin square design to create four parallel lists of 200 experimental sentences such that no one subject saw more than one sentence from each set. The 200 filler questions were added to each list, and then each list was pseudorandomized and divided into 20 sets of 20 sentences each.

Procedure

Subjects were run in two sessions lasting about 2.5 hours each. Subjects were seated in a reclining chair in a sound-attenuated room and wore an elastic cap mounted with tin electrodes. An illuminated rectangular border appeared uninterruptedly in the middle of the screen during presentation of stimulus sentences for purposes of fixation. Stimuli were presented on a computer screen in Japanese characters basically one *bunsetsu* at a time with 650 ms duration and 650 ms[5] stimulus onset asynchrony. A *bunsetsu* consists of one free morpheme (lexical word or pronoun) and the bound morpheme/s associated with it (particles modifying the noun/verb such as the object and subject markers), and will be referred as a “word” hereafter. The interstimulus interval between sentences was three seconds, and subjects were given as much rest as they wished between sets of sentences.

In order to maintain subjects' attention, comprehension questions were inserted in the stimuli. Every five sentences on average but at a semi-random interval (at least three and at most seven comprehension questions were inserted in a set of 20 sentences), subjects were asked to answer a comprehension question regarding the immediately preceding sentence. Comprehension questions appeared after an interval of 1000 ms following the offset of the sentence-final word, and remained on the screen until subjects responded to them by pressing one of two buttons held in their hands. The next sentence began 3000 ms after the button press. Before beginning with the first experimental set, subjects were given a practice set of 20 sentences.

Electrophysiological Recording

The electroencephalogram (EEG) was recorded from 19 positions, including all standard positions of the international 10/20 system, using tin electrodes mounted in an elastic cap. Reference electrodes were positioned on the two mastoid processes, and the EEG was algebraically referenced off-line to the mean of the activity at these two electrodes. Additional electrodes were placed beneath the right eye and at the outer canthi of the two eyes to detect lateral eye-movements and blinks for later correction. The EEG was amplified with a bandpass of 0.01 to 100 Hz, digitized with four-ms resolution, and stored on hard disk for off-line analysis. A blink-correction algorithm based on a spatial filter was used to retain as many trials as possible for averaging.

Data Analysis

Measurements were taken of single-word averages for phasic effects and of two-word and four-word averages for longer-lasting effects. Single-word averages consisted of 1000 ms epochs including a 100 ms prestimulus baseline; two-word averages consisted of 2000 ms epochs (2 x 650 ms stimulus onset asynchrony [SOA] plus a 100 ms prestimulus baseline); four-word averages consisted of 3000 ms epochs (4 X 650 ms SOA plus a 400 ms prestimulus baseline), unless otherwise noted.

In order to see effects of scrambling, overall ANOVAs treating movement (*in-situ* vs. scrambled) and type (demonstrative vs. *wh*) as factors were conducted. In addition to the two condition factors, the ANOVAs included additional topographical factors. The statistical analyses were done separately on midline (Fz, Cz, and Pz), parasagittal (Fp1/2, F3/4, C3/4, P3/4, O1/2), and temporal (F7/8, T3/4, T5/6) electrodes. Midline analyses consisted of three-way ANOVAs with three within-group factors, including movement, type, and three levels of anterior/posterior sites. Parasagittal analyses consisted of four-way ANOVAs with four within-group factors, including movement, type, two levels of hemisphere, and five levels of anterior/posterior sites. Temporal analyses consisted of four-way ANOVAs with four within-group factors, including movement, type, two levels of hemisphere, and three levels of anterior/posterior sites. Pairwise comparisons were undertaken between *in-situ* vs. scrambled demonstratives and between *in-situ* vs. scrambled *wh* in the face of significant interactions between movement and type in some electrode array/s for a particular sentence position. In such cases, additional ANOVAs were run on the three electrode arrays, treating movement, hemisphere, and anteriority as factors. An alpha level of .05 was used for all statistical tests, with a *p*-value of .10 considered marginally significant[6]. The Huynh-Feldt correction for lack of sphericity was applied whenever applicable. Original degrees of freedom are reported with the corrected probability level.

Results

The mean correct response rate to comprehension questions across subjects was 91% (range: 80%-97%). Thus no subject's data were excluded from the ERP analyses based on poor comprehension. The blink rejection rate (subsequently corrected by the blink-correction algorithm) was approximately 13%. The artifact rejection rate for other reasons was approximately 17%, probably due to the length of the stimulus sentences.

Table 1 shows the English gloss of the sample stimuli. Recall from the introduction that we planned to investigate the effects of holding a filler and of filling a gap.

[INSERT TABLE 1 ABOUT HERE]

More specifically, to observe effects of filler storage on the ERP record, we planned to investigate the sentence positions between the displaced object 'that/what-ACC' and the presumed gap before the verb, i.e. 'that reckless adventurer finally', as shown in Table 1. As for gap-filling effects, there were two possible sentence positions to check for such effects, namely, before and after the posited gap position. For example, in a question with a scrambled object like Table 1b, the gap would presumably be between 'finally' and 'discovered-Q', and comparisons were made at both positions. However, the parser might also try to dispose of the filler as soon as possible (Frazier & Clifton, 1989), and since a sentence like (5a) (which places the demonstrative object right after the subject, omitting the adverbial 'finally') is also possible, comparisons were also made at 'adventurer-NOM'.

(5)

a. Canonical word order

The local newspaper-to according that reckless adventurer-NOM ***that-ACC*** discovered-Q.

'According the local newspaper, did that reckless adventurer discover that?'

b. Scrambled word order

The local newspaper-to according ***that-ACC*** that reckless adventurer-NOM (__) discovered-Q.

'According the local newspaper, did that reckless adventurer discover that?'

Therefore, 'adventurer-NOM finally' was treated as the pre-gap region, while 'discovered-Q' was treated as the post-gap region and examined separately.

Effects of Holding a Filler

Between Filler and Gap

This section examines the effects of filler storage by comparing the ERPs to sentence positions between filler and gap, i.e. 'that reckless adventurer-NOM finally' in Table 1. Visual inspection of the waveforms elicited by this four-word string indicated that scrambled demonstrative and *wh*-objects elicited slow potential effects of increased bilateral anterior negativity (AN) in comparison to their *in-situ* counterparts (Figure 2). In addition, a distinct positivity was observed near the end of the first word in the string ('that') for scrambled demonstratives (Figure 3), but not for scrambled *wh*.

First, to confirm the observation of slow potentials, overall ANOVAs were run on mean amplitude measurements between 950 and 2600 ms, covering the second half of 'reckless' and all of 'adventurer-NOM finally'[7]. In this latency window, there was a marginal main effect of movement in parasagittal electrodes [$F(1,19) = 3.99$, $p = .06$] and a significant interaction between movement and anteriority in parasagittal [$F(4,76) = 3.59$, $p = .038$] and temporal [$F(2,38) = 6.18$, $p = .005$] electrode arrays. These indicated that the ERPs to the scrambled conditions were more negative than those to the *in-situ* conditions at anterior electrodes (see Figure 2).

[INSERT FIGURE 2 ABOUT HERE]

Second, to test the effect of late positivity at the first word of the string ('that'), separate overall ANOVAs were run in the latency window of 500 and 800 ms. Supporting the observation of a late positivity in response to scrambled demonstratives, there was a marginal interaction between movement and type in the midline array [$F\{1, 19) = 4.09$, $p = .057$], as well as a marginal movement x type x anteriority interaction in the parasagittal array [$F(4, 76) = 2.46$, $p = .10$]. Due to the statistical interactions involving type, pairwise comparisons were made between *in-situ* vs. scrambled demonstratives, and between *in-situ* vs. scrambled *wh*. When scrambled demonstratives were compared to *in-situ* demonstratives, there was a significant main effect of movement in the midline [$F(1, 19) = 11.24$, $p = .003$] and parasagittal [$F(1, 19) = 4.59$, $p = .045$] arrays, supporting the visual impression that scrambled demonstratives were more positive in this latency window (see Figure 3)[8]. There was no statistically reliable difference in the scrambled *wh* vs. *wh-in-situ* comparison.

[INSERT FIGURE 3 ABOUT HERE]

Effects of Filling a Gap

Pre-Gap Region

This section examines the ERPs to the pre-gap region, 'adventurer-NOM finally' in Table 1, for possible gap-filling effects. Visual inspection of the waveforms suggested that compared to their *in-situ* counterparts, scrambled conditions (demonstrative and *wh*) elicited a left frontal negativity roughly around 300-600 ms poststimulus onset of both 'adventurer-NOM' and 'finally'. In addition, there was a positivity in response to scrambled demonstratives around 500-800 ms poststimulus onset of both 'adventurer-NOM' and 'finally'. Overall ANOVAs showed complex interactions which corroborated these observations (see Table 2). Since these comparisons yielded complex interactions involving type (demonstrative vs. *wh*), we proceeded to examine pairwise comparisons between *in-situ* vs. scrambled demonstratives, and between *in-situ* vs. scrambled *wh*.

[INSERT TABLE 2 ABOUT HERE]

Pre-gap region: In-situ vs. scrambled demonstratives. Visual inspection of the waveforms suggested that scrambled demonstratives in comparison to *in-situ* demonstratives elicited greater left frontal negativity between 300-600 ms poststimulus onset of both 'adventurer-NOM' and 'finally', as well as a positivity near the end of both 'adventurer-NOM' and 'finally' (see Figure 4).

[INSERT FIGURE 4 ABOUT HERE]

To test the observed LAN effects in response to scrambled demonstratives, ANOVAs were performed in the 300-600 ms region of 'adventurer-NOM' (latency window: 300-600 ms) and in the 300-600 ms region of 'finally' (latency window: 950-1250 ms). At 'adventurer-NOM', there were significant [*midline*: $F(2, 38)=5.29$, $p = .029$[9]; *parasagittal*: $F(4, 76)=3.79$, $p = .03$] as well as marginal [*temporal*: $F(2, 38)=3.11$, $p = .06$] interactions between movement and anteriority, indicating that scrambled demonstratives were more negative at anterior regions compared to *in-situ* demonstratives. At 'finally', there was a marginal interaction between movement and hemisphere at temporal electrodes [$F(1, 19)=4.05$, $p = .059$]: the ERPs to scrambled demonstratives were more negative over the left hemisphere but more positive over the right hemisphere compared to *in-situ* demonstratives. Additionally, there was a marginal movement x hemisphere x anteriority interaction in the temporal [$F(2, 38)=3.12$, $p = .056$] array, because scrambled demonstratives compared to *in-situ* demonstratives were more positive over the right hemisphere, but more negative over the front and more positive over the back of the left hemisphere.

The observation of a late positivity (P600) in the scrambled demonstrative condition was for the most part supported by ANOVAs run in both the 500-800 ms region of 'adventurer-NOM' (latency: 500-800 ms) and in the 500-800 ms region of 'finally' (latency: 1150-1450 ms). For 'adventurer-NOM', there was a marginal main effect of movement in the scrambled condition at midline electrodes [$F(1, 19)=3.82$, $p = .065$]. For 'finally', there was a significant interaction between movement and hemisphere at temporal electrodes [$F(1, 19)=11.74$, $p = .003$], because scrambled demonstratives showed greater late positivity especially over the right hemisphere in response to the pre-gap region. In addition to the above, at 'finally', there was a significant movement x hemisphere x anteriority interaction at temporal electrodes [$F(2, 38)=3.61$, $p = .037$], due to the fact that scrambled demonstratives relative to *in-situ* demonstratives were consistently more positive over the right hemisphere, but more negative over the front and more positive over the back of the left hemisphere.

