Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Brain Research 22 (2005) 407–428

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/cogbrainres

Research report

ERP effects of the processing of syntactic long-distance dependencies

Colin Phillips[a,b,*], Nina Kazanina[a], Shani H. Abada[c]

[a]Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, United States
[b]Neuroscience and Cognitive Science Program, University of Maryland, United States
[c]School of Communication Sciences and Disorders, McGill University, Canada

Accepted 30 September 2004
Available online 19 November 2004

Abstract

In behavioral studies on sentence comprehension, much evidence indicates that shorter dependencies are preferred over longer dependencies, and that longer dependencies incur a greater processing cost. However, it remains uncertain which of the various steps involved in the processing of long-distance dependencies is responsible for the increased cost of longer dependencies. Previous sentence comprehension studies using event-related potentials (ERPs) have revealed response components that reflect the construction [J. King, M. Kutas, Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. Journal of Cognitive Neuroscience, 7, (1995) 376–395.] and completion [E. Kaan, A. Harris, E. Gibson, P. Holcomb, The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes, 5, (2000) 159–201.] of long-distance *wh*-dependencies. This article reports one off-line rating study and one ERP study that manipulated both the presence of *wh*-dependencies and the length of the dependencies (one clause vs. two clauses), with the aim of clarifying the locus of length-sensitivity and the functional role of associated ERP components. Results of the off-line study confirm that longer *wh*-dependencies incur greater processing cost. Results of the ERP study indicate that both a sustained anterior negativity that follows the initiation of the *wh*-dependency and also a late posterior positivity (P600) that marks the completion of the dependency are sensitive to the presence of a *wh*-dependency, but do not show amplitude variations reflecting the length of the dependency. However, the P600 is delayed when it marks the completion of a longer *wh*-dependency. This suggests that both the sustained negativity and the P600 reflect length-insensitive aspects of the construction of syntactic dependencies. In addition, an N400 component is elicited in the middle of the two clause *wh*-dependency, upon encountering a verb with an argument structure that prevents completion of the dependency.

Theme: Neural basis of behavior
Topic: Cognition

Keywords: Event-related potentials; Sentence processing; *Wh*-questions; Long-distance dependencies; P600; Sustained negativity; N400

1. Introduction

1.1. Filler-gap dependencies

A key property of natural language is its ability to build vast numbers of different expressions from a relatively small store of memorized words. This is made possible by the ability to combine words into hierarchically organized sentence structures. A hallmark of hierarchical sentence structure is the existence of long-distance dependencies between words and phrases that are not adjacent in the surface sequence of words. Long-distance dependencies can arise in at least two ways. They can arise in situations where elements are structurally close to one another and in their canonical positions, but are separated by additional material, as in subject–verb agreement relations such as '*The man who always delivers the newspaper drives a big black truck*'. Long-distance dependencies also arise in situations where a phrase is displaced from its canonical position in

* Corresponding author. Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, United States. Fax: +1 301 405 7104.
E-mail address: colin@umd.edu (C. Phillips).

0926-6410/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2004.09.012

order to mark scope, focus, topichood, or other grammatical functions. For example, English uses the position of *wh*-phrases to mark questions as direct questions or indirect questions. The canonical position of a direct object NP is immediately after the verb that assigns its thematic role (1a), but the direct object is moved to the front of an embedded clause in an indirect question (1b), or to the front of the main clause in a direct question (1c). Following standard psycholinguistic terminology, we refer to the fronted *wh*-phrase as a *filler*, and to the canonical position of the fronted NP, marked by underlining in (1b–c), as a *gap*. The combination of the two is known as a *filler-gap dependency* or *wh-dependency* [18].

(1a) The father knows that his son likes *pizza* for breakfast.
(1b) The father knows *what* his son likes ___ for breakfast.
(1c) *What* does the father know that his son likes ___ for breakfast?

In this study we use off-line complexity ratings and evidence from event-related potentials to investigate the time-course of the construction of filler-gap dependencies in language comprehension, with a focus on which sub-parts of this process are sensitive to the distance between the filler and the gap.

In sentence comprehension the parsing of a fronted *wh*-phrase initiates a search for a possible gap position. Successful identification of the gap position is necessary in order to determine which thematic role the *wh*-phrase receives and which verb assigns that thematic role. Semantic integration of the *wh*-phrase and verb is necessary, in order for compositional interpretation of the sentence to proceed. Although the parser could, in principle, wait for direct evidence of a gap position, in the form of a verb with an unfilled argument slot [42,88], much evidence now indicates that the parser pursues a more active approach. It posits a gap as soon as a potential gap site is identified, and does not wait for evidence that the gap site is not already filled. In head-initial languages such as English, potential gap sites can be identified using information from verb argument structure, and there is a good deal of evidence that direct object gaps are posited as soon as an appropriate transitive verb is encountered. Evidence for dependency-formation at verb positions comes from a variety of different sources, including filled gap effects in reading-time studies [9,80], implausibility detection studies using eye-movements [83] or ERP measures [28], complexity-based arguments [72], antecedent reactivation effects ([66,82] but cf. [60,67]), or patterns of anticipatory eye-movements [81]. An unresolved issue in the linguistic and psycholinguistic literature involves the question of whether *wh*-dependencies genuinely involve gap positions or whether fronted phrases are directly associated with verbs. However, since the focus of this paper is on *wh*-dependencies in English, where gap-based and gap-free accounts make very similar predictions about the time-course of dependency formation [11,33,72,77], our findings are compatible with either approach. We will refer to filler-gap dependencies, for ease of exposition, but nothing in this study hinges on the representational status of gaps.

1.2. Effects of dependency length

The grammar of English and other languages allows *wh*-dependencies to be arbitrarily long, potentially spanning many clauses and much intervening linguistic material. However, it is also clear from the psycholinguistic literature that shorter *wh*-dependencies are preferred over longer *wh*-dependencies, all other things being equal. In sentence comprehension, this is reflected in the finding that speakers attempt to complete *wh*-dependencies at the first potential position after a fronted *wh*-phrase [9,19,20,80]. In sentence production, it is reflected in the fact that attempts to elicit object relative clauses often lead to high levels of passive subject relative clauses, presumably because the shorter *wh*-dependency of a subject-gap relative clause is favored over the longer *wh*-dependency of an object-gap relative clause [10]. In judgments of perceived complexity, the length effect is reflected in higher ratings of difficulty for center-embedded sentences that contain longer *wh*-dependencies [31].

There are a number of different accounts of why shorter *wh*-dependencies are favored. Under one view, the length effect reflects a specific subroutine of the parser, known as the *Active Filler Strategy*, which attempts to create a gap site as soon as possible after the *wh*-phrase [20]. Under this approach, the creation of a gap is an end in itself. Under a second view, the preference for shorter dependencies is a consequence of the parser's attempt to link the *wh*-phrase as soon as possible to the verb that assigns it a thematic role [1,34,74]. A third approach derives the preference for early completion of *wh*-dependencies from the claim that longer dependencies consume a greater proportion of the working memory resources available for sentence processing [31,32]. This approach is the focus of the current study.

Gibson's approach distinguishes the costs associated with multiple stages in the construction of a long-distance dependency. Processing a *wh*-filler leads to the creation of an incomplete syntactic dependency, and the prediction of a thematic-role assigner, typically a verb that will allow completion of the dependency. This syntactic prediction must be maintained in working memory, and it is assumed that the cost of maintaining this prediction (*memory cost* or *storage cost*) increases as additional linguistic material is processed, leading to a distance-sensitive cost function. When the thematic-role assigning verb is processed, the filler must be integrated with the verb to complete the dependency. It is assumed that the process of integration requires a fixed amount of resources for performing the syntactic and semantic integration itself, plus an amount of resources for first reactivating the filler in working memory that is proportional to the distance between the filler and the verb. Thus, the completion of the filler-gap dependency is also assumed to involve a distance-sensitive cost function, but this is assumed to result from lexical reactivation processes, rather than from the compositional interpretation of the *wh*-phrase and the verb ([31], p.11; see also Ref. [41]). The preference for shorter *wh*-dependencies under this approach is therefore a special case of a general attempt to minimize the usage of working memory resources at each step of processing.

A key piece of evidence for the assumption that processing costs are sensitive to the length of syntactic dependencies comes from a contrast in ratings of intuitive complexity for sentences such as (2a–b).

(2a) The fact that the employee who the manager hired stole the office supplies worried the executive.
(2b) The executive who the fact that the employee stole office supplies worried hired the manger.

Sentence (2a), in which a relative clause is embedded inside a sentential complement, is judged to be significantly easier to process than (2b), in which the sentential complement is embedded inside the relative clause [31]. This contrast, first pointed out by Cowper [8], can be explained by the fact that the *wh*-dependency in the relative clause, marked by highlighting of the critical words, is longer in (2b) than in (2a). On the other hand, the contrast is unexpected under approaches that simply limit the number of center-embedded clauses [45,55] or the maximum number of incomplete syntactic dependencies [30,40,55,56,61,89]. Of course, examples (2a) and (2b) differ in more than just the length of the *wh*-dependency, and thus the complexity difference may be amenable to alternative explanations. In Experiment 1 below we provide a more direct test of the notion that increasing the length of a *wh*-dependency from one to two clauses increases perceptual complexity.

Note also that the memory-based approach to the dependency-length effect is not at odds with the previous two approaches. The Active Filler Strategy and the thematically driven approach offer theories of when a given dependency counts as complete or incomplete, and the memory-based model provides a specific measure of the costs that those dependencies incur. Thus, it is possible, in principle, to implement Gibson's claims about memory cost functions for a number of different parsing algorithms.

Contrasts in complexity ratings for sentences like (2a–b) suggest that distance matters to the cost of processing long-distance dependencies, but they do not provide information about the time-course of such effects, or about which of the stages of dependency construction is length sensitive. Any theory in which at least one component of dependency construction is length sensitive could account for the contrast, and in fact different claims about length-sensitivity can be found in different versions of Gibson's model (cf. [31], p. 65; [32]). Even in cases where word-by-word reading-time evidence indicates that processing difficulty is greatest around the position of the verb that completes the long-distance dependency [46], this could be attributed to a number of alternative processes, all of which are expected to reach maximum cost around the same word, and any of which could be distance sensitive. These processes include the cost of storing an incomplete dependency in working memory, the cost of reactivating the filler, and the cost of integrating the verb and the filler. Thus, a primary motivation for the current study was to obtain more fine-grained information on the locus of distance-sensitive effects in processing long-distance dependencies. To this end, we used event-related brain potentials (ERPs), which provide detailed information about sentence comprehension processes, as reflected in both the time-course and the scalp distribution of brain activity.