Pre-gap region: Scrambled vs. in-situ wh. Visual inspection of the waveforms showed an effect of left anterior negativity in response to sentences with scrambled *wh* between 300-600 ms poststimulus onset of 'adventurer-NOM' and 'finally' (see Figure 5). Unlike the demonstrative comparison, there was no apparent late positivity effect (see Figure 5).

[INSERT FIGURE 5 ABOUT HERE]

To test the LAN effect, ANOVAs were run in the 300-600 ms region of 'adventurer-N' (latency window: 300-600 ms). The ANOVAs revealed no significant movement-related effect, except for a movement x hemisphere x anteriority interaction in the temporal array [$F(2, 38) = 7.7$, $p = .002$], which appeared to be due to a frontal negativity and a posterior positivity in response to scrambled *wh* over the right hemisphere. Separate ANOVAs were run in the 300-600 ms region of 'finally' (latency window: 950-1250 ms). There was a significant [*temporal*: $F(2, 38)=4.76$, $p = .016$] as well as marginal [*parasagittal*: $F(4, 76)=2.61$, $p = .09$] interaction between movement and anteriority, indicating that scrambled *wh* in comparison to *wh-in-situ* conditions were more negative over the front than over the back of the head. In addition, there was a marginal movement x hemisphere x anteriority interaction at temporal electrodes [$F(2, 38)=2.94$, $p = .066$], because the ERPs to scrambled *wh* compared to *wh-in-situ* conditions were more negative over left anterior regions but more positive over right posterior regions.

Summary of pre-gap region. In summary, overall ANOVAs showed complex interactions among the factors of movement, type, hemisphere, and anteriority in response to the pre-gap region. When pairwise comparisons were examined, LAN effects were elicited in both scrambled vs. *in-situ* demonstrative and scrambled vs. *wh-in-situ* comparisons. Although LAN effects were visually apparent at both words for both comparisons, the effect at the subject noun position ('adventurer-NOM') for the *wh* comparison was not statistically supported. P600 effects were elicited at both the subject noun and adverbial positions for the scrambled vs. *in-situ* demonstrative comparison, whereas there was no apparent late positivity effect for the scrambled vs. *wh-in-situ* comparison.

Post-Gap Region

In order to test if there was any further gap-filling effect at the post-gap position in an SOV language like Japanese, single-word averages were made of ERPs to the post-gap region, 'discovered-Q' in Table 1. Visual inspection of the waveforms suggested an anterior negativity in response to both scrambled demonstrative and scrambled *wh* conditions starting around 300 ms poststimulus onset and continuing to the end of the epoch, although the impression was stronger for scrambled demonstratives. Overall ANOVAs performed in the latency window of 300-900 ms supported the observation. Corroborating the observed anterior negativity, there was a significant main effect of movement in the temporal array [$F(1,19) = 9.53$, $p = .006$], as well a significant interaction between movement and anteriority in the parasagittal array [$F(4,76) =$

3.68, p = .033]. Reflecting the observed difference between scrambled demonstratives and *wh*, there was a significant main effect of type in temporal electrodes [$F(1,19) = 6.63$, $p = .019$], as well as a significant [*midline*: $F(1,19) = 5.09$, $p = .036$] or marginal [*temporal*: $F(1,19) = 3.36$, $p = .082$] interaction between movement and type. Because of the statistical interactions involving type (demonstratives vs. *wh*), pairwise comparisons were made between *in-situ* vs. scrambled demonstratives and between *in-situ* vs. scrambled *wh.*

Post-gap region: Scrambled vs. in-situ demonstratives. The anterior negativity observed in response to scrambled demonstratives was tested by ANOVAs run in the latency window of 300 to 900 ms[10], and yielded a significant main effect of movement in all electrode arrays [*midline*: $F(1, 19)=8.30$, $p = .01$; *parasagittal*: $F(1, 19)=9.62$, $p = .006$; *temporal*: $F(1, 19)=21.94$, $p = .001$]. In addition, in the parasagittal array there was a marginal movement x anteriority interaction [$F(4, 76)=2.75$, $p = .074$]. This indicated that scrambled demonstratives tended to elicit greater negativity over anterior regions compared to *in-situ* demonstratives at the sentence-final verb position (Figure 6).

[INSERT FIGURE 6 ABOUT HERE]

Post-gap region: Scrambled vs. in-situ wh. Visual inspection, also for this comparison, suggested an anterior negativity starting around 300 ms poststimulus onset, especially over the left hemisphere, in response to scrambled *wh*. The observation of an anterior negativity was marginally supported by ANOVAs done in the latency window of 300 to 900 ms[11]. There was a marginal interaction between movement and anteriority at parasagittal electrodes [$F(4, 76)=2.39$, $p = .105$], indicating that scrambled *wh* elicited more negativity at anterior sites compared to *in-situ wh* at the sentence-final verb (see Figure 7).

[INSERT FIGURE 7 ABOUT HERE]

Summary of post-gap region. Both yes/no-questions with scrambled demonstrative objects as well as *wh*-questions with scrambled *wh*-objects elicited greater anterior negativity at the verb position relative to their *in-situ* counterparts, although only very marginally in the *wh*-comparison.

General Discussion

The results of our ERP study of filler-gap dependencies in Japanese were remarkably similar to the results of previous ERP studies of filler-gap dependencies in English and other European languages (see Table 3):

[INSERT TABLE 3 AROUND HERE]

Japanese questions containing scrambled object fillers elicited bilateral slow anterior negative potentials between filler and gap, P600 effects immediately following fillers and immediately preceding gaps (scrambled demonstrative condition only), phasic LAN effects in the gap region, and bilateral anterior negativity (AN) at the sentence-final verb position. As will be discussed in greater detail below, we assume that the slow negative potentials indexed storage of the filler in working memory, the P600 effects syntactic integration of the filler-gap dependency, the phasic LAN effects the retrieval of the filler from working memory for purposes of gap assignment, and the AN effects at the final verb an ongoing effect of working memory related to processing of a non-canonical structure. The overwhelming familiarity of these results is striking in the face of major differences between Japanese and English in typology (strictly head-final vs. head-initial languages), syntactic operation (scrambling via adjunction [or A-movement] vs. *wh*-[Ā-]movement via substitution), and apparent processing requirements: due to the head-final nature of Japanese, sentence processing in this language entails the need to accommodate massive amounts of syntactic ambiguity on a regular basis, while the processing of a head-initial language like English requires more limited amounts of occasional tolerance for syntactic ambiguity (see Inoue and Fodor, 1995). In the remainder of this section we discuss the nature of our results and their individual idiosyncrasies, compare them to the results of other related studies, and finally address their larger implications for sentence processing.

Slow Potentials Between Filler and Gap (Figure 2)

When the subject NP and following adverbial in the scrambled O **S** __V configuration were compared to the same four words in the canonical word order **S** O V configuration, the former showed greater bilateral, slow potential anterior negativity. The most ready interpretation of this effect was that the parser was holding an extra (object) NP in the O **S** __V configuration at the points of comparison, causing an extra verbal working memory load. These slow potential effects in response to questions with scrambled constituents were consistent with previous reports in the literature that such effects are caused by the necessity of holding a filler in working memory pending its assignment to a gap. This has been demonstrated both in English and in German in response to processes of *wh*-movement (i.e. *wh*-question and relative clause formation: King & Kutas 1995; Müller et al., 1997; Münte, Schwirtz, Wieringa, Matzke, & Johannes, 1997; Kluender & Münte, 1998; Fiebach et al., 2001), as well as in German in response to clause-internal scrambling (Rösler et al., 1998). The distribution of this effect has been inconsistent across studies, however: in the present study and in the King and Kutas study of English relative clauses, the effect had a symmetric anterior maximum; in Münte et al. (1997), which involved two separate replications of King and Kutas using English and German materials, the effect exhibited a symmetric centroparietal maximum; a left anterior maximum was reported in two independent studies of German *wh*-object vs. *wh*-subject questions (Kluender & Münte, 1998; Fiebach et al., 2001) and in the Rösler et el. study of clause-internal scrambling in German, while Müller et al., using auditory presentation to replicate King and Kutas in English, reported a right anterior maximum for the slow potential effect. Given this diversity in distribution of the slow negative potential effect across studies of filler-gap dependencies, we are not in a position to sort out the functional significance of its variability, but believe that our results are nonetheless fully compatible with those of previous studies.

A possible objection to our interpretation of the slow potential effect in our data is that since Japanese is a *pro*-drop language, i.e. it freely allows null pronominal elements, participants might have interpreted the accusative-marked NP object in the OSV configuration as following a null subject (*pro* **O**...V), rather than as a displaced object scrambled from its canonical position. Under such an interpretation, the nominative-marked NP, which was the point of comparison (O **S** __V), could have been interpreted as the subject of an embedded clause instead (*pro* O [**S**...] V). On this interpretation, the slow anterior negative potentials could have been caused by the need to open a new clause (see Garnsey et al., 2001, fn.3). Note however that the negative slow potential difference between scrambled and *in situ* conditions began already at the adjective ('reckless') of the subject NP ('that reckless adventurer-NOM'), i.e. before the nominative case-marking on the final head noun of the subject NP had been registered, and thus before the need to open a new clause even arose. Since determiners and modifiers are not case-marked in Japanese, at the adjective position ('reckless') it was not entirely clear that this NP would be marked nominative. Moreover, even if the nature of our stimulus materials led subjects to expect nominative-marked subjects following initial accusative-marked direct objects, Uehara and Bradley (1998) showed in their sentence completion study that Japanese speakers overwhelmingly prefer to interpret a nominative-marked NP preceded by an accusative-marked NP (O **S**....) as the overt subject of the highest clause. In other words, Japanese speakers will opt for scrambled word order when the alternative is an empty subject in canonical position in the highest clause, followed by an overt subject in an embedded clause (*pro* O [**S**....] V). Therefore, it seems reasonable to assume that participants in the present experiment anticipated the scrambled structure when presented with an initial accusative-marked NP, resulting in an increase in working memory load.