1.3. ERP reflections of well-formed dependencies

Although research on event-related potentials and sentence processing has for a long time been dominated by studies of different kinds of syntactic and semantic violations [21,39,51], a number of recent findings have demonstrated that ERPs can be used to investigate the mechanisms that underlie the processing of normal, well-formed sentences.

An important early result was the finding that the processing of grammatically correct but semantically anomalous sentences elicits a late centro-parietal negativity, known as the N400 [52,53]. Although the N400 was initially associated with the processing of semantically anomalous sentences (e.g., *He spread his warm bread with socks*), later work showed that the N400 is associated with fully plausible but unexpected sentence continuations (e.g., *The girl put the sweet in her pocket after the lesson*, [37,84,85]). It is now understood that all content words elicit an N400 response, and that the amplitude of this response varies inversely with the cloze probability of the word [3,53,54].

A different ERP component, which is frequently observed in studies of syntactic processing, is the late posterior response component known as the P600 [24,38,65,69]. The understanding of the P600 has evolved in a similar fashion to that of the N400. The P600 was initially observed in the context of parsing breakdown associated with 'garden path' sentences and a variety of different syntactic violations, and was considered to reflect the detection of syntactic anomalies. More recent studies have led to a view of the P600 as also reflecting successful syntactic processing, either in the form of diagnosis and reanalysis processes [25,26,63,64,70], or the detection of ambiguity [27], or in the form of successful syntactic integration operations [13,15,16,44]. Most relevant to the current study, Kaan et al. [44] showed that a

P600 response is observed at the verb that allows the completion of a *wh*-dependency, relative to a control condition in which no *wh*-fronting occurs, using sentences such as (3a–b). A similar finding has been reported for English relative clauses [35].

(3a) Emily wondered whether the performer in the concert had imitated a pop star for the audience's amusement.
(3b) Emily wondered which pop star the performer in the concert had imitated for the audience's amusement.

Kaan and colleagues suggest that the P600 response elicited at the completion of a *wh*-dependency reflects the processing cost associated with syntactic integration of the *wh*-phrase and the verb, and therefore conclude that the amplitude of the P600 is an index of syntactic integration difficulty. Importantly, they group together a number of sub-processes in this notion of 'syntactic integration', including reactivation of the filler *wh*-phrase, syntactic confirmation of thematic role assignment, and compositional semantic interpretation of the verb and the *wh*-phrase. Their account suggests that P600 amplitude should reflect not only the number of syntactic integration operations performed at a given word, but also the length of the syntactic dependencies that are completed at that point. The inclusion of a dependency-length manipulation in the current study makes it possible to test this prediction.

A different ERP response component has been observed in connection with the storage of a fronted *wh*-phrase in working memory during the processing of long-distance dependencies. King and Kutas [47] recorded single-word and multi-word ERPs while participants read subject and object relative clauses. In the object relative clauses, which involve a longer dependency, multi-word ERPs revealed a *sustained anterior negativity* (SAN), relative to the subject relative clause control condition. This effect began shortly after the head of the relative clause and continued to the verb of the relative clause four words later. A similar effect has been observed in an auditory version of the same task [62]. King and Kutas argue that the sustained negativity reflects the working memory load of holding an incomplete syntactic dependency in memory, and propose that it is related to the anterior negativity observed in a previous ERP study of working memory [76]. Based on this analysis, we can derive the following prediction, which is tested in the current study. If the sustained anterior negativity (SAN) is an index of working memory load, and if the cost of holding incomplete syntactic dependencies in memory is length-sensitive, then manipulation of the length of a *wh*-dependency should lead to an increase in the amplitude of the SAN towards the end of the dependency. Note that the effect of dependency-length on both the sustained negativity and the P600 have been investigated previously in a study of German by Fiebach et al. [16], a study that we discuss in more detail below.

In addition, some studies have reported a more transient left anterior negativity associated with the onset of a *wh*-dependency. Kluender and Kutas compared the response to a subject pronoun at the beginning of a *wh*-question, as in *What have you*... with the same pronoun at the beginning of a yes–no question, as in *Have you*... [48,49]. They observed a left-anterior negativity and a bilateral posterior positivity from around 300–500 ms after the onset of the pronoun in the *wh*-question condition, and interpret this effect as a reflection of storing the *wh*-filler in working memory. However, McKinnon and Osterhout [59] did not observe such a contrast at the non-pronominal subject (e.g. *the candidate*) that followed a *which*-NP or *whether* in a study of constraints on *wh*-dependencies.

1.4. Dependency length manipulation

The goals of the current study were to test the functional role of the sustained negativity that has been associated with holding incomplete syntactic dependencies in working memory, and the P600 that has been associated with syntactic integration at the completion of a long-distance dependency. In doing so, the aim was to clarify theoretical questions about the locus of distance-sensitive effects in sentence processing.

To this end, we compared single-word and multi-word ERPs elicited by short-distance and long-distance *wh*-dependencies with control conditions that lacked long-distance dependencies, as illustrated in Table 1. The critical words that mark the beginning and end of filler-gap dependencies, and their counterparts in the control conditions, are highlighted in the example sentences. The design of the stimulus materials is discussed in more detail in Section 3.

In both of the *wh*-dependency conditions ((b) and (d)), the processing of the *wh*-phrase should initiate a search for a verb that assigns a thematic role to the *wh*-phrase, which in turn allows compositional semantic interpretation to proceed. In the short-distance conditions, an appropriate verb is found in the first clause searched, whereas in the long-distance conditions, the first verb after the *wh*-phrase (e.g.

Table 1
Design of stimulus materials. Critical verbs and *wh*-phrases are highlighted

Short	Control (a)	The detective hoped that the lieutenant knew that the shrewd witness would **recognize** the accomplice in the lineup.
	wh (b)	The detective hoped that the lieutenant knew **which accomplice** the shrewd witness would **recognize** in the lineup.
Long	Control (c)	The lieutenant knew that the detective hoped that the shrewd witness would **recognize** the accomplice in the lineup.
	wh (d)	The lieutenant knew **which accomplice** the detective hoped that the shrewd witness would **recognize** in the lineup.

hope) is thematically inappropriate, and an appropriate verb is found in the second clause searched.

We can therefore formulate the following predictions. First, if the processing cost of integrating a verb and a fronted *wh*-phrase is length-sensitive, and if Kaan et al. [44] are correct in their claim that P600 amplitude is an index of syntactic integration cost, then we expect to observe a higher amplitude P600 at the most deeply embedded verb in the comparison of the long-distance *wh*-dependency condition with its control, relative to the same comparison in the short-distance conditions. Such a finding would parallel observations of P600 amplitude variation in the processing of garden-path and ungrammatical sentences [70]. If, on the other hand, only certain sub-processes of syntactic integration are length sensitive, as suggested by Gibson ([31], p. 11), and if the P600 does not reflect all sub-processes of syntactic integration, then P600 amplitude might not vary as a function of the length of the *wh*-dependency.

Second, if the memory cost of maintaining an incomplete *wh*-dependency is length-sensitive and if the SAN is an index of memory cost, then we should observe a higher amplitude in the comparison of conditions (c) and (d) than in the comparison of conditions (a) and (b). In the clause that immediately follows the *wh*-phrase, the SAN should show similar amplitude in short-distance and long-distance conditions. The effect of length should become evident only in the second clause of the long-distance condition.

Similar questions have been investigated by Fiebach et al. [16] in a study of German scrambling constructions, although that study differs from the current study in a number of respects. Fiebach et al. presented a number of different types of indirect *wh*-questions, comparing canonical subject-initial word order and non-canonical object-initial word order, and also varying the distance between the *wh*-word and the following argument NP. A sample object-initial sentence is shown in (4), with parentheses used to indicate the adverbial phrases that were present only in the long-distance dependency condition. They assumed a syntactic analysis in which the object-initial word order involves a syntactic dependency between the surface position of the *wh*-phrase and a gap in the canonical object position, but no comparable dependency in the subject-initial conditions. They were interested in the ERP correlates of the presence of the additional dependency in the *wh*-object conditions, and any effects of the length of this dependency.

(4) Thomas fragt sich, wen$_i$ am Dienstag [nachmittag nach dem Unfall]
Thomas asks himself, who.acc on Tuesday afternoon after the accident der Doktor ___$_i$ verständigt hat.
the doctor.nom called has
'Thomas asks himself who the doctor called on Tuesday afternoon after the accident.'

Fiebach et al. report a sustained left anterior negativity in the interval between the object *wh*-phrase and its canonical position in (4), relative to the same interval in the subject-initial condition. They argue that the amplitude of this negativity increases over the course of the dependency, and that this reflects the increasing cost of holding the *wh*-phrase in working memory. They also report a late centro-parietal positivity at the second argument NP of the embedded clause (e.g., der Doktor) in the object-initial conditions relative to the subject-initial conditions, and no interaction with length. They argue that this is a P600 response that reflects the completion of the *wh*-dependency in advance of the verb, and interpret the lack of a length effect as evidence of a length-insensitive integration process. No P600 effect was observed at the sentence-final verb position.

Although the current study shares a number of the objectives of the Fiebach et al. [16] study, the two studies differ in a number of respects. Most importantly, the use of English makes it possible to clearly evaluate ERP effects at the point of completion of the *wh*-dependency, since there is broad agreement that *wh*-dependencies like those illustrated in Table 1 are completed at the verb position [2,28,33, 80,83]. In verb-final languages such as German there is some evidence that a *wh*-dependency may be constructed before the verb is encountered [1,21], but it is impossible for semantic integration of the *wh*-phrase and the verb to proceed until the verb is processed. Fiebach et al. observed no P600 effect at the verb position, and thus could not evaluate the effects of dependency length on ERPs associated with syntactic integration. This issue is addressed further in the Discussion section. They did observe a P600 response at the pre-verbal subject NP in their object-initial conditions and suggested that this response reflected completion of the *wh*-dependency. However, this effect could also reflect the non-canonical position of the definite subject NP, which does not normally appear following a series of adverbial phrases in German. Other design improvements in the current study include the close matching across conditions of the linear position of the critical word, the use of a Latin Square design such that repetition of stimulus items is avoided, the use of multiple statistical measures of the SAN, and the use of longer-duration syntactic dependencies (2000 and 4000 ms in this study, 700 and 2600 ms in Fiebach et al. [16]).

A small number of other studies have investigated dependency-length effects using ERPs by manipulating the length of subject-verb agreement dependencies. Münte et al. [63,64] compared ERPs elicited by subject-verb agreement violations in German when the subject and the verb were adjacent in a main clause, and when the subject and the verb were separated by three words inside an embedded relative clause. They found that the P600 elicited by agreement violations had a larger amplitude in the relative clause conditions. In contrast, Kaan found no effect of dependency length on ERPs to both grammatical and ungrammatical subject–verb agreement dependencies in Dutch, when comparing dependencies with two or five intervening words [43]. Kaan suggests that the larger P600 response observed in the 'complex' conditions of Münte et al. [63,64] may reflect increased syntactic complexity rather than dependency length per se. Unlike the current study, both of these previous studies investigated the effect of dependency length on responses to violations, or, in the case of Kaan's grammatical conditions, on responses to dependencies that are not known to elicit well-defined ERP components.

In summary, evidence from behavioral studies suggests that the length of syntactic dependencies impacts processing cost, although many questions remain about the source and temporal profile of dependency length effects. Meanwhile, existing ERP studies provide evidence for at least two different response components that are sensitive to different stages in the construction of long-distance dependencies (SAN, P600), although it remains unclear exactly which sub-processes these ERP components reflect. The goal of the current study is to investigate how these ERP components are affected by a dependency-length manipulation, and to use this to help to identify the locus of dependency-length effects in models of parsing.

2. Experiment 1: complexity rating study

2.1. Rationale

The aim of this study was to use an off-line measure of complexity to test the prediction that two-clause *wh*-dependencies incur a greater processing cost than one-clause *wh*-dependencies. Although such a contrast is straightforwardly predicted by theories such as Gibson [31], it has not to our knowledge been tested previously. There is a good deal of evidence from behavioral and ERP studies that speakers prefer to associate a fronted *wh*-phrase with the first available verb ([1,28,80], Frazier and Clifton, 1989), but such findings do not entail that longer dependencies are more difficult to process. It is important to test this prediction, since it is a central premise of our ERP study.

2.2. Subjects

Subjects in this study were 24 undergraduate students at the University of Maryland (17 females). All subjects gave informed consent and were paid US$5.00 for their participation, which lasted around 30 min.

2.3. Materials and procedure

Materials for Experiment 1 (complexity rating task) and Experiment 2 (ERP study) were drawn from the same pool. We therefore describe the design used in both studies here. The experimental design consisted of 4 conditions, organized in a 2×2 factorial design, manipulating the presence or absence of a *wh*-dependency and the length of the dependency, comparing one clause vs. two clause dependencies. Materials consisted of 160 sets of the four target conditions, of which 24 sets were selected for use in Experiment 1.

A sample set of target items is shown in Table 1. The full set of materials for Experiments 1 and 2 is available from http://www.ling.umd.edu/colin/research. All sentences were declarative statements, including the sentences with *wh*-dependencies, which had the form of indirect questions with a *wh*-phrase at the beginning of an embedded clause. It is standard to use indirect questions in psycholinguistic studies of *wh*-dependencies, because they are more natural in a reading task, because they make it possible to use a yes/no question comprehension task, and because they can be embedded more inconspicuously among a set of filler items. Each sentence contained three clauses. Within each set of items, all lexical material was held constant, with differences only in word order and in the replacement of the determiner *the* in the control conditions with *which* in the *wh*-dependency conditions. All target items contained 3 verbs. One of the verbs, which allowed an indirect question as complement (e.g., *know*), appeared as the second verb in the short-distance conditions and as the first verb in the long-distance conditions. Another of the verbs, which allowed a sentential complement but not a noun phrase complement (e.g., *hope*), appeared as the first verb in the short-distance conditions and as the second verb in the long-distance conditions. It was important that this verb disallow a noun phrase complement, in order to prevent premature completion of the *wh*-dependency when it appeared as the intermediate verb in the long *wh*-dependency condition. The third verb was the same across all conditions, and was a strongly transitive verb (e.g., *recognize*). The mean transitivity value for the embedded verbs used in the 160 experimental items was 81%, based on the proportion of occurrences of each verb that was immediately followed by an NP in counts of the British National Corpus [78].

Twenty-four sets of four conditions were distributed among four lists in a Latin Square design. Each list was combined with 24 additional items of varying length and complexity to create four questionnaires of 48 items each. Participants were asked to rate the sentences on a scale from 1 to 5, where a score of 1 indicated that the sentence was very easy to understand and 5 indicated that the sentence was very difficult to understand.

2.4. Results and discussion

Mean ratings for the four experimental conditions were entered into a 2×2 repeated measures ANOVA, with the factors WH (*wh*-dependency vs. control) and Length (short-distance vs. long-distance). Mean ratings are shown in Table 2.

Results of the ANOVA confirmed the prediction that two-clause *wh*-dependencies are perceived as more difficult to understand than one-clause *wh*-dependencies. The main effects of WH and Length and the WH×Length interaction

Table 2
Mean complexity ratings in Experiment 1 on a scale from 1 (easy) to 5 (difficult)

	Control	WH
Short-distance	2.60 (0.41)	2.71 (0.65)
Long-distance	2.74 (0.69)	3.51 (0.51)

Standard deviations are shown in parentheses.

were all significant by subjects and items (WH: $F1(1,23)=22.9$, $p<0.0001$, $F2(1,23)=15.22$, $p<0.001$; Length: $F1(1,23)=30.27$, $p<0.0001$, $F2(1,23)=12.85$, $p<0.01$; WH×Length: $F1(1,23)=13.33$, $p<0.01$, $F2(1,23)=7.52$, $p<0.05$). Planned comparisons within each level of the WH factor showed that the difference between the two control conditions was not reliable ($t(23)=1.21$, $p=0.24$), but that the difference between the two *wh*-dependency conditions was highly reliable ($t(23)=5.83$, $p<0.001$).

In sum, results of Experiment 1 confirm that two-clause *wh*-dependencies are perceived as more difficult to process than are one-clause *wh*-dependencies. This contrast is expected under accounts that claim that longer syntactic dependencies entail a greater processing cost [31]. However, whole-sentence complexity ratings of this kind do not indicate the locus of the dependency length effect. The length effect could be due to any of the different factors discussed in the Introduction, including storage of the filler in working memory, reactivation of the filler at the completion of the dependency, or integration of the filler with the verb. Experiment 2 takes advantage of the high temporal sensitivity of ERP measures to investigate the detailed time course of the dependency length effect.

3. Experiment 2: materials and methods

3.1. Participants

There were 20 participants in the study (10 females), all of whom were undergraduate students at the University of Maryland. All participants were classified as strongly right-handed based upon the Edinburgh handedness inventory [68], had normal or corrected-to normal vision, and no known history of neurological disorder. All participants gave informed consent for their participation, and were paid US$10 per hour. Data from four participants (1 female) were excluded from the analyses, due to high levels of artifacts in the recordings, as described in more detail below.

3.2. Materials

The experimental sentences for the ERP study followed the same design used in Experiment 1 and illustrated in Table 1. There were four conditions, organized in a 2×2 factorial design, manipulating the factor WH (control vs. *wh*-dependency) and Length (one clause vs. two clause *wh*-dependency). The design of the experimental materials is discussed in more detail in Section 2.3. Of particular relevance to the on-line ERP study, the *wh*-phrase in the *wh*-dependency conditions was always a complex *which-N* phrase. This was chosen instead of the simpler *what* for two reasons. First, Kaan et al. observed a larger P600 response at the completion of dependencies headed by *which-N* than dependencies headed by *what* [44]. Second, *which-N* is unambiguously nominal, whereas *what* allows both a nominal and a clausal interpretation (e.g., *what did you say?*), and would therefore allow the possibility of premature completion of the *wh*-dependency at the intermediate verb in the long-distance condition. Third, *which-N* phrases have been argued on linguistic grounds to more easily allow true multi-clause dependencies than phrases such as *who* [7,71]. Although the use of a *which-N* phrase had the consequence that there was a one-word difference in the linear position of words in the control and *wh*-dependency conditions, statistical analyses showed that this linear offset was unlikely to have been responsible for differential ERP responses. For example, if differences observed at the embedded verb position were merely artifacts of the fact that this verb was the 13th word in control conditions and the 14th word in *wh*-dependency conditions, similar differences should have been observed at earlier positions in the same clause. The results reported below exclude this possibility.

Materials for Experiment 2 consisted of 160 sets of the four target conditions, distributed among four lists in a Latin Square design, such that each list contained only one item from each set, and 40 examples of each experimental condition. The full set of materials is available from http://www.ling.umd.edu/colin/research. Each subject read the items from one of the four lists, interspersed with 320 filler items in pseudorandom order. Each target sentence was followed by at least one filler sentence. Fillers contained one to three clauses, and were similar to the target items in length and complexity. There was some structural and lexical overlap between fillers and targets, in the form of clauses with *wh*-fronting, relative clauses headed by *who*, *which* or *that*, modifier clauses headed by *when* or *where*, etc. The effect of this was to make it difficult for participants to detect the target structures or to develop special strategies.

3.3. Procedure

Participants were comfortably seated in a dimly lit testing room, with a computer monitor approximately 100 cm in front of them. A fixation point appeared in the center of the screen at the beginning of each trial. The subject pressed a button to initiate presentation of the sentence, which began 1200 ms later. Sentences were presented word by word in the center of the screen at a rate of 500 ms per word (300 ms word, 200 ms blank screen). Words appeared in black lowercase letters on a white background. The stimuli subtended a visual angle of less than 3° horizontally. The final word of each sentence was marked with a period. A yes/no comprehension question appeared 500 ms after the final word of each sentence (e.g., for example (b) in Table 1, *Was the witness expected to identify an accomplice in the lineup?*). Subjects responded by pressing a key on a button box, and visual feedback was provided in the case of incorrect responses.

Due to the large number of test sentences and comprehension questions (480 in total), the experiment was divided across two recording sessions per subject. Each testing session began with a short practice block, followed by 5 blocks of 48 experimental sentences, each lasting 13–14 min. Participants were given short breaks between the blocks. Total recording time per session was approximately 75 min. Each session lasted 2 1/2–3 h including set-up time.

3.4. EEG recordings

Continuous EEG was recorded from 30 Ag/AgCl electrodes, mounted in an electrode cap (Electrocap International) and referenced to linked mastoids: midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, O1/2. Additional electrodes were placed above and below the left eye to monitor eye movements. The EEG and EOG recordings were amplified by a SynAmpsTM Model 5083 EEG amplifier, using a DC to 100 Hz low-pass filter, and sampled at a frequency of 1 kHz. Impedances were kept below 5 kΩ.