A further objection to the verbal working memory interpretation of the slow anterior negative potential could be based on the suggestion that our scrambled conditions were somewhat unnatural. In a corpus study of written Japanese texts (Yamashita, in press), more than 95% of the scrambled sentences in the corpus involved heavy syntactic elements like subordinate clauses and referential NPs. There were no scrambled *wh*-words in the corpus, but in general it seems clear that relatively light elements like demonstratives and *wh*-pronouns are rarely scrambled in written Japanese. Thus the slow anterior negative potential effect could have merely been a response to the unusual nature of our scrambled stimulus materials. However, a subsequent acceptability judgment study of the experimental materials[12] seems to militate against such an interpretation. The fact that light rather than heavy sentence constituents were scrambled in our design does not seem to have caused problems, as conditions with *in-situ* demonstratives and *wh-in-situ* received ratings of 3.1 and 2.7, respectively (five-point scale, average 2.9), remarkably close to those assigned to scrambled versions of the same sentences (both 3.0 on the same scale) [*in-situ demonstratives vs. scrambled demonstratives*: $t(19) = .373$, $p = .713$; *wh-in-situ vs. scrambled wh-pronouns*: $t(19) = -.780$, $p = .445$]. Thus in terms of general acceptability, there was no reliable difference in acceptability between scrambled (3.0) and *in-situ* (2.9) conditions [*in-situ vs. scrambled collapsed across demonstratives and wh*: $t(19) = -.780$, $p = .445$], and differences in acceptability therefore fails as a plausible explanation for the slow potential effects. Given the similarity of the slow potential effects in this study to those in other ERP studies of filler-gap dependencies, we will assume in what follows that they index differences in verbal working memory load across conditions.

This interpretation has the additional advantage of accounting for the discrepancy in results between our ERP study of Japanese scrambling and that of Mazuka et al. (2001), who reported no slow anterior negative potential effects between filler and gap in their comparison of OSV and SOV structures. We would like to suggest that this discrepancy may be due to a simple difference in length of the stimulus materials. The materials used by Mazuka et al. consisted of three-word sentences ('husband-ACC wife-NOM __ awaited'), while those in the present study consisted of sentences ten words long. Fiebach et al. (2001) showed that German *wh*-object fillers separated from their gaps by eight words (presented in five phrases) elicited more sustained left-lateralized anterior negativity across sentence positions than *wh*-object fillers separated from their gaps by only four words (presented in two phrases). While the shorter condition also elicited transient left anterior negativity at the first phrase following the filler (relative to the same phrase in a *wh*-subject question of equivalent length), this effect disappeared at the second intervening phrase, which immediately preceded the gap position. In a similar fashion, we intentionally lengthened the region between the scrambled object and its gap in our stimulus materials to include four words, the subject NP ('that reckless adventurer-NOM') and the adverb ('finally'). This manipulation appears to have effectively exacerbated the verbal working memory demands of our scrambled sentences, which in turn elicited the slow potential effects. Mazuka et al., on the other hand, seem to have intentionally kept their stimulus materials as short as possible, inserting no intervening material whatsoever between the scrambled object filler and the subject position preceding the gap, in effect keeping the working memory demands of their stimulus materials to a bare minimum. This interpretation further predicts that if Mazuka et al. had inserted intervening material between the scrambled object filler and its gap in their design, comparisons made at such intervening sentence positions would have also showed slow negative potential differences, as has been now demonstrated in numerous other studies.

P600 Effects in Response to Scrambled Demonstrative Fillers and Their Gaps (Figures 3 and 4)

In scrambled demonstrative questions, the word immediately following the scrambled demonstrative object (i.e. the first word of the subject NP) elicited a marked P600 effect relative to the same sentence position in the in-situ demonstrative condition (Figure 3). While this is to our knowledge the first reported instance of a P600 effect tied to the appearance of a filler, Mazuka, et al. (2001) also reported a P600 effect to the scrambled direct object NP ('husband-ACC') of their **O** S __V ('husband-ACC wife-NOM __ awaited') sentences. We are not sure why the P600 effect would be contemporaneous with the appearance of the filler itself in Mazuka et al., but delayed by one word in our data, but this may have had something to do with the fact that the scrambled direct objects in Mazuka et al.'s design were full referential noun phrases, while ours were instead out-of-the-blue, closed-class (and therefore light) anaphoric demonstratives.

Additionally, we observed an effect of late positivity both at the last word of the subject NP (i.e. the head noun, 'adventurer-NOM') and at the adverb ('finally') of the scrambled demonstrative condition (Figure 4). As discussed at greater length in the next section, we were unable to predict the location of the gap in our scrambled sentences with 100% accuracy. For this reason, we cannot say with absolute certainty whether both of these sentence positions constituted the pre-gap region, or whether the head noun of the subject NP constituted the pre-gap region, and the adverbial the post-gap region. However, this is not of major consequence in this case, since P600 effects related to gap-filling have been elicited in response to the processing of both the pre-gap (Kaan et al., 2000) and the post-gap region (Phillips, Kazanina, Wong, & Ellis, 2001) of object *wh*-questions in English. In addition, Fiebach et al. (2001) reported P600 effects (superimposed on a slow left anterior negative potential, much like in the present study) at the subject position immediately preceding the gap of verb-final German *wh*-object questions. Most closely related to the present study, Mazuka et al. (2001) reported P600 effects not only in response to the sentence-initial scrambled object NP ('husband-ACC') of their experimental sentences ('husband-ACC wife-NOM __ awaited'), as discussed in the preceding paragraph, but also to the immediately adjacent subject NP preceding the gap ('wife-NOM __ ').

There are two arguments for interpreting these P600 effects as an index of syntactic integration costs rather than verbal working memory costs, as first proposed by Kaan et al. (2000). First, the entry of a filler into working memory seems to be reliably indexed by the onset of a slow negative potential, rather than a P600 effect, as demonstrated in the many ERP studies of filler-gap dependencies cited in the previous section. Second, and more crucially, recall that the filler and its gap were separated from each other in our data by at least three words ('that reckless adventurer-NOM [__] finally __'), while in Mazuka et al. (2001), the filler and pre-gap positions were immediately adjacent to each other ('husband-ACC wife-NOM __ awaited'). In the previous section, we attributed the lack of a slow negative potential effect in the Mazuka et al. study to this difference in the length of the stimulus materials. Note, however, that despite these differences in length, the P600s elicited by scrambled fillers and their gaps were quite similar across the two studies. Likewise, Fiebach et al. (2001) reported that the subject NP preceding the gap elicited a P600 effect in both their long and their short *wh*-object conditions. P600 effects seen immediately following (or, in Mazuka et al.'s data, simultaneous with) a filler and immediately preceding a gap thus seem to be independent of filler-gap distance, and hence of verbal working memory demands. The most parsimonious conclusion, then, based on the results of previous ERP studies as well as the present study, is that P600 effects in these contexts must be related to the necessity of incorporating a filler-gap dependency into the syntactic parse, rather than an index of verbal working memory demands *per se*.

Why we did not elicit similar effects in our scrambled *wh*-condition is not entirely clear, but the upside of this is that our study and that of Mazuka et al. (2001) were consistent in the elicitation of P600 effects in response to scrambled NP fillers and their gaps. Note further that so far, P600s in response to (scrambled) fillers have been elicited only in Japanese sentence processing; prior to this, P600 effects had been observed only around gap positions, and only in European languages (Kaan et al., 2000; Phillips et al., 2001; Fiebach et al., 2001). Whether this is a reliable cross-linguistic difference is something that must be determined by further research.

(L)AN Effects in the Gap Regions (Figures 4 through 7)

There were two sentence positions relevant for potential gap-filling effects, namely both before and after the posited gap position. However, as mentioned in the previous section, given the relatively free word order of Japanese, it was difficult to pinpoint the exact location of the gap in our experimental sentences. Examples of our *in-situ* and scrambled demonstrative conditions are repeated for convenience in (6a) and (6b), respectively.

(6)
a. *in-situ* demonstrative

The local newspaper-to according that reckless adventurer-NOM finally ***that-ACC*** discovered-Q.

'According to the local newspaper, did that reckless adventurer finally discover that?'

b. scrambled demonstrative

GAP FILLING?
↓
The local newspaper-to according ***that-ACC*** that reckless adventurer-NOM (__) finally__discovered-Q.

'According to the local newspaper, did that reckless adventurer finally discover that?'

As (6b) demonstrates, the position of the gap was ambiguous: it could have either preceded or followed the adverb. As discussed in the Results section, an active filler strategy (Frazier & Clifton, 1989) approach to gap-filling predicts that the parser would have tried to dispose of the filler as quickly as possible, and the earliest possible place to posit an object gap would have been at the head noun of the subject NP ('adventurer-NOM'). However, there would have been nothing to prevent the parser from confirming this assignment at the following adverbial ('finally'). In a cross-modal priming study, Nicol (1993) reported filler reactivation at two separate sentence positions: the subcategorizing verb, as proposed by the direct association hypothesis (Pickering & Barry, 1991; Pickering, 1993), and again at the purported gap position later in the string. It thus seems plausible there would have been multiple attempts at gap-filling in response to our stimuli as well. And indeed, there were P600 effects and LAN effects alike at both the head noun of the subject NP and again at the adverb.

The effects at the sentence-final verb appeared to be different in character. First, there were no P600 effects in the ERPs to the verb. This was the pattern of results in Mazuka et al. (2001) as well: although both of the first two words in their three-word scrambled OSV sentences elicited P600 effects, there were no significant effects at all at the final verb. Second, while the LAN effects in response to the head noun of the subject NP and to the adverb in the present study were clearly left-lateralized, the effects of anterior negativity to the final verb were not. Therefore, while we are ultimately unsure how to divide three sentence positions (head noun of the subject NP, adverb, and sentence-final verb) into two well-defined regions, one preceding the gap and one following it, in what follows we will treat the LAN effects at the head noun of the subject NP and the adverb ('adventurer-NOM finally') separately from the effects of anterior negativity at the final verb ('discovered-Q'), and consider them qualitatively different.

Summarizing the results in the gap region, scrambled demonstratives elicited LAN effects (which were followed by P600 effects, as discussed in the previous section) at both the subject noun and adverbial positions, while scrambled *wh* elicited LAN effects (but no subsequent P600 effect) at the adverbial position only. At the final verb, both scrambled conditions elicited anterior negativities; however, the response to the final verbs of scrambled *wh* was statistically tenuous.

One dilemma that we face here is that scrambled demonstratives yielded more robust and reliable effects than did scrambled *wh*. We have no ready explanation for why this should be the case. However, our acceptability study again assured us that these differences were not caused by differences in acceptability: questions containing scrambled demonstratives received the exact same acceptability rating as those containing scrambled *wh*, namely 3.0. At this point, the reason for the quantitative differences in reliability of effects between scrambled demonstratives and scrambled *wh* is not transparent, and we will continue to investigate it. However, since both scrambled conditions elicited similar (L)AN effects relative to the corresponding *in-situ* conditions, we shall treat these as equivalent in what follows.

Another concern is that LAN effects have typically been reported at post-gap positions in previous studies (Kluender & Kutas, 1993a,b; King & Kutas, 1995) rather than at pre-gap positions. In the present study, we obtained somewhat the opposite pattern of results: the head noun of the subject NP ('adventurer-NOM') certainly preceded the gap, and elicited both LAN and P600 effects, the adverb ('finally'), which could have either preceded or followed a gap, also elicited both LAN and P600 effects, but the final verb ('discovered-Q'), which must have followed the gap, instead elicited bilateral anterior negativity that continued throughout the epoch. What should be made of this pattern of effects is unclear, and for lack of a better argument we will merely assume that if P600 effects can apparently be elicited at both pre-gap (Kaan et al., 2000) and post-gap positions (Phillips et al., 2001) across studies, there is no obvious reason why LAN effects should not exhibit the same variability.