3.5. Data analysis

Trials with ocular and other large artifacts were removed based on visual screening prior to any further analyses. This procedure affected 7.3% of trials. In order to guard against artifacts caused by the slow voltage drifts that are commonly observed in DC EEG recordings, a series of additional procedures were applied. First, we applied a detrending algorithm similar to the one employed by Fiebach et al. [16], in order to correct for the common linear component, which showed a positive drift in most data sets. Raw data files were segmented into 11s time intervals, subject to the constraint that target sentences always fell within a single interval. The intervals used for detrending were thus almost three times the length of the long-distance *wh*-dependencies. Within each interval, a linear regression was computed for each electrode, and subtracted from the original recording. Second, any detrended intervals that showed a range of more than 100μV were excluded. This procedure affected a further 12.8% of trials. Third, we examined averages for each subject in each condition in the intervals where the target sentences were identical, such that no systematic differences should be expected between the *wh*-dependency and control conditions, i.e., words 1–7 in the short-distance conditions, and words 1–3 in the long-distance conditions. Based on this procedure, four subjects were excluded who showed divergences of greater than 10 μV at 12 or more electrode sites in these intervals. Among the remaining 16 subjects, only 2% of electrode sites showed divergences on this scale in this interval. Subsequent analyses indicated that differences among conditions in the ERP responses were unlikely to be artifacts of the detrending procedure. Further details are provided in Section 4. After the artifact reduction procedures had been applied, 83.7% of the initial total number of trials from the 16 included subjects remained available for subsequent analysis. The inclusion rate was closely matched across the four experimental conditions (range: 83.1–84.3%).

For statistical analyses, six regions of interest (ROIs) were used in the ANOVAs, consisting of groups of three electrodes at each ROI: left anterior (F3, FC3, C3), anterior midline (FZ, FCZ, CZ), right anterior (F4, FC4, C4), left posterior (CP3, P3, O1), posterior midline (CPZ, PZ, OZ), right posterior (CP4, P4, O2). These ROIs were organized into the two topographic factors *laterality* (left, midline, right), and *anterior–posterior*.

Subsequent analyses were of two types. *Single-word analyses* were based on 1100 ms intervals surrounding each word of the target sentences, consisting of a 100-ms prestimulus baseline and a 1000-ms post-stimulus interval. At each word of interest, ANOVAs were computed with the within-subjects factors *wh-dependency* (*wh*-dependency vs. control) and *length* (short-distance vs. long-distance), and the two topographic factors described above. All *p*-values reported below reflect application of the Greenhouse-Geisser correction where appropriate, to control for violations of the sphericity assumption [36], together with the original degrees of freedom. At all word-intervals, ANOVAs were calculated based on mean voltages within a series of 200 ms time-windows (−100–100 ms, 100–300 ms, 300–500 ms, 500–700 ms, 700–900 ms). Due to the large number of possible interactions in this design, we report as significant only those interactions for which subsequent analyses yielded significant contrasts within the levels of the interacting factors. Grand average waveforms were low-pass filtered at 10 Hz for purposes of visualization, but all analyses were based on unfiltered data. In order to check for possible list-specific effects, an additional set of analyses was conducted at the two critical verb positions including *list* as a factor in the ANOVA. Among all analyses tested in this way, only one significant interaction (discussed below) was found to be paired with a significant interaction with the list factor, and this interaction does not impact the overall conclusion of the study in any way.

Multi-word analyses were based on a series of 500 ms intervals corresponding to each of the words that appeared between the start and finish of the *wh*-dependency. Each interval lasted from 100 ms after the presentation of a word until 600 ms after the presentation of the word. Two types of analysis were performed on each interval. In the cumulative analyses, which were most similar to the multi-word measures employed by Fiebach et al. [16], mean voltages were computed for the entire 500 ms interval, relative to a common baseline established to a word at the beginning of the dependency, i.e., the complementizer *that* in the control conditions and a noun (e.g. *accomplice*) in the *wh*-dependency conditions. The cumulative analyses provide a measure of the total divergence between *wh*-dependency and control conditions since the onset of the *wh*-dependency, but do not provide an accurate measure of the contribution of each successive word to the sustained negativity. The non-cumulative analyses were identical to the cumulative analyses, except that mean voltages were computed relative to a baseline established at the beginning of each individual interval. This type of multiword analysis is more similar to that used in King and Kutas [47]. It provides a measure of the local degree of divergence between *wh*-dependency and control conditions specifically at that time interval. If the amplitude of the sustained negativity reflects the storage cost of holding a *wh*-filler in working memory, and if this cost increases over the course of a dependency, then successive intervals during the *wh*-dependency should show increased anterior negativity in the non-cumulative multi-word analysis. For both types of multi-word analysis, the baseline interval lasted from 25 to 65 ms after the onset of the word, an interval that was centered around the peak of the P50 response to words in the grand average. This brief post-stimulus baseline interval was selected instead of a pre-stimulus interval for two reasons. First, lexical differences in the region that preceded the onset of the *wh*-dependency made it undesirable to use a pre-stimulus interval as a baseline. Second, a brief baseline interval was used in order to minimize baseline artifacts resulting from the SAN response.

4. Results

4.1. Comprehension question accuracy

Among the 16 participants included in the analysis, comprehension accuracy was 88.5% for target sentences.

4.2. Single-word ERP analyses

4.2.1. Early regions

All conditions were identical up to the main clause verb (word 3), and the short-distance conditions were identical up to the second verb (word 7). In the overall ANOVA for words 1–3 there were no significant main effects or interactions. There were also no significant main effects or interactions in the analysis of short-distance conditions alone at words 4–7.

4.2.2. Onset of wh-dependency

The onset of the *wh*-dependency consisted of a *which-N* phrase followed by a determiner in both short-distance and long-distance conditions alike. The corresponding position in the control conditions contained the complementizer *that*. Since the *wh*-dependency began at the start of the second clause in the long-distance conditions and at the start of the third clause in the short-distance conditions, results for these two regions are reported separately.

In the long-distance conditions, the word following the main clause verb was different between the control condition (*that*) and the *wh*-dependency condition (*which*). Planned comparisons involving only the long-distance conditions revealed a number of significant effects. There was a positivity in the *wh*-dependency condition, which was strongest over anterior channels, with a slight left hemisphere dominance. This produced an interaction of the *wh*-dependency factor and the anterior–posterior factor that was marginally significant at the 100–300 ms interval, $F(1,15)=4.17$, $p<0.06$, and significant at the 300–500 ms interval, $F(1,15)=9.91$, $p<0.01$, and at the 500–700 ms interval, $F(1,15)=5.36$, $p<0.05$. At the 300–500 ms interval there was also a significant interaction of the *wh*-dependency factor and laterality, $F(2, 30)=3.96$, $p<0.05$. At the 700–900 ms interval there was also a significant main effect of the *wh*-dependency factor, $F(1,15)=4.77$, $p<0.05$, due to a broadly distributed negativity in the *wh*-dependency condition. However, this effect may have been due to lexical differences at the following word position, where there was a common noun in the *wh*-dependency condition (e.g. *accomplice*), and the determiner *the* in the control condition.

For all subsequent regions, responses to words in the *wh*-dependency condition are compared to identical words in the control condition. At the determiner *the* at the beginning of the second clause there was a broadly distributed positivity in the long-distance *wh*-dependency condition, which lasted from shortly after the onset of the word until the 500–700 ms interval. However, this was probably an artifact, due to the fact that the baseline interval for this region coincided with an interval when the responses to the preceding words differed the most. Consistent with this interpretation, the effect disappears in an analysis that uses a 25–65 ms post-stimulus interval as the baseline interval (see Fig. 3). In the 700–900 ms interval there was a negativity with a left anterior focus, which gave rise to an interaction of the *wh*-dependency factor with the anterior posterior factor, $F(1,15)=10.04$, $p<0.01$, and an interaction of the *wh*-dependency factor with laterality, $F(2,30)=4.05$, $p<0.05$. This effect resembles the transient left-anterior negativity observed at the onset of *wh*-dependencies in some previous studies [48,49]. However, it is important to note that although the negativity was observed at left anterior regions only, other regions of interest showed a reduction or disappearance of the positivity that had been present at earlier time intervals. Thus, it is possible that a more broadly distributed negativity was masked by the earlier positivity. This interpretation is consistent with analyses at the 100–300 ms time interval for the following noun (*detective*), where the main effect of the *wh*-dependency factor was marginally significant, $F(1,15)=4.40$, $p=0.053$, due to a negativity in the *wh*-dependency condition. There was also a significant interaction of the *wh*-dependency factor with laterality and the anterior–posterior factor, $F(2,30)=3.85$, $p<0.05$, due to the fact that the effect of the *wh*-dependency was stronger over left and midline anterior regions. The scalp distribution of this negativity is shown in Fig. 1B.

In the short-distance conditions, the *wh*-dependency began at the start of the third clause. The results of analyses of the short-distance conditions in these regions are similar to results from the start of the *wh*-dependency in the long-distance conditions. At the first word of the third clause, which involved a lexical difference between the complementizer *that* and the *wh*-word *which*, no main effects or interactions were significant until the 700–900 ms interval, where there was a negativity with a posterior focus in the *wh*-dependency condition. This produced a main effect of the *wh*-dependency factor, $F(1,15)=5.46$, $p<0.05$, and a marginally significant interaction of the *wh*-dependency factor with the anterior–posterior factor, $F(1,15)=3.81$, $p=0.07$. As in the corresponding region in the long-distance conditions, this late effect may reflect the lexical difference between the common noun (e.g. *accomplice*) and the determiner *the* at the following

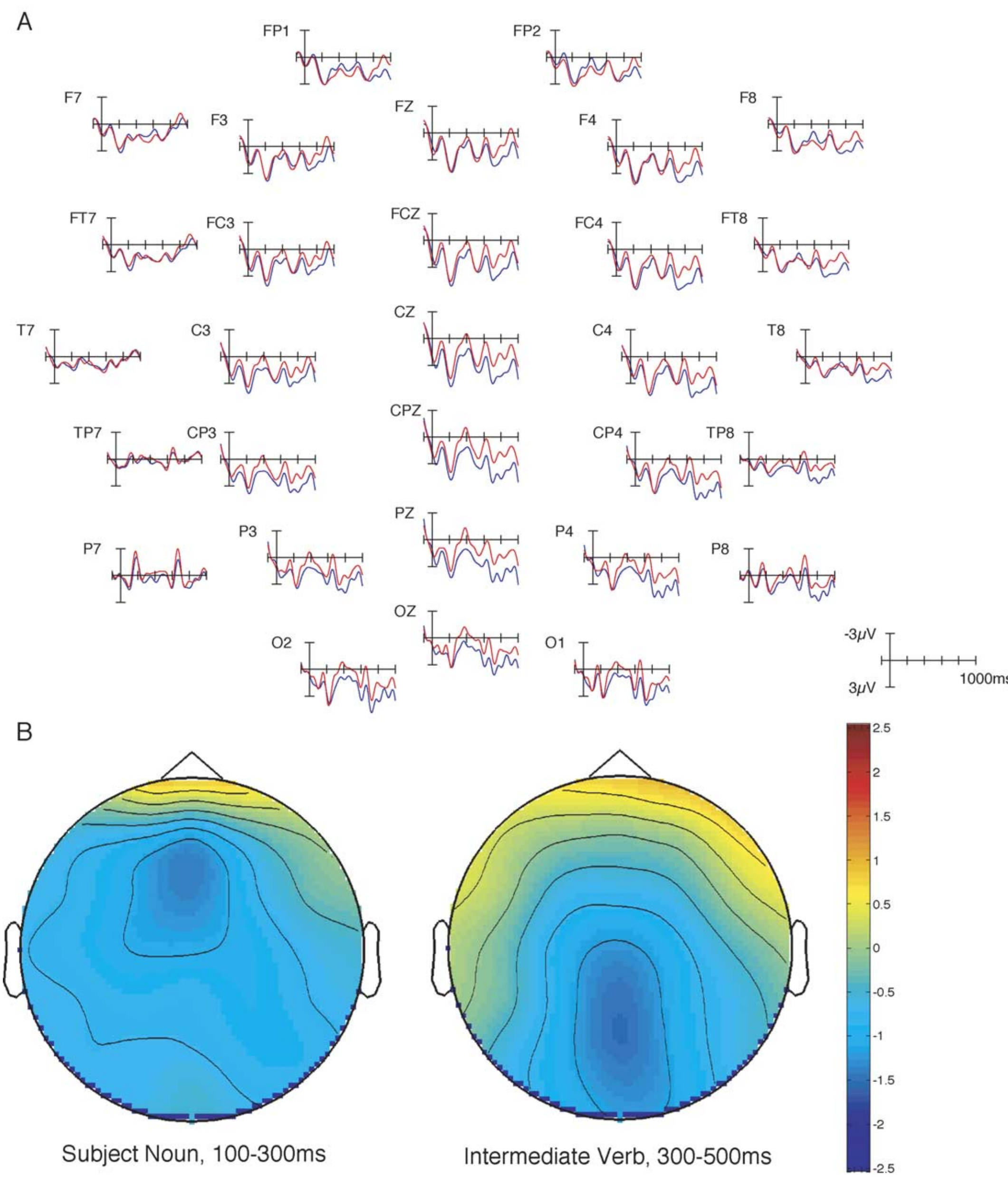

Fig. 1. (A) Grand average ERP response at the intermediate verb position to the long-distance *wh*-dependency condition (red line) and its control (blue line). (B) Topographic scalp voltage maps, comparing the centro-parietal N400 response elicited at the intermediate verb (right) with the anterior negativity observed at the noun in the preceding region (left).

region. At the determiner *the* at the beginning of the third clause, there was a positivity with a slight posterior focus, which in the 100–300 ms interval produced a main effect of the *wh*-dependency factor, $F(1,15)=5.66$, $p<0.05$, and in the 300–500 ms interval produced a marginally significant main effect of the *wh*-dependency, $F(1,15)=3.72$, $p=0.07$, and an interaction of the *wh*-dependency factor with laterality and the anterior–posterior factor, $F(2,30)=5.86$, $p<0.01$. Analyses at individual ROIs showed that this positivity was significant or marginally significant at all posterior ROIs (left posterior: $F(1,15)=4.45$, $p=0.052$; midline posterior: $F(1,15)=5.78$, $p<0.05$; right posterior: $F(1,15)=5.67$, $p<0.05$) and at no anterior ROIs. As in the corresponding position at the onset of the long-distance *wh*-dependency, this positivity may reflect a baseline artifact due to the lexical differences between conditions at the previous word.

4.2.3. Intermediate verb

At the position of the intermediate verb, visual inspection revealed a negativity in the long-distance *wh*-dependency condition relative to its control, with the broad distribution and the timing profile characteristic of an N400 response, as shown in Table 3 and Fig. 1. No corresponding effect was observed in the comparison of the short-distance *wh*-dependency condition and its control. The absence of any effect in the short-distance conditions is unsurprising, since those conditions were identical up to this region. We first report the results of the overall ANOVA, and then proceed to describe planned comparisons within each level of the dependency-length and anterior–posterior factors. In the discussion section we suggest that the N400 in the long-distance conditions reflects the fact that the verb prevents immediate completion of the *wh*-dependency.

Table 3
Summary of ANOVA *f*-values at intermediate verb in long-distance conditions

	100–300 ms	300–500 ms	500–700 ms	700–900 ms
ANOVA, Long conditions				
wh (1,15)	–	3.09†	2.84†	5.27*
wh×ant (1,15)	–	4.77*	3.65†	–
wh×lat (2,30)	–	–	–	3.84*
wh×ant×lat (2,30)	–	–	–	–
Effect of wh-dependency at individual ROIs				
Left anterior (1,15)	–	–	–	3.35†
Midline anterior (1,15)	–	–	–	5.37*
Right anterior (1,15)	–	–	–	4.89*
Left posterior (1,15)	–	4.93*	4.24†	–
Midline posterior (1,15)	–	6.55*	3.51†	5.62*
Right posterior (1,15)	–	4.72*	–	5.00*

Factors: *wh*—*wh*-dependency; *ant*—anterior–posterior; *lat*—laterality.
* $p<0.05$.
† $0.05<p<0.11$.

Results of the overall ANOVA at the intermediate verb showed a main effect of the *wh*-dependency at the 300–500 ms interval, $F(1,15)=5.24$, $p<0.05$, and at the 700–900 ms interval, $F(1,15)=5.39$, $p<0.05$, due to greater negativity in the *wh*-dependency conditions, but not at the 500–700 ms interval ($F<1$). In contrast, the dependency-length×*wh*-dependency interaction was marginally significant at the 500–700 ms interval, $F(1,15)=3.46$, $p=0.082$, but did not approach significance at the 300–500 ms interval or the 700–900 ms interval. The only other significant effects in the overall ANOVA at this word were the anterior×dependency length×*wh*-dependency interaction at the 300–500 ms interval, $F(1,15)=5.21$, $p<0.05$, and the laterality×dependency-length×*wh*-dependency interaction at the 500–700 ms interval, $F(2,30)=3.88$, $p<0.05$, and at the 700–900 ms interval, $F(2,30)=4.60$, $p<0.05$.

However, planned comparisons within each level of the dependency-length factor revealed no significant main effects or interactions at any interval in the short-distance conditions, but a number of significant contrasts involving the long-distance conditions, due to a negativity in the *wh*-dependency condition that showed a posterior focus in the 300–500 and 500–700 ms intervals and a broader distribution in the 700–900 ms interval. The primary effects in the long-distance conditions are summarized in Table 3. The main effect of the *wh*-dependency factor was marginally significant at the 300–500 ms interval, $F(1,15)=3.09$, $p=0.099$, and the 500–700 ms interval, $F(1,15)=2.84$, $p=0.11$, and significant at the 700–900 ms interval, $F(1,15)=5.27$, $p<0.05$. The interaction of the *wh*-dependency and anterior–posterior factors was significant at the 300–500 ms interval, $F(1,15)=4.77$, $p<0.05$, and marginally significant at the 500–700 ms interval, $F(1,15)=3.65$, $p=0.075$, but not at the 700–900 ms interval, $F<1$. The laterality×*wh*-dependency interaction was significant only at the 700–900 ms interval, $F(2,30)=3.84$, $p<0.05$.

Separate analyses of short-distance and long-distance conditions at the nine posterior electrodes and the nine anterior electrodes revealed topographic differences. In the short-distance conditions there were no significant effects at any location or interval. In the long-distance conditions the only significant effect at the anterior electrodes was a main effect of the *wh*-dependency at the 700–900 ms interval, $F(1,15)=4.96$, $p<0.05$. The late onset of this effect may reflect the fact that this effect was in fact an anterior negativity elicited by the complementizer *that* following the verb. At posterior electrodes, however, the effect of the *wh*-dependency factor was significant at the 300–500 ms interval, $F(1,15)=5.72$, $p<0.05$, and the 700–900 ms interval, $F(1,15)=4.57$, $p<0.05$, and was marginally significant at the 500–700 ms interval, $F(1,15)=3.48$, $p=0.08$. The laterality×*wh*-dependency interaction was significant at the 500–700 ms interval, $F(2,30)=3.74$, $p<0.05$, and at the 700–900 ms interval, $F(2,30)=8.16$, $p<0.01$, due to a more pronounced negativity at midline and right-hemisphere electrodes than at left-hemisphere electrodes.

None of the significant main effects, interactions or planned comparisons at any interval showed a significant interaction with the factor *list* when that factor was included in the ANOVA.

In consideration of the possibility that the negativity observed at the intermediate verb might be related to the anterior negativity associated with the storage of a *wh*-filler in working memory, Fig. 1B compares the scalp distribution of the negativity in the long-distance conditions at the intermediate verb and at the preceding noun. Although both words elicited a negativity in the *wh*-dependency condition relative to the control condition, the scalp distribution of this negativity differs. The negativity at the intermediate verb shows the centro-parietal distribution characteristic of an N400 response, whereas the negativity at the preceding word has a left anterior focus.

4.2.4. Pre-final words of wh-dependency

The final three words of both the short-distance and the long-distance *wh*-dependencies consisted of an adjective, a common noun, and an auxiliary verb. There were no significant main effects or interactions involving the *wh*-dependency factor at these regions. Planned comparisons within each level of the dependency-length factor also showed no significant effects of the *wh*-dependency. The lack of differences at these regions is important, since it indicates that any differences observed at the following verb position are likely to be genuine effects of the experimental manipulations, rather than artifacts of the one word offset between the control and *wh*-dependency conditions.

4.2.5. Final verb

The final verb of the sentence marked the completion of the *wh*-dependency. Analyses of single-word ERPs at this region revealed a late positive component in both *wh*-dependency conditions relative to their respective controls. This component showed the posterior scalp distribution characteristic of a P600 response. This component showed an earlier onset and a broader scalp distribution in the short-distance conditions than in the long-distance conditions, with the positivity reaching significance in the 300–500 ms interval in the short-distance conditions and in the 500–700 ms interval in the long-distance conditions. Waveforms and topographic maps are shown in Fig. 2, and results of the ANOVA are summarized in Table 4.