This raises the further issue of the precise nature of the phasic (L)AN effects reported around the gap region: were these merely shorter time slices of (an increase in) the longer-lasting slow negative potential ranging across several sentence constituents, or were they truly phasic effects specifically elicited by particular words in the string over and above the slow potential effect? The answer to this question will vary according to sentence position. First, as to LAN effects in response to the subject head noun ('adventurer-NOM') and the adverb ('finally') of scrambled conditions, the answer must be yes: these anterior negativities were truly lateralized to the left hemisphere (Figures 4 and 5), as evidenced by interactions between condition, anteriority, and hemisphere in the temporal array that just missed significance. The slow anterior potential effect (Figure 2), on the other hand, was clearly symmetric, showing interactions only with anterior to posterior electrode position, but not with hemisphere. If the phasic effects at the subject head noun and the adverb were merely the result of (an increase in) the slow anterior negative potential, one would expect them to be bilateral as well. Second, the negativity to the subject head noun ('adventurer-NOM') in scrambled demonstratives (Figure 4) showed every sign of being a phasic effect: at left frontal electrodes (Fp1, F3, and F7), the ERPs to the subject head noun of the two conditions aligned almost perfectly at the N100 and P200 peaks, subsequently diverged in the N400 peak, and then realigned in the N100 and P200 components of the following word. If this difference were merely due to pre-existing differences in the slow negative potential, it would not show this realignment of componentry at the onset of the adverbial 'finally'. Note that, even though the second argument from componentry realignment does not go through for the LAN effects in response to the adverbial (Figures 4 and 5), the first argument from left-lateralization still holds, which we believe suffices to distinguish them from the slow negative potential effects.

At the sentence-final verb, however, the argument from left-lateralization also fails: the phasic effects of anterior negativity elicited in response to the final verb of scrambled sentences were not reliably left-lateralized. Thus the AN to the sentence-final verb could indeed have been caused by the slow negative potential seen in response to earlier sentence positions (Figure 2), and as such would not itself be a gap-filling effect. Under this scenario, we would be looking at something entirely different from gap-filling at the final verb position. But even if this is the case, we find it nonetheless significant that there is any difference at all at the final verb, for reasons to be elucidated below.

Before addressing what the AN to the final verb actually indexed, however, we can say with a fair degree of confidence what the effect at sentence end was not. It was not an anticipatory response to a following comprehension question for two reasons. First, the majority (80%) of stimulus sentences that participants saw did not have comprehension questions associated with them, and there would be no reason to expect any difference between the scrambled and *in-situ* conditions on this dimension, as both were followed by the same total number of comprehension questions, i.e. 20% of the time. Moreover, the anterior negativity was unlikely a mere sentence wrap-up effect as seen in previous ERP studies (Osterhout & Holcomb, 1992; Hagoort, Brown, & Groothusen, 1993), which have consistently been reported as an N400 difference: neither the morphology nor the distribution of our sentence-final effects was consistent with that of an N400 effect (see Figures 6 & 7). This ERP effect at sentence end therefore seems more likely to have something to do with the scrambling of prior constituents.

As discussed earlier, the bilateral slow anterior negative potentials elicited between filler and gap in scrambled conditions could have simply been an effect of holding an extra NP object in working memory at the point of comparison (O **S** __V vs. **S** O V), rather than a scrambling effect *per se*. However, the comparisons undertaken at the sentence-final verb position of the scrambled conditions provide clearer evidence that scrambling operations have their own associated processing costs independent of the number of NP arguments being currently held in the memory store. The sentence-final effects of anterior negativity in response to scrambling were not confounded by comparing different sentence positions or different lexical items, and cannot be attributed to different numbers of NP arguments being held in working memory across conditions, either. By this point, the parser had had to hold the same number of NPs in both conditions, and the negative anterior difference at sentence end can therefore only have been a consequence of scrambling itself.

In their self-paced reading time study, Mazuka et al. (in press) found that the subject of scrambled O **S** __V configurations took longer to read than the object of canonical S **O** V configurations, which also indicated that scrambling itself incurs a processing cost, independent of the number of NPs that must be held in working memory. Pechmann et al. (1994) and Rösler et al. (1998) likewise demonstrated costs of scrambling in acceptability, comprehension response time, and ERP studies using similar comparisons in German. However, here again we face a discrepancy with the results reported by Mazuka et al. (2001) in their ERP study of Japanese scrambling. As discussed earlier, Mazuka et al. (2001) elicited no slow anterior negative potential differences between filler and gap. They likewise elicited no phasic LAN effects at gap positions, and no differences of anterior negativity at the sentence-final verb of scrambled sentences. We surmise that the explanation for the lack of phasic LAN effects and of anterior negativities at sentence end will be the same as that for the lack of slow negative potentials in their data, namely, differences in length of the stimulus materials. In other words, we believe it is consistent to claim that all effects of negative voltage deflection in our data were tied in one way or another to the verbal working memory demands of our scrambled sentences: the slow negative potentials to the entry and maintenance of a filler in working memory, the phasic LAN effects to the retrieval of this filler from working memory for purposes of gap assignment, and the anterior negativities at sentence end to the global working memory demands of having had to process a non-canonical scrambled structure. This latter point will be discussed at greater length in the following section.

We think that this interpretation of our results can also help to make sense of a long-standing conundrum in the ERP literature regarding the functional significance of LAN vs. P600 effects. Recall that in our data, both effects were elicited in response to the subject head noun ('adventurer-NOM') of our scrambled demonstrative condition. We have interpreted these in terms of the costs associated with retrieving a filler from verbal working memory for purposes of gap-filling (LAN) and syntactically integrating this filler-gap relationship in the ongoing parse (P600). This is far from the first time that this particular pattern of results has been observed, however. In fact, the very first study of P600 effects (Osterhout & Holcomb, 1992) reported the exact same combination of effects (i.e. LAN plus P600) in response to *was* in the sentence **The broker hoped to sell the stock was sent to jail*. As originally suggested in Kluender and Kutas (1993b: 622-624), the P600 can in this case be interpreted as an index of syntactic processing difficulty, while the LAN can be interpreted as a desperate (and ultimately futile) attempt on the part of parser to seek an appropriate, previously occurring discourse referent in verbal working memory to assign as the subject of the verb *was*. This interpretation is also consistent with the results reported in Urbach (1993) in response to *was* in the garden path sentence *The cook helped in the kitchen was busy*. In this instance, the elicited LAN effect would index a (this time) successful attempt on the part of the parser to locate an appropriate, previously occurring discourse referent for the verb *was* in verbal working memory, while the P600 effect would reflect the syntactic processing difficulty involved in reanalysis of the string 'helped in the kitchen' as a reduced relative clause. If this account is accurate, it would provide a fortuitous index of simultaneous but dissociable working memory and syntactic integration costs in the ERP record (cf. Gibson, 1998).

Larger Implications

Slow anterior negative potentials are familiar from previous studies of filler-gap dependencies (King & Kutas, 1995; Münte et al., 1997; Müller et al., 1997; Kluender & Münte, 1998; Fiebach et al., 2001), as are LAN effects at post-gap positions (Kluender & Kutas, 1993 a, b; King & Kutas, 1995), and P600 effects at gap positions have been reported more recently (Kaan et al., 2000; Phillips et al., 2001; Fiebach et al., 2001). As noted earlier, Mazuka et al. (2001) also elicited P600 effects tied to the scrambled filler in Japanese **O** S V contexts. In this sense our results simply confirm what has generally come to be expected in filler-gap contexts. However, it should at the same time be pointed out that it is not entirely obvious that Japanese should necessarily show psychologically real effects of storing fillers, maintaining filler-gap dependencies, or filling gaps in scrambling contexts. In our experimental materials, we in essence simply permuted two adjacent sentence constituents, placing the object NP in front of the subject NP (which we had however intentionally lengthened). Since Japanese is a case-marking language, the grammatical functions of these two NPs, and consequently their respective thematic roles, were easy to determine, as we never presented examples of sentences with unusual linking between thematic roles and grammatical functions. Nonetheless, the ERPs to our scrambled sentences showed very real consequences of this simple permutation: the effects we elicited seemed to indicate that the brains of our participants were actively engaged in attempts to assign the scrambled NP to a gap position and thereby put things back in canonical SOV order, even though it is not completely obvious that this would be necessary to insure successful sentence comprehension.

These results are incompatible with a model of sentence processing based on head-driven parsing (Pritchett, 1992). Specifically, phasic LAN and P600 effects elicited at the subject head noun and adverb are difficult to account for under this model, as it predicts no effects of gap-filling before the sentence-final verb. More generally, the pattern of ERP effects in this study associated with entering scrambled fillers in working memory, holding them in the memory store, and assigning them to gap positions cannot be attributed to thematic ambiguity of the filler, as suggested by King and Kutas in their (1995) study of relative clauses in English. Since Japanese is strictly head-final, on the assumptions of a head-driven parser, *all* NPs, scrambled or otherwise, remain thematically ambiguous until sentence end, and therefore the order in which they were received should not make any particular difference in thematic role assignment.

Aside from this model, our results seem consistent with a variety of processing models pertaining to filler-gap dependencies, including the active filler strategy (Frazier & Clifton 1989), the minimal chain principle (De Vincenzi, 1991), and proposals related to verbal working memory (e.g. Schlesewsky, Fanselow, Kliegl, & Krems, 2000). However, since these models link the problem of filler-gap resolution to structural positions (e.g. gaps, members of a syntactic chain, or phrase structural projections), they do not have anything direct to say about end-of-sentence effects. In other words, if filler-gap processing is tied directly to structural positions, then once the relevant structural positions have been parsed, one might reasonably expect processing difficulty to end: the region of difficulty ought to be localized to the span of the filler-gap dependency itself. Yet it has been known since early ERP studies of filler-gap dependencies (Kluender & Kutas, 1993a; King & Kutas, 1995) that while extra processing costs generally begin after the filler has been entered into the parse, they continue well past the gap position, the end of the syntactic chain, or the subcategorizing verb, as the case may be. Thus, even though filler-gap models based on structural positions are certainly compatible with residual end-of-sentence effects, they do not predict them.