Results of the overall ANOVA at the final verb showed few significant main effects, due to the fact that the positivity elicited by the *wh*-dependency conditions showed the posterior distribution characteristic of the P600 response. The only significant main effect was an effect of dependency-length in the 300–500 ms interval, $F(1,15)=7.90$, $p<0.05$. Also, the *wh*-dependency×dependency-length interaction did not reach significance in any interval. However, a number of interactions involving the anterior–posterior factor were significant. The *wh*-dependency×anterior–posterior interaction was marginally significant at the 500–700 ms interval, $F(1,15)=4.14$, $p=0.06$, and was significant at the 700–900 ms interval, $F(1,15)=4.75$, $p<0.05$, reflecting the posterior focus of the positivity in the *wh*-dependency conditions. The *wh*-dependency×anterior–posterior×laterality interaction was significant at both the 500–700 ms interval, $F(2,30)=4.49$, $p<0.05$, and the 700–900 ms interval, $F(2,30)=7.63$, $p<0.01$. In further analyses this last three-way interaction also participated in a significant four-way interaction with the list factor, suggesting the possibility of item-specific or participant-specific effects. However, since this late interaction does not impact the overall conclusions of the study, it is not considered further. The dependency-length×anterior–posterior×laterality interaction was marginally significant at both the 300–500 ms interval, $F(2,30)=2.89$, $p=0.074$, and at the 700–900 ms interval, $F(2,30)=2.94$, $p=0.083$. These interactions, together with the experimental hypothesis of a positivity with a posterior focus, motivate additional analysis of the topographic distribution of the ERP effects.

Planned comparisons within each level of the anterior–posterior factor revealed clear topographic differences, with many more significant effects appearing at posterior than at anterior regions. At anterior electrodes the only significant effects were a *wh*-dependency×laterality interaction in the 100–300 ms interval, $F(2,30)=4.65$, $p<0.05$, and in the 500–700 ms interval, $F(2,30)=3.41$, $p<0.05$. However, additional comparisons within each level of the laterality factor revealed no significant effects, and thus these effects are not considered further.

At posterior electrodes, the 100–300 ms interval showed a significant *wh*-dependency×dependency-length interaction, $F(1,15)=5.08$, $p<0.05$, due to a marginally significant effect of the *wh*-dependency factor in the short-distance conditions, $F(1,15)=2.93$, $p<0.11$, but no such effect in the long-distance conditions, $F<1.1$. However, given the marginal nature of the effect of *wh*-dependency in the short-distance conditions, we consider it premature to conclude that the short-distance and long-distance conditions differ in this time interval.

At the 300–500 ms interval at posterior electrodes, there was a significant main effect of dependency-length, $F(1,15)=10.04$, $p<0.01$, and a main effect of the *wh*-dependency, $F(1,15)=4.96$, $p<0.05$. Although the *wh*-dependency×dependency-length interaction did not reach significance, $F(1,15)=2.16$, $p=0.16$, pairwise comparisons within each level of the dependency length factor showed that there was a marginally significant effect of the *wh*-dependency in the short-distance conditions, $F(1,15)=4.38$, $p=0.054$, but no effect of the *wh*-dependency in the long-distance conditions, $F(1,15)=1.86$, $p=0.19$. Fig. 3 compares the mean voltages at posterior electrodes at this interval. We conclude from this comparison that the posterior positivity reaches significance earlier in the short-distance conditions.

At the 500–700 ms interval at posterior electrodes, the main effect of the *wh*-dependency factor was significant,

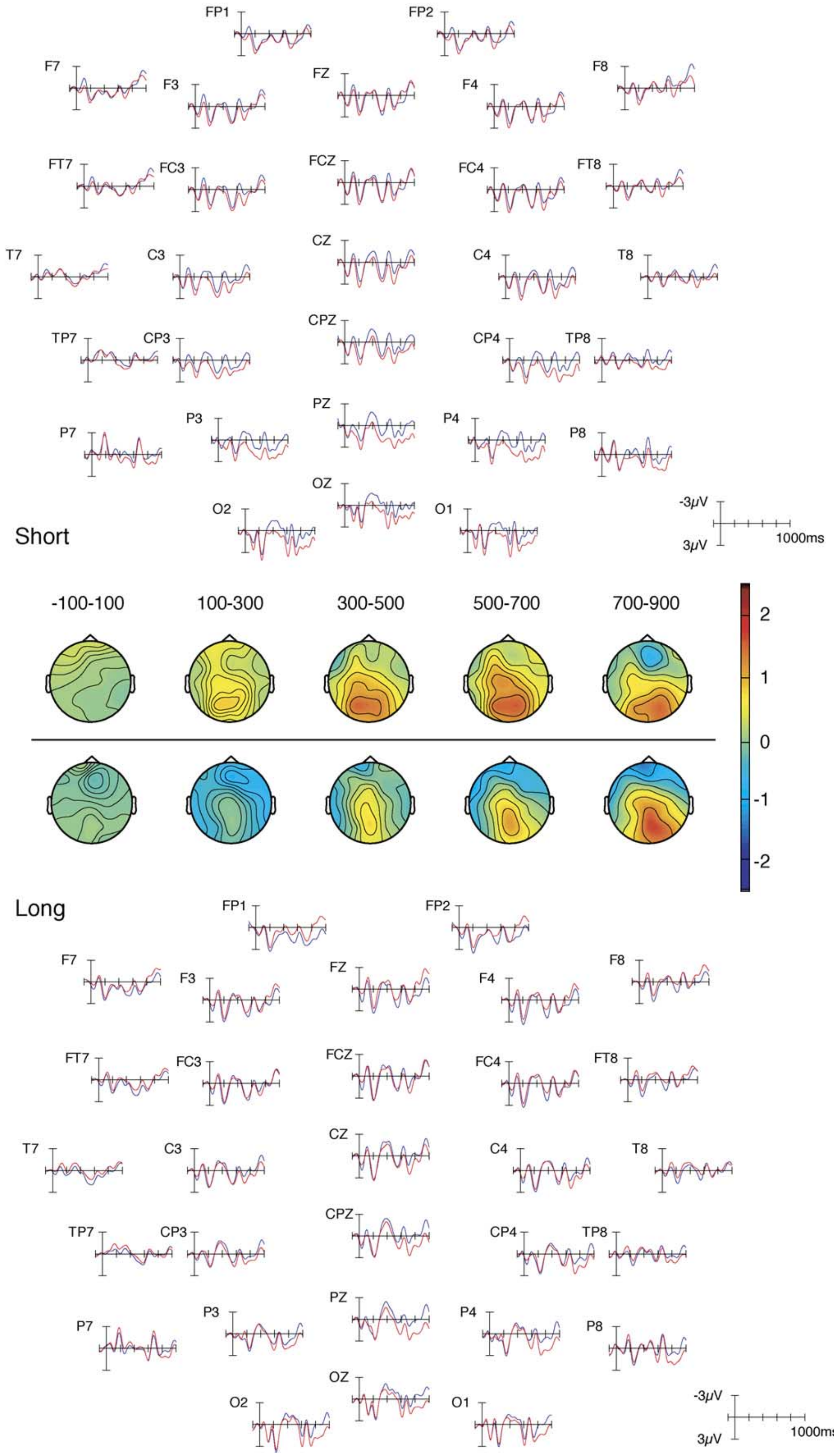

Fig. 2. Grand average ERP responses at the final verb, which marked the completion of the *wh*-dependency, for short-distance conditions (top) and long-distance conditions (bottom). *Wh*-dependency conditions are shown in red, control conditions in blue. Topographic maps (center) are based on average voltage differences between the *wh*-dependency and control conditions at successive 200 ms time intervals.

Table 4
Summary of ANOVA *f*-values from analyses at the final verb position

	100–300 ms	300–500 ms	500–700 ms	700–900 ms
Overall ANOVA				
wh (1,15)	–	–	–	–
lg (1,15)	–	7.90*	–	–
wh×*lg* (1,15)	–	–	–	–
wh×*ant* (1,15)	–	–	4.14†	4.75*
wh×*lat* (2,30)	–	–	–	–
wh×*ant*×*lat* (2,30)	–	–	4.49*	7.63**
wh×*lg*×*ant* (1,15)	–	–	–	–
wh×*lg*×*lat* (2,30)	–	–	–	–
lg×*ant*×*lat* (2,30)	–	2.89†	–	2.94†
wh×*lg*×*ant*×*lat* (2,30)	–	–	–	–
Anterior regions only				
wh (1,15)	–	–	–	–
wh×*lg* (1,15)	–	–	–	–
wh×*lat* (2,30)	4.65*	–	–	3.41*
wh×*lg*×*lat* (2,30)	–	–	–	–
Posterior regions only				
wh (1,15)	–	4.96*	5.57*	6.68*
lg (1,15)	–	10.04**	4.14†	–
wh×*lg* (1,15)	5.08*	–	–	–
wh×*lat* (2,30)	–	–	3.60*	3.71*
wh×*lg*×*lat* (2,30)	–	–	–	2.87†
Short–distance conditions				
left posterior (1,15)	–	4.40†	5.48*	–
midline posterior (1,15)	–	4.48†	5.91*	–
right posterior (1,15)	–	3.89†	5.16*	4.23†
Long-distance conditions				
left posterior (1,15)	–	–	–	
midline posterior (1,15)	–	3.31†	3.73†	5.10*
right posterior (1,15)	–	–	–	4.51†

Factors: *wh*—*wh*-dependency; *lg*—dependency-length; *ant*—anterior–posterior; *lat*—laterality. Analyses of individual regions of interest are shown for posterior regions only, due to the absence of any significant effects in anterior regions.
* $p<0.05$.
** $p<0.01$.
† $p<0.1$.

$F(1,15)=5.57$, $p<0.05$, and the main effect of dependency-length was marginally significant, $F(1,15)=4.14$, $p=0.06$, but there was no *wh*-dependency×dependency length interaction, $F=1.36$. There was also a *wh*-dependency×laterality interaction, $F(2,30)=3.60$, $p<0.05$. Additional comparisons within individual ROIs revealed that these effects reflected topographic differences between the short-distance and long-distance conditions. The effect of the *wh*-dependency was significant in the short-distance conditions when all three levels of the laterality factor were combined, $F(2,30)=5.76$, $p<0.05$, and individually within each level of the laterality factor (left: $F(1,15)=5.48$, $p=0.034$; midline: $F(1,15)=5.91$, $p=0.028$; right: $F(1,15)=5.16$, $p=0.038$). In the long-distance conditions, on the other hand, there was a significant *wh*-dependency×laterality interaction, $F(2,30)=4.10$, $p<0.05$, due to the fact that there was a marginal effect of the *wh*-dependency at midline electrodes, $F(1,15)=3.73$, $p=0.073$, but no effect at either left-posterior or right-posterior regions. Summarizing, at this time interval the posterior positivity was present in both short-distance and long-distance conditions alike. There was some evidence for a difference in the topographic extent of the two effects, but since these differences are not strong enough to motivate separate generators, they are not discussed further here.