An alternative possibility worth considering is a notion of "canonicity" or "canonical word order", which may be defined as the most statistically frequent and contextually neutral word order (Comrie, 1981). In Japanese, scrambled sentences are reported to occur very infrequently, less than 1% of the time in transcriptions of magazine articles in various registers of formality (N = 2,635: Yamashita, in press). In regard to contextual neutrality, as discussed earlier, Yamashita reported that 95% of the scrambled sentences in her corpus could be accounted for either by reference to the immediately preceding context or by "heaviness" in terms of syntactic complexity. Thus it seems safe to conclude that scrambled word orders are indeed non-canonical in Japanese. It is possible that the human parser simply prefers canonical word order as a path of least effort and resistance, and further that it can only comprehend sentences with non-canonical word order by linking "displaced elements" with their canonical positions[13]. However, on this view, non-canonical order is a more global, constructional problem not tied to specific structural positions: once the parser recognizes that it is not dealing with a canonical structure, it continues to bear the resulting processing costs all the way through to sentence end. In ERP studies of filler-gap dependencies in English and German, it certainly looks as if the brain continues to fret all the way home about having had to deal with non-canonical word order. Interestingly, Münte, Schiltz, and Kutas (1998) reported that sentences like *Before the scientist submitted the paper, the journal changed its policy*, in which events are described in reverse chronological order, elicited slow left-lateralized anterior negative potentials compared to sentences with events described in chronological order of occurrence like *After the* *scientist submitted the paper, the journal changed its policy*. Here the processing problem cannot be one of linking positions in a syntactic phrase structural configuration, since both sentences exhibit identical phrase structures, but must instead be related to expectations of canonical temporal order in event structure. Moreover, the amplitude of this slow potential difference was highly correlated with working memory span: those with high memory spans showed much larger effects than those with low memory spans. Presumably, high span readers have the working memory capacity required to put the two clauses of *before* sentences back in canonical chronological order, as indexed by the slow potentials in their ERPs, while low span readers are already at ceiling and show little evidence of carrying out the same operation (Carpenter, Miyake, & Just, 1995). In light of these findings, we suspect that models of sentence processing that can accommodate notions of canonicity may be of most use in making sense of the ERP findings related to filler-gap dependencies as well.

Conclusion

Despite substantial differences in language family, orthography, basic word order, and the potential for syntactic ambiguity, this study uncovered no entirely new or unexpected ERP components in response to Japanese language materials. One may thus conclude that language-related ERP components are universal across languages: despite some discrepancies, the results of this experiment in general followed similar patterns seen previously in English and German, and now also in Japanese. The word immediately following scrambled demonstratives elicited a P600; this was similar to Mazuka et al.'s (2001) report of a P600 at the scrambled NP direct object of **O** S __V sentences. Slow anterior negative potentials were elicited by the words intervening between scrambled elements and their canonical positions, similar to slow potential effects reported in King and Kutas (1995), Münte et al. (1997), Kluender and Münte (1998), and Fiebach et al. (2001). At the head noun of the subject NP, both phasic LAN and P600 effects were seen. The latter were reminiscent of P600 effects at the pre-gap verb position of *wh*-movement sentences (Kaan et al., 2000). Phasic LAN effects have not been previously reported at pre-gap positions, but may well be related to post-gap LAN effects reported in English (Kluender & Kutas, 1993a, b; King & Kutas, 1995), and an effect of anterior negativity persisted at the sentence-final verb, as has also been the case in earlier English studies. To our knowledge, this is the first demonstration that this complex of documented object gap-filling ERP effects can be elicited in scrambling contexts as well as in standard *wh*-movement contexts.

This seems a fairly remarkable result. On the one hand, it simply provides further evidence for a model of gap-filling that is already widely accepted. On the other hand, there is no *a priori* reason why the Japanese parser should be required or expected to adhere to this model, and the fact that it does is heartening news, in that one would certainly hope that the human parser would operate on the basis of universal parsing principles. At the same time, the evidence from Japanese allows us to refine our universal theories of parsing: it suggests that processing difficulties associated with filler-gap dependencies across the world's languages may be due to deviations from canonical representations of word order, based in part on statistical frequencies of occurrence, and possibly even derived from more general representations of canonicity in language.

References

Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: Evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.

Basilico, D., Piñar, P., & Anton-Méndez, I. (1995). Canonical word order and the processing of verbal traces in Spanish. *The Proceedings of the Eighth Annual CUNY Conference on Human Sentence Processing,* Tucson, Arizona.

Carpenter, P.A., Miyake, A., & Just, M.A. (1995). Language comprehension: sentence and discourse processing. *Annual Review of Psychology*, 46, **91**-120.

Caplan, D., Alpert, N., & Waters, G. (1998). Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. *Journal of Cognitive Neuroscience*, **10**(4), 541-552.

Caplan, D., Alpert, N., & Waters, G. (1999). PET studies of syntactic processing with auditory sentence presentation. *Neuroimage*, **9**(3), 43-51.

Caplan, D., Alpert, N., Waters, G., & Olivieri, A. (2000). Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Human Brain Mapping*, **9**(2), 65-71.

Comrie, B. (1981). *Language Universals & Linguistic Typology*. Chicago: The University of Chicago Press.

De Vincenzi, M. (1991). Filler-gap dependencies in a null subject language: Referential and nonreferential WHs. *Journal of Psycholinguistic Research*, **20**, 197-213.

Fiebach, C.J., Schlesewsky, M., & Friederici, A.D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, **30**, 321-338.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, **9**, 427-473.

Frazier, L. & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes*, **4**, 93-126.

Garnsey, S., Yamashita, H., Ito, K., & McClure, M. (2001). *What when the verb at the end is happens? An ERP study of Japanese sentence comprehension.* Poster session presented at the 14th annual CUNY conference on human sentence processing, Philadelphia, PA.

Gibson, E. (1990). Recency Preference and Garden-Path Effects. In *Program of the Twelfth Annual Conference of the Cognitive Science Society* (pp. 372-379). Hillsdale, NJ: Lawrence Erlbaum.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, **68**, 1-76.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language & Cognitive Processes*, **8** (4), 439-483.

Holmes, V.M., & O'Regan, J.K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior*, **20**, 417-430.

Inoue, A. & Fodor, J.D. (1995). Information -paced parsing of Japanese. In R. Mazuka & N. Nagata (Eds.), *Japanese Sentence Processing*. Hillsdale, NJ: Erlbaum.

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., & Thulborn, K.R. (1996). Brain activation modulated by sentence comprehension. *Science*, **274**, 114-116.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, **15**(2), 159-201.

Keppel, G. (1982). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

King, J.W. & Just, M.A. (1991). Individual differences in syntactic: The role of working memory. *Journal of Memory and Language*, **30**, 580-602.

King, J.W. & Kutas, M. (1995). Who did what to when?: Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, **7**(3), 376-395.

Kluender, R. & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, **5**(2), 196-214.

Kluender, R. & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, **8**(4), 573-633.

Kluender, R., Münte, T.F. (1998). *ERPs to grammatical and ungrammatical wh-questions in German: Subject/object asymmetries.* Poster session presented at the 11th annual CUNY conference on human sentence processing, Newark, NJ.

Love, T., & Swinney, D. (1998). The influence of canonical word order on structural processing. *Syntax and Semantics*, **31**, 153-166.

MacWhinney, B., & Pleh, Cs. (1988). The processing of restrictive relative in Hungarian. *Cognition*, **29**, 95-141.

Mahajan, A.K. (1992). *The A/Ā-distinction and movement theory*. Doctoral dissertation, Massachusetts Institute of Technology.

Mazuka, R., Itoh, K., & Kondo, T. (2001). *Event-related potentials for scrambled word order in Japanese*. Poster session presented at the 14th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.

Mazuka, R., Itoh, K., & Kondo, T. (In press). Cost of scrambling in Japanese sentence processing. In M. Nakayama (Ed.), *Papers from International East Asian Psycholinguistics Workshop*. Stanford, CA: Center for the Study of Language and Information.

Müller, H., King, J. W., & Kutas, M. (1997). Event related potentials elicited by spoken relative clauses. *Cognitive Brain Research,* **5**, 193-203.

Münte, T.F., Schwirtz, O., Wieringa, B.M., Matzke, M., & Johannes, S. (1997). Elektrophysiologie komplexer Sätze: Erignskorrelierte Potentiale auf der Wort- und Satzebene. EEG-EMG, **28**, 11-17.

Münte, T.F., Schiltz, K., & Kutas M. (1998). When temporal terms belie conceptual order. *Nature*, **395**, 71-73.

Nakagome, K., Takazawa, S., Kanno, O., Hagiwara, H., Nakajima, H., Ito, K., & Koshida, I. (2001). A topographical study of ERP correlates of semantic and syntactic violations in the Japanese language using the multichannel EEG system. *Psychophysiology*, **38**, 304-315.

Nicol, J. L. (1993). Reconsidering reactivation. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting* (pp. 321-347). Mahwah, NJ : Erlbaum.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, **31**, 785-806.

Pechmann, T., Uszkoreit, H., Engelkamp, J., & Zerbst, D. (1994). *Word order in the German middle field*. Report 43, Computational Linguistics at the University of Saarland.

Phillips, C., Kazanina, N., Wong, K., & Ellis, R. (2001). *ERP evidence on the time course of processing demands in wh-dependencies*. Poster session presented at the 14th annual CUNY conference on human sentence processing, Philadelphia, PA.

Pickering, M. & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, **6**(3), 229-259.

Pickering, M. (1993). Direct association and sentence processing: A reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes*, **8**(2), 163-196.

Pritchett, B. (1992). *Grammatical competence and parsing performance*. Chicago: The University of Chicago Press.

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: Word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, **38**(n2), 150-176.

Saito, M. (1985). *Some asymmetries in Japanese and their theoretical implications*. Doctoral dissertation, Massachusetts Institute of Technology.

Saito, M. (1992). Long distance scrambling in Japanese. *Journal of East Asian Linguistics*, **1**, 69-118.

Schlesewsky, M., Franselow, G., Kliegl, R., & Krems, J. (2000). The subject preference in the processing of locally ambiguous wh-questions in German. In B. Hemforth and L. Konieczny (Eds.), *German Sentence Processing* (pp. 65-93). Dordrecht, Netherlands : Kluwer Academic.

Stamenov, M., & Andonova, A. (1997). Lexical access and co-reference processing in Bulgarian. In D. Hillert (Ed.), *Cross Linguistic Studies of Language Processing*. San Diego: Academic Press.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, **52** (3), 452-473.

Uehara, K. & Bradley, D. (1998). *Processing scrambled argument structures in Japanese*. Paper presented at the annual meeting of the Linguistic Society of America, New York.

Urbach, T.P. (1993). *Access to non-syntactic information during parsing: Evidence from event-related potentials (ERPs) in garden-path sentences.* Poster session presented at the 6th annual CUNY conference on human sentence processing, Amherst, MA.

Welbelhuth, G. (1989). *Syntactic Saturation Phenomena and the Modern Germanic Languages*, Doctoral dissertation, University of Massachusetts.

Yamashita, H. (1997). The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research*, **26**(2), 163-188.

Yamashita, H. (in press). Scrambled sentences in Japanese: Linguistic properties and motivations for production. *TEXT*.

Footnotes

[1] [1] There has been a debate in the literature over the years (Saito, 1985, 1992; Welbelhuth, 1989; Mahajan, 1992) as to whether scrambled elements occupy positions assigned thematic roles (a so-called A[rgument]-position) or not (a so-called Ā-position). To oversimplify this debate slightly for purposes of exposition, Mahajan and Saito (1992) proposed that clause-internal scrambling can be either to A- or to Ā-positions, while clause-external scrambling is of necessity only to Ā-positions. This debate is largely irrelevant to our concerns here, however, as we were interested in the processing effects of syntactically displacing constituents (including *wh*-constituents) from their canonical word order positions in the surface representation of sentences by whatever means. We will therefore refer to any such displaced element in Japanese as having been "scrambled", without intending any particular theoretical claims by doing so. Since *wh*-elements in Japanese generally remain *in situ*, i.e. in their canonical positions, and do not undergo processes of *wh*-movement to clause-peripheral positions as do *wh*-phrases in languages like English, we will also refer to syntactically displaced *wh*-elements in Japanese as having been "scrambled" in what follows. Scrambled *wh*-objects were originally included in the experimental design to most closely approximate the manipulations undertaken in previous ERP studies of filler-gap dependencies in European languages via *wh*-movement. As it turned out, however, this would not have been necessary, since our clearest results came from scrambled demonstrative phrases anyway (see Results and Discussion sections).

[2] These effects were essentially replicated in Müller, King, and Kutas (1997) using an auditory version of the King and Kutas (1995) stimulus materials: object relative clause sentences elicited greater anterior negativity starting from the relative clause region (Regions 1 and 2 in King & Kutas) and becoming noticeably larger after the gap (Region 3 in King & Kutas). However, these slow potentials were right- instead of left-lateralized, as had been the case in King and Kutas, in which stimuli were presented visually. The authors speculated that this was probably due to some mechanism of auditory perception, such as more involvement of the right hemisphere.

[3] Garnsey, Yamashita, Ito, and McClure (2001) also conducted a Japanese ERP study examining sentences with semantic violations, inflectional violations, and clausal reanalysis. In the clausal reanalysis comparison, biclausal sentences with the structure S-IO-[OSV]-V (in

which the embedded clause direct object is scrambled) were compared to similar sentences with the structure S-IO-[SOV]-V (in which the embedded clause direct object is in canonical position). The embedded clause subjects of both word orders elicited greater bilateral anterior negativity, and the embedded clause subject of the scrambled OSV condition elicited a P600. Garnsey et al. interpreted the bilateral negativity as indexing the opening of a new clause, and the P600 as indexing a garden path effect: participants were led to believe that the scrambled direct object in the OSV condition belonged to the main clause (i.e. canonical word order S-IO-DO) and had to revise this interpretation when the embedded subject came in.

[4] Also see Nakagome et al. (2001) for an earlier ERP study of Japanese sentence processing. Nakagome et al. examined the processing of mismatches in semantic selectional restrictions and *wh*-[±]Q particles. Semantic violations elicited a larger N400 relative to well-formed control sentences, while *wh*-Q violations yielded a P600 effect.

[5] The presentation rate of 650 ms per *bunsetsu* is longer than the standard 500 ms per word for English, yet was decided on as the best rate after consulting three native speakers of Japanese. Similarly, Garnsey et al. (2001) used 700 ms per *bunsetsu* in their Japanese ERP study.

[6] Following a modified Bonferroni procedure (Keppel, 1982), an alpha level of .05 was also used for the pairwise comparisons.

[7]Although the subject NP /adverb complex between filler and gap was preceded by words with roughly equivalent ERPs in three out of four conditions (i.e. 'according' in the *in-situ* conditions, and 'that-ACC' in the scrambled demonstrative condition), and therefore could have served as a suitable prestimulus baseline, the scrambled *wh*-objects 'what-ACC' in the scrambled *wh*-condition elicited a larger P200. Since a pre-stimulus baseline of 400 ms would have included this difference between conditions in the P200 region, measurements were taken with a 100 ms post-stimulus baseline instead, which allowed reliable alignment of the post-stimulus onset ERPs.

[8] There were also marginally significant differences in the P200 region (150-250 ms) at midline electrodes [$F(1, 19) = 4.02$, $p = .059$]. However, we believe that this cannot be the sole cause of the P600 effect in response to this word of the sentence. First, there were also significant P600 effects, but no accompanying significant P200 effects, in the parasagittal array

in response to the same word. Second, at Fz in the midline, the ERPs to the two conditions diverge in the P200 region, realign in the N400 region, and then diverge again in the P600 region. In addition, there was an effect of scrambling on mean amplitudes at Fz that just missed significance in the P600 region (500-800 ms) [$t(19) = 1.68$, p =.054], whereas there was no significant or marginal effect of scrambling in the P200 region (150-250 msec) [$t(19) = 1.19$, p = .124].

[9]The movement x anteriority interaction at midline electrodes was probably due to P600 effects rather than LAN effects. As seen in Figure 3, the midline region shows clear positivity at posterior sites starting around 400 ms poststimulus onset of ‘adventurer-NOM’.

[10] 300-900 ms is longer than the typical latency window for LAN effects (300-500 or 300-600 ms), yet was chosen since there was no word following the sentence-final verb. To be consistent with the standard window, separate tests were also run in the latency window of 300-600 ms. In the 300-600 ms window, there was a significant [*temporal*: $F(1, 19)$=14.23, p < .001] and marginal [*midline*: $F(1, 19)$=3.31, p = .085] main effect of movement, indicating that scrambled demonstratives elicited greater negativity at the final verb.

[11] Again, separate ANOVAs were run in a latency window of 300-600 ms, the standard latency window for LAN effects. In the 300-600 ms window, there was a marginal interaction between movement and anteriority at parasagittal [$F(4, 76)$=2.60, p = .072] electrodes, indicating that scrambled *wh* elicited greater negativity over the front of the head, but greater positivity over the back of the head relative to *wh-in-situ.*

[12] We ran a posthoc acceptability judgment study on some of the experimental materials. Eight native Japanese speakers (different from those for the ERP study) were instructed to read silently 20 stimulus and 20 filler sentences and rate their “naturalness” on a five-point scale (5: most natural-1: least natural) by marking scantrons.

[13] On the basis of evidence from cross-modal priming studies, Love and Swinney (1998) suggest that the role of canonical word order in filler-gap dependencies may vary cross-linguistically. In particular, cross-modal priming results from Spanish (Basilico, Piñar, & Anton-Méndez, 1995) suggest that the parser attempts to conform to an underlying SVO order, just as it does in English, even though Spanish surface word order is more flexible than that of English. On the other hand, Bulgarian object relative clauses do not show standard reactivation effects at

gap positions (Stamenov & Andonova, 1998), and this is attributed to the fact that Bulgarian allows much freer word order in such constructions. Since we have attributed the ERP results of this study to the attested (Yamashita, in press) low-frequency occurrence of scrambling constructions in Japanese, the account of the Bulgarian cross-modal priming data would likewise crucially depend on the frequency of occurrence of various word orders in spoken and written text. If it did turn out to be the case that Bulgarian truly allows multiple word order options of equivalent frequency, this would predict that filler-gap dependencies in Bulgarian (or other Slavic languages with similar word order flexibility) would not elicit the same pronounced ERP responses to marked, non-canonical word orders as has been demonstrated in English, German, and now Japanese as well.

Table 1

Point of Comparisons

pre-gap region ↓ post-gap region ↓

a. The local newspaper-to according that reckless adventurer-N finally **that/what-A** discovered-Q.

b. The local newspaper-to according **that/what-A** that reckless adventurer-N finally ____ discovered-Q.

FILLER GAP

↑ between filler and gap

‘According the local newspaper, did that reckless adventurer finally discover that/what?’

Table 2

Results of Overall ANOVAs for the Pre-Gap Region

	'Adventurer-NOM'						'Finally'					
	LAN: 300-600 ms			P600: 500-800 ms			LAN: 300-600 ms (950-1250 ms)			P600: 500-800 ms (1150-1450 ms)		
	Mid.	Para.	Tem.	Mid.	Para.	Tem.	Mid.	Para.	Tem.	Mid.	Para.	Tem.
Movement												
Type			*									*
Movement x Type												
Movement x Hemisphere												**
Movement x Anteriority	**	**	**		*	**			*		**	**
Movement x Type x Hemisphere		*										
Movement x Type x Anteriority									*			*
Movement x Hemisphere x Anteriority											*	
Movement x Type x Hemisphere x Anteriority			**			**			*			**

*p < .10. **p < .05.

Table 3

Summary of Results

<table>
<tr><th></th><th>Scrambled Demonstratives
vs.
In-Situ Demonstratives</th><th>Scrambled Wh
vs.
Wh-in-Situ</th></tr>
<tr><td rowspan="2">Between filler and gap
(subject NP + adverbial)</td><td colspan="2">bilateral anterior negativity
(slow potentials)</td></tr>
<tr><td>P600
(phasic)</td><td>--</td></tr>
<tr><td rowspan="2">Pre-gap (?) region
(head noun of subject NP + adverbial)</td><td>P600
(phasic)</td><td>--</td></tr>
<tr><td>LAN
(phasic)</td><td>LAN
(phasic)</td></tr>
<tr><td>Post-gap region
(sentence final verb)</td><td>anterior negativity
(phasic?)</td><td>anterior negativity
(phasic?)</td></tr>
</table>

Figure Captions

Figure 1. Phrase structure analysis of scrambling (Saito 1985).

Figure 2. ERPs (n=20) at frontal electrodes (Fp1, Fp2, Fz, F7, F8) to sentence positions between filler and gap ('that reckless adventurer-NOM finally' in Table 1) of scrambled vs. *in-situ* conditions, collapsed across demonstrative and *wh* conditions, and relative to a 100 ms poststimulus baseline.

Figure 3. ERPs (n=20) at midline electrodes (Fz, Cz, Pz) to sentence positions between filler and gap ('that reckless adventurer-NOM finally' in Table 1) of scrambled vs. *in-situ* demonstrative conditions, relative to a 400 ms prestimulus base line.

Figure 4. ERPs (n=20) at left hemisphere (Fp1, F7, F3, T3, C3, P3) and midline (Fz, Cz, Pz) electrodes to the pre-gap position ('adventurer-NOM finally' in Table 1) of scrambled vs. *in-situ* demonstratives.

Figure 5. ERPs (n=20) at left hemisphere (Fp1, F7, F3, T3, C3, P3) and midline (Fz, Cz, Pz) electrodes to the pre-gap position ('adventurer-NOM finally' in Table 1) of scrambled *wh* vs. *wh-in-situ*.

Figure 6. ERPs (n=20) at all electrodes to the post-gap verb position ('discovered-Q' in Table 1) of scrambled vs. *in-situ* demonstratives.

Figure 7. ERPs (n=20) at all electrodes to the post-gap verb position ('discovered-Q' in Table 1) of scrambled *wh* vs. *wh-in-situ*.