At the 700–900 ms interval at the posterior electrodes, there was a significant main effect of the *wh*-dependency, $F(1,15)=6.68$, $p<0.05$, and no interaction of *wh*-dependency and dependency-length. However, there was a significant *wh*-dependency×laterality interaction, $F(2,30)=3.71$, $p=0.05$, and a marginally significant *wh*-dependency×dependency-length×laterality interaction, $F(2,30)=2.87$, $p=0.073$. Additional comparisons within individual ROIs revealed that these effects were due to the fact that the effect of the *wh*-dependency in the short-distance conditions did not reach significance at any level of the laterality factor, but in the long-distance conditions was significant at the posterior–midline region, $F(1,15)=5.10$, $p<0.05$, and at the right-posterior region, $F(1,15)=4.51$, $p=0.051$. Summarizing, at this interval we find similar evidence for a continued posterior positivity in short-distance and long-distance conditions alike. Given that this effect was already present at earlier intervals, any differences found here play no role in the overall conclusions of the study.

An additional analysis was conducted to compare the peak amplitude of the late positivity in the short-distance and long-distance conditions. Due to the earlier onset of the positivity in the short-distance conditions, this required comparison of different time intervals in different conditions. For each level of the dependency-length factor, the peak amplitude of the P600 was selected by finding the largest difference between the *wh*-dependency and control conditions in moving averages based on 200 ms intervals computed every 5 ms. For the short-distance conditions the

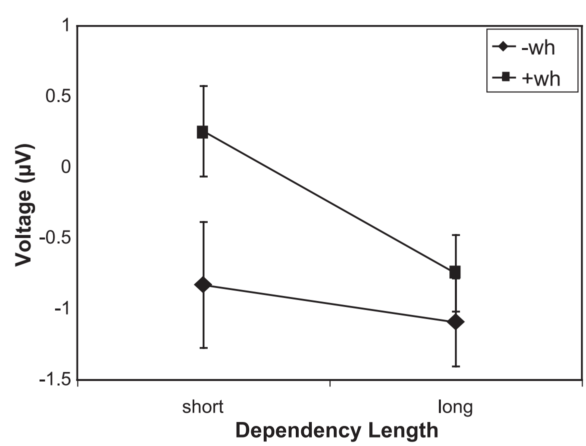

Fig. 3. Mean voltages at posterior electrode sites in the 300–500 ms interval following the final verb, showing a positivity in the short-distance conditions but not in the long-distance conditions.

peak amplitude of 1.73μV was reached at electrode PZ in the 570–770 ms interval, for the long-distance conditions the peak amplitude of 1.68 μV was reached at electrode PZ in the 700–900 ms interval. Mean amplitudes from these intervals were entered into the same repeated measures ANOVA used for all other single word analyses. Results showed significant effects of the *wh*-dependency factor. This is unsurprising, since the time intervals were selected around the maximum difference between *wh*-dependency conditions and their controls. More importantly, there were no significant interactions of the *wh*-dependency factor with dependency length, both in the overall ANOVA and in separate analyses conducted at anterior and posterior channels and at each of the six individual ROIs (all F's<1). Therefore, there is no evidence for a difference in peak amplitude of the P600 in short-distance and long-distance conditions.

4.3. Multi-word analyses

4.3.1. Beginning of sentence

Due to the different number of words in the *wh*-dependency in the short-distance and long-distance conditions, multi-word analyses were conducted separately within each level of the dependency length factor.

In order to test for the possibility of artifactual sustained effects, both cumulative and noncumulative multi-word analyses were conducted on the opening words of the sentences, prior to the onset of the *wh*-dependency. These analyses showed that there were no significant differences at any word position prior to the start of the *wh*-dependency.

4.3.2. Long distance conditions

The results of the cumulative analysis, which measures the divergence between conditions since the beginning of the *wh*-dependency, indicated a sustained anterior negativity (SAN) that remained highly significant throughout the *wh*-dependency. The SAN was reflected both in the main effect of the *wh*-dependency and by the interaction of the *wh*-dependency with the anterior posterior factor. The interaction was due to the fact that the *wh*-dependency condition was more negative than the control condition at anterior sites. There was an effect of the *wh*-dependency at posterior sites only at the position of the complementizer *that* at the beginning of the third clause. Results of the ANOVA in the cumulative analysis are shown in Table 5 and the amplitude and scalp distribution of the SAN is illustrated in Fig. 4A.

However, in the non-cumulative analysis, which measures the separate contribution of each individual word region to the SAN, there was no main effect of the *wh*-dependency in any region other than at the complementizer *that* [$F(1, 15)=7.16$, $p<0.05$]. This effect is likely due to the N400 response elicited at the preceding verb as described in the single word analyses. There were no significant interactions that involved the *wh*-dependency factor and showed significant effects within the levels of the interaction. The amplitude and scalp distribution of the sustained negativity in the non-cumulative analyses is shown in Fig. 4B.

In light of the finding that later words in the *wh*-dependency made little contribution to the sustained negativity, additional analyses were conducted in order to test whether the sustained negativity was artificially reduced by the linear detrending procedure that was used to correct for slow drifts in the raw EEG recordings. If a larger SAN effect were masked by the detrending procedure, then it should be more apparent in analyses of the raw data. However, cumulative analyses of raw and detrended data at electrode CZ for each word of the long-distance conditions indicated that the detrending procedure did not mask a SAN. Analyses at the words preceding the onset of the *wh*-dependency showed that detrending removed a spurious difference between conditions. Analyses at each successive word of the *wh*-dependency showed that whereas the cumulative SAN was present throughout the dependency in the detrended data, as shown in Fig. 4A, it was present only at the beginning of the dependency in the raw data. Taken together, this suggests that the detrending procedure successfully removed artifacts from the data.

As an additional test of the effect of position on the sustained negativity, we conducted analyses that included a *position* factor in order to compare the cumulative measure of the sustained negativity at the beginning of the two clauses of the *wh*-dependency. These analyses focused on anterior channels and thus the anterior–posterior factor was not included in this ANOVA. An analysis involving the determiner *the* at the beginning of the second and third clauses showed a main effect of *wh*-dependency, $F(1, 15)=16.5$, $p<0.001$, an interaction of *wh*-dependency and position, $F(1, 15)=9.21$, $p<0.01$, and a marginally signifi-

Table 5
Summary of ANOVA *f*-values in the cumulative multi-word analysis of the *wh*-dependency interval in the long-distance conditions

	the	lieutenant	knew	that	the	shrewd	witness	would
wh (1,15)	–	6.63*	6.52*	15.74**	9.71**	10.58**	4.45†	6.40*
wh×*ant* (1,15)	6.68*	15.36**	10.28**	22.58**	13.71**	21.95**	13.85**	11.00**
Anterior regions only: *wh* (1,15)	5.71*	14.27**	11.35**	25.50**	25.21**	26.06**	10.42**	10.82**
Posterior regions only: *wh* (1,15)	–	–	–	6.39*	–	–	–	–

* $p<0.05$.
** $p<0.01$.
† $p<0.1$.

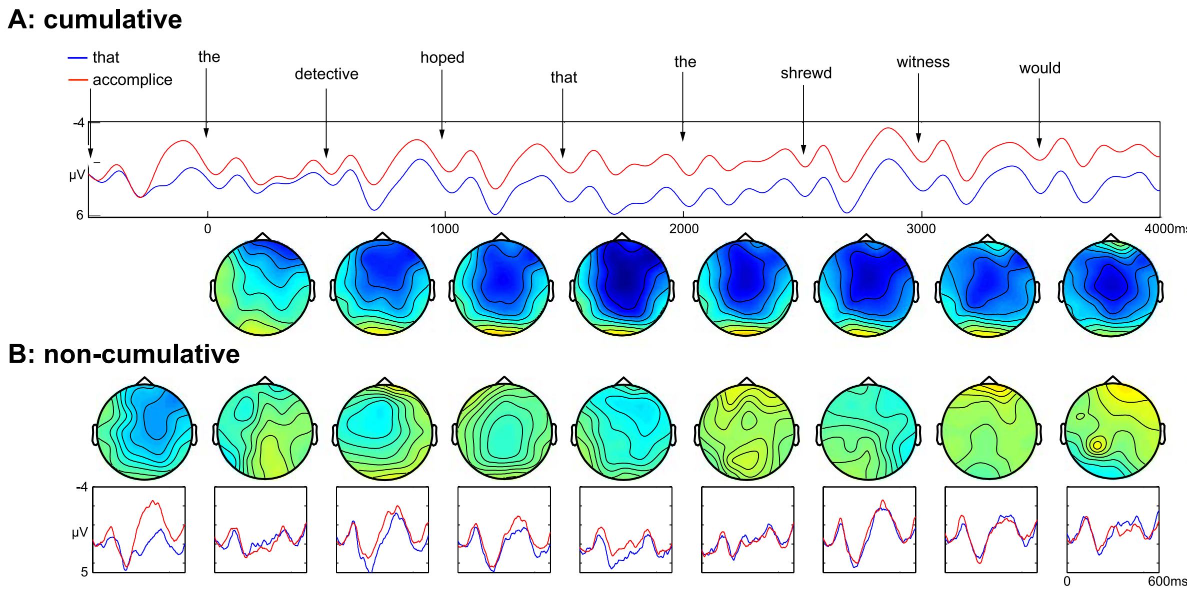

Fig. 4. Sustained negativity elicited in the comparison of the long-distance *wh*-dependency condition (red) and its control (blue), as reflected in waveform plots for midline electrode CZ, and topographic scalp maps based on mean voltage differences in successive 500 ms intervals. (A) Cumulative analysis, similar to Fiebach et al. [16] suggests that a sustained negativity persists throughout the *wh*-dependency. (B) Non-cumulative analysis, similar to King and Kutas [47], indicates that all words of the dependency do not contribute equally to the negativity. The topographic map for the first word of the cumulative analysis is not shown, since it is identical to the corresponding region in the non-cumulative analysis.

cant interaction of *wh* dependency and laterality with position, $F(1,15)=2.81$, $p=0.08$. An analysis involving the immediately following word, e.g., *accomplice* vs. *shrewd*, showed a main effect of the *wh* dependency, $F(1, 15)=21.45$, $p<0.001$, but no interaction of *wh* dependency and position and no interaction of *wh* dependency with laterality and position, F's<1.5. Thus, these analyses confirm the conclusion that the sustained negativity grows steeply at the beginning of the *wh* dependency but shows little change thereafter.