Figure 1

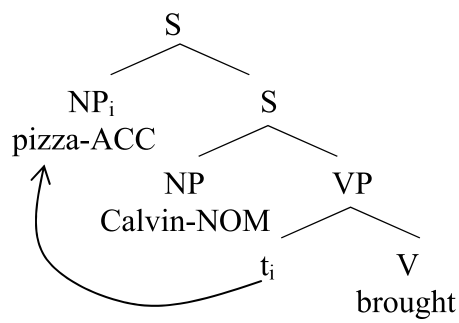


Figure 2

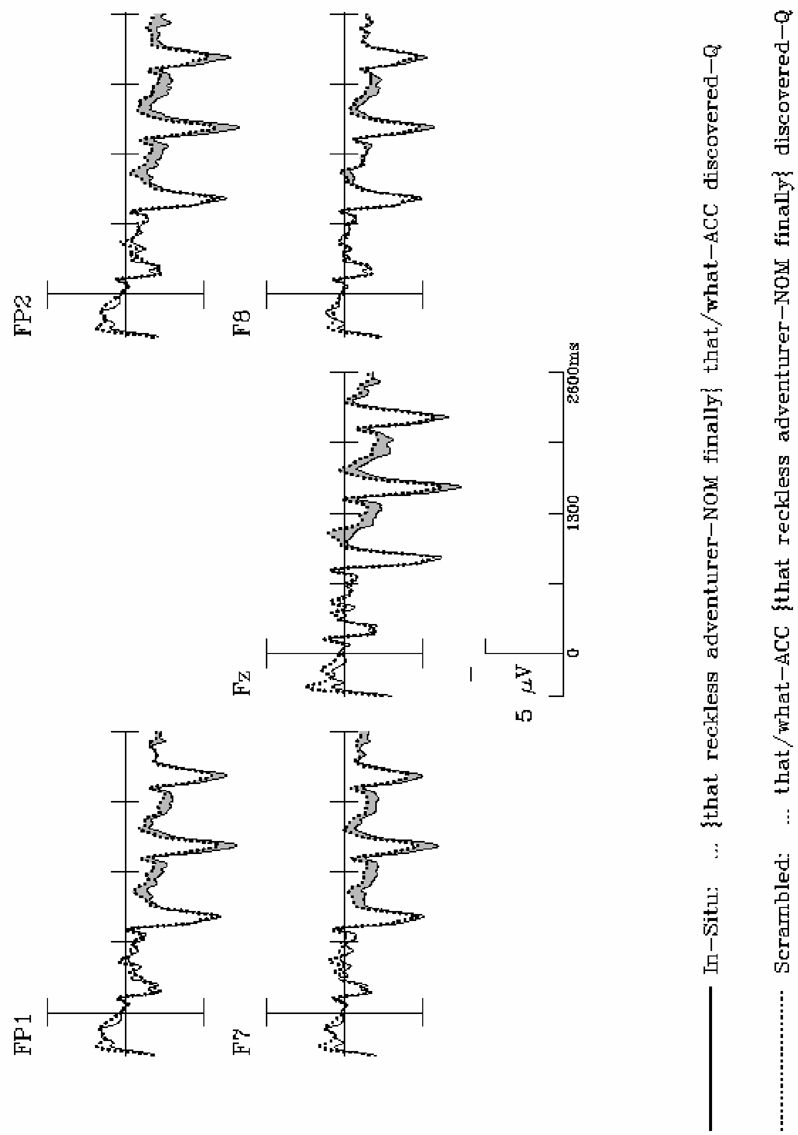

FP1
FP2
F7
Fz
F8
5 μV
0
1300
2600ms
In-Situ: ... {that reckless adventurer-NOM finally} that/what-ACC discovered-Q
Scrambled: ... that/what-ACC {that reckless adventurer-NOM finally} discovered-Q

Figure 3

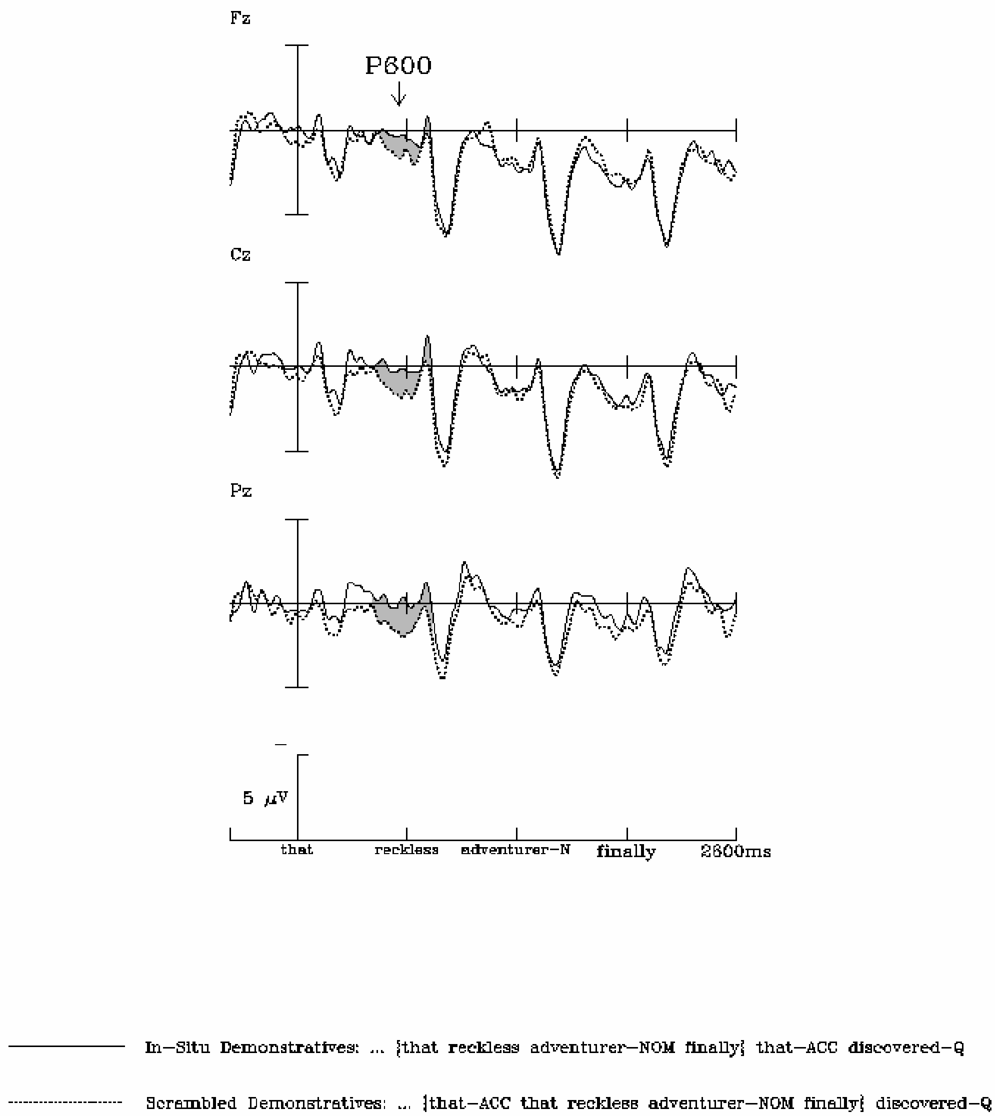

Fz
P600
Cz
Pz
5 μV
that
reckless
adventurer–N
finally
2600ms
In–Situ Demonstratives: ... {that reckless adventurer–NOM finally} that–ACC discovered–Q
Scrambled Demonstratives: ... {that–ACC that reckless adventurer–NOM finally} discovered–Q

Figure 4

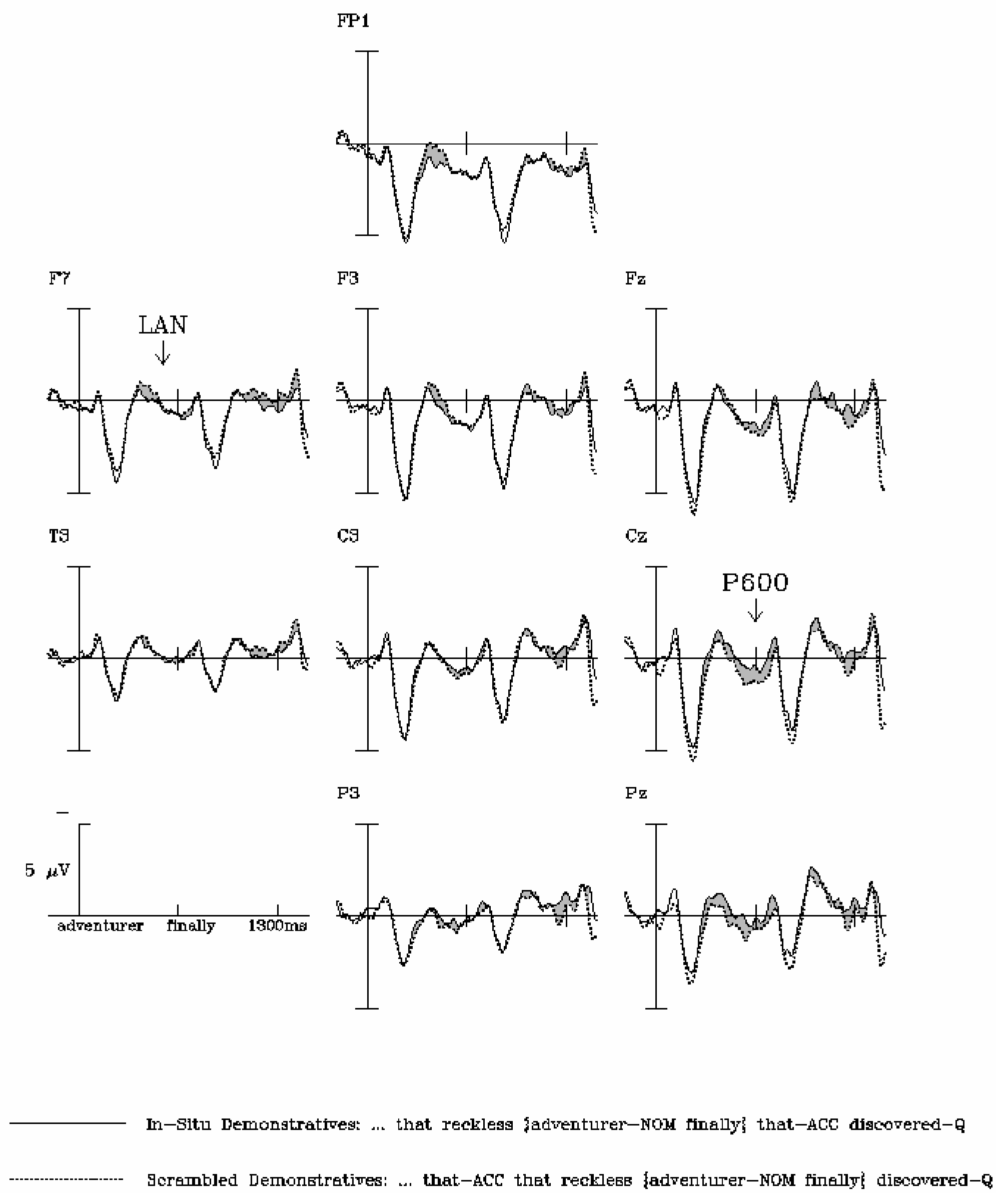

FP1
F7
LAN
F3
Fz
T3
C3
Cz
P600
5 μV
adventurer
finally
1300ms
P3
Pz
In–Situ Demonstratives: ... that reckless {adventurer–NOM finally} that–ACC discovered–Q
Scrambled Demonstratives: ... that–ACC that reckless {adventurer–NOM finally} discovered–Q

Figure 5

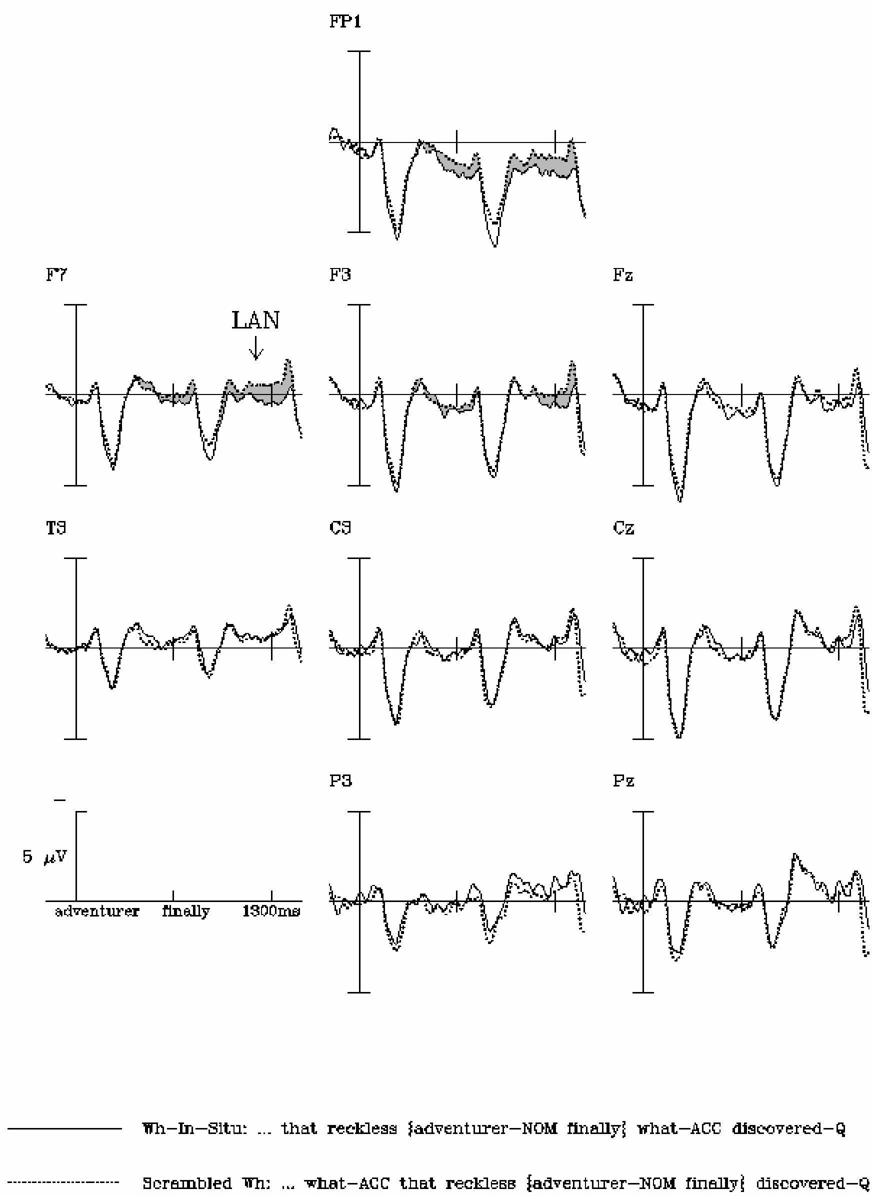
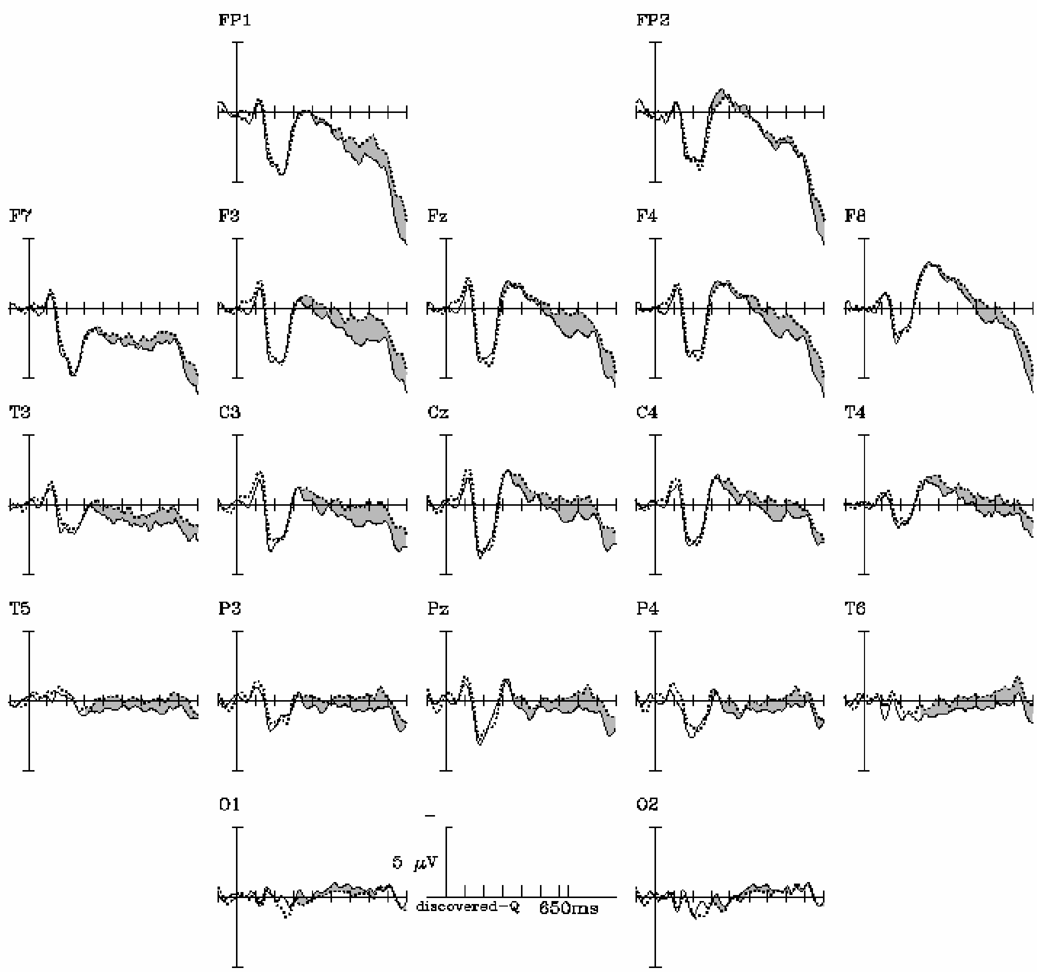

FP1
F7
LAN
F3
Fz
T3
C3
Cz
5 μV
adventurer
finally
1300ms
P3
Pz
Wh–In–Situ: ... that reckless {adventurer–NOM finally} what–ACC discovered–Q
Scrambled Wh: ... what–ACC that reckless {adventurer–NOM finally} discovered–Q

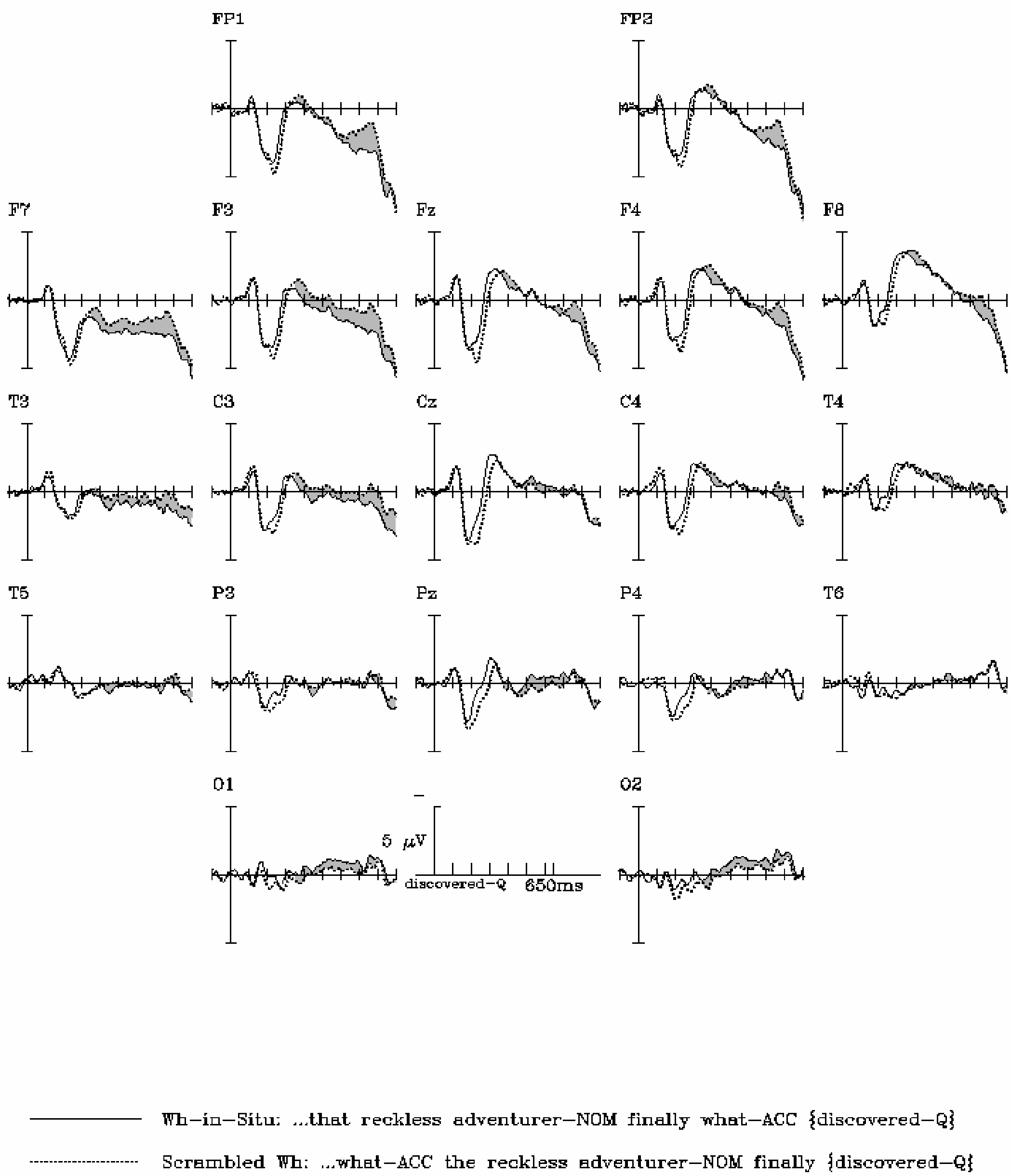
Figure 6



——— In–Situ Demonstratives: ...that reckless adventurer–NOM finally that–ACC {discovered–Q}

········· Scrambled Demonstratives: ...that–ACC that reckless adventurer–NOM finally {discovered–Q}

Figure 7


FP1
FP2
F7
F3
Fz
F4
F8
T3
C3
Cz
C4
T4
T5
P3
Pz
P4
T6
O1
O2
5 μV
discovered–Q
650ms
Wh–in–Situ: ...that reckless adventurer–NOM finally what–ACC {discovered–Q}
Scrambled Wh: ...what–ACC the reckless adventurer–NOM finally {discovered–Q}