4.3.3. Short distance conditions

In the cumulative analysis for the short-distance conditions there was a main effect of the *wh*-dependency at the adjective and the following noun, as shown in Table 6, due to a negativity that was strongest at midline posterior channels. As shown by the comparison of cumulative and non-cumulative analyses in Fig. 5, this effect may reflect a

Table 6
Summary of ANOVA f-values in the cumulative multi-word analysis of the *wh*-dependency interval in the short-distance conditions

	the	shrewd	witness	would
wh (1,15)	–	5.04*	4.25†	–
ant (1,15)	–	–	–	4.25†
wh×*ant* (1,15)	6.75*	3.97†	–	–
Anterior regions only: *wh* (1,15)	–	–	–	–
Posterior regions only: *wh* (1,15)	7.08*	10.08**	4.55*	4.58*

* $p<0.05$.
** $p<0.01$.
† $p<0.1$.

negativity caused by a lexical category contrast (*that* vs. *accomplice*) at the beginning of the dependency. A similar effect is visible in the corresponding region in the non-cumulative analysis of the long-distance conditions. This negativity remained significant at posterior channels throughout the dependency. The peak amplitude of the negativity was around 2 μV less than the peak amplitude of the negativity in the corresponding analysis of the long-distance conditions. A 2×2 ANOVA based on the word-ROI pairings that showed the greatest negativity in the two conditions (long distance: midline-anterior at complementizer *that*; short-distance: midline–posterior at determiner, i.e. *the*) showed a main effect of the *wh*-dependency factor, $F(1,15)=26.82$, $p<0.0001$, no effect of length, $F<1.4$, and a marginally significant *wh*-dependency×length interaction, $F(1,15)=3.29$, $p=0.09$, due to greater negativity in the long-distance conditions. Note that since the amplitude difference between conditions was already apparent in comparison of the first clause of each *wh*-dependency, it would be inappropriate to view this as a reflection of holding a *wh*-filler in working memory for longer.

In the non-cumulative analysis, results differ from the cumulative analysis only from the second region onwards. At the adjective there was a main effect of the *wh*-dependency, $F(1,15)=5.49$, $p<0.05$, and a significant interaction of *wh*-dependency, laterality, and the anterior–posterior factor, $F(2,30)=4.14$, $p<0.05$, due to the fact that the *wh*-dependency condition was more negative than the control at left anterior sites and more positive at posterior sites. There were no other main effects or significant interactions that involved the *wh*-dependency factor and showed significant effects within the levels of the

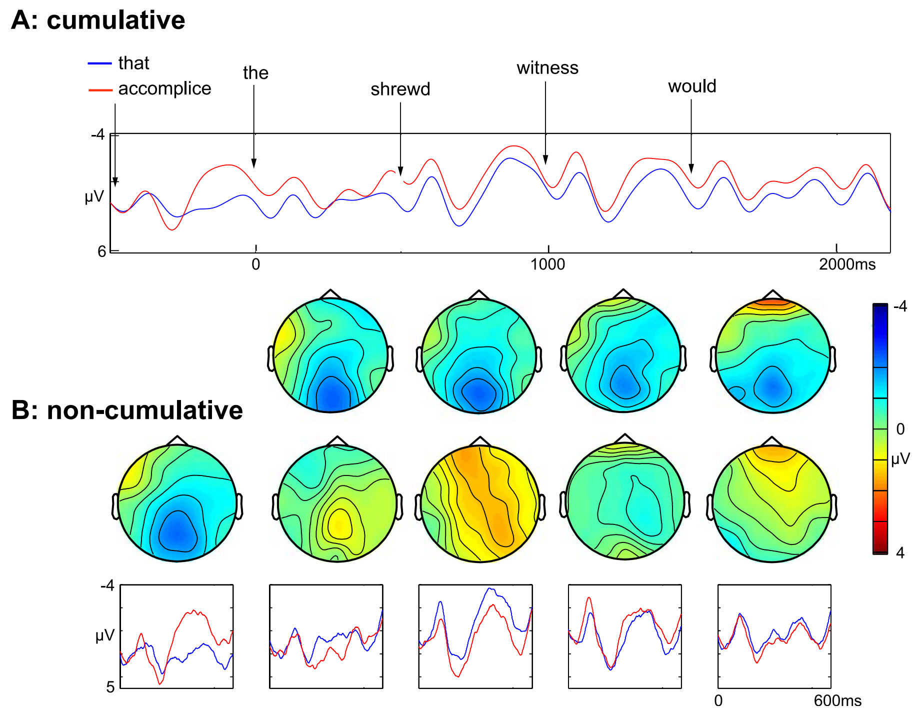

Fig. 5. Sustained negativity elicited in the comparison of the short-distance *wh*-dependency condition (red) and its control (blue). All other properties are identical to Fig. 4. The topographic map for the first word of the cumulative analysis is not shown, since it is identical to the corresponding region in the non-cumulative analysis.

interaction. A comparison of the sustained negativity in the cumulative and non-cumulative analyses is shown in Fig. 5.

5. Discussion

The aim of the present study was to clarify the locus of dependency-length effects in the processing of filler-gap dependencies. Although previous behavioral studies provide evidence that longer syntactic dependencies incur greater processing cost, and this conclusion is strengthened by the results of Experiment 1, it is unclear what the source of this increased cost is. Dependency-length effects could be due to the increasing cost of holding an incomplete dependency in working memory, or to the increased cost of reactivating the head of the dependency when it is far away from the gap site, or to the increased cost of syntactically integrating phrases that are far apart in surface word order, or to a combination of more than one of these factors. Based on previous findings that holding an incomplete *wh*-dependency elicits a sustained anterior negativity [16,47,62,63], and that completion of a *wh*-dependency elicits a P600 [16,44], we investigated whether these ERP components are affected by manipulation of the length of the *wh*-dependency. We found effects of dependency length on both of these ERP components, but in both cases we suggest that the results lead to specific conclusions about which sub-processes of building long-distance dependencies are length-sensitive.

We first discuss the ERP response components that are directly related to the dependency length manipulation, and then discuss other ERP effects.

5.1. Incomplete dependencies and the sustained negativity

We found that holding an incomplete *wh*-dependency in memory elicits a sustained negativity. This negativity persists throughout the *wh*-dependency, until the point of completion of the dependency, and may be interpreted as a reflection of the cost of holding the *wh*-phrase in working memory [16,47].

In the long-distance conditions, the sustained negativity showed an anterior focus, similar to the distribution observed in previous studies. However, we were able to gain a clearer view of the time-course of the sustained negativity by comparing the results of two analyses, a cumulative analysis that used a common baseline for all words, along the lines of Fiebach et al. [16], and a non-cumulative analysis that used a separate baseline for all words, in order to assess the contribution of each successive word to the sustained effect, along the lines of King and Kutas [47]. Although the cumulative analysis indicates that the negativity persists throughout the two-clause *wh*-dependency, the non-cumulative analysis reveals that the growth of the negativity is confined to the first clause of the dependency plus the complementizer *that* at the beginning of the second clause. In the second clause of the dependency, successive words do not contribute greater negativity to the effect. In fact, the last four words of the dependency make no contribution at all to the sustained negativity, contrary to the predictions of an account in which the sustained negativity reflects gradually increasing memory costs associated with the incomplete *wh*-dependency.

In the short-distance conditions, cumulative analyses revealed a weaker sustained negativity than in the long-distance conditions, in this case with a posterior focus. Comparison of the cumulative and non-cumulative analyses suggests that this posterior negativity was contributed primarily by a negativity associated with a single-word at the start of the dependency, resulting from a lexical class difference between the complementizer *that* and the common noun in the *wh*-phrase. Subsequent words made no further contribution to the negativity.

In sum, there was a difference between the sustained negativity in the short-distance and long-distance conditions, with a greater amplitude negativity in the long-distance conditions. However, it would be inappropriate to conclude from this that these results show that the cost of storing an incomplete *wh*-dependency is length-sensitive, as argued by Fiebach et al. [16]. The difference between the two conditions was apparent already in the first clause of each *wh*-dependency, where the difference could not be attributed to length sensitivity, and the second clause of the long-distance dependency made no additional contribution to the *wh*-dependency.

The fact that the short-distance and long-distance conditions differed already in the first clause of the *wh*-dependency was not predicted. The difference may have resulted from the fact that in controlling overall sentence-length and the position of the completion of the *wh*-dependency it was necessary to vary the position of the start of the short-distance and long-distance *wh*-dependencies. We tentatively suggest that in the same way that N400 amplitudes are known to be reduced overall and less susceptible to word-frequency variation at later positions in a sentence [79,86], it is possible that negativities related to the presence of incomplete syntactic dependencies are also attenuated at later positions in a sentence, where contextual constraints imposed by the earlier portion of the sentence are greater.

Our conclusion that we find no evidence for length-sensitivity in the SAN appears to be at odds with the conclusion of Fiebach et al. that the SAN reflects length-sensitive processes. We should emphasize that the results of the two studies are not incompatible. Fiebach et al. report a growing SAN in a cumulative analysis of a long-distance dependency that spans 2600 ms. This corresponds roughly to the first clause of the long-distance dependency in the current study, where our cumulative analysis also suggests an increasing trend in the amplitude of the SAN. However, based on the greater length of our long-distance dependency and based upon the stricter criterion of the non-cumulative analysis, we conclude that the SAN does not provide evidence that the cost of storing the *wh*-phrase in memory increases over the course of the dependency. Of course, the lack of clear evidence for a length effect on the SAN could simply indicate that the SAN is an insufficiently sensitive measure of working memory load in sentence processing.

5.2. Syntactic integration and the P600

At the embedded verb that marked the completion of the *wh*-dependency, a late positivity was observed in the *wh*-dependency conditions, relative to the control conditions. In both the long-distance and short-distance conditions, this response showed the timing and the posterior scalp distribution characteristic of a P600 response. There were no differences in the amplitude of the P600 response at the completion of short-distance and long-distance *wh*-dependencies, but the P600 had an earlier onset in the short-distance condition than in the long-distance condition, appearing in the 300–500 ms interval in the short-distance conditions and in the 500–700 ms interval in the long-distance conditions. It should be noted that given the temporal granularity of our analyses, the time difference between the conditions may be smaller than 200 ms. This finding has a number of implications for the interpretation of what the P600 reflects.

First, the fact that the P600 was elicited by the completion of a well-formed filler-gap dependency lends further support to the claim that the P600 is not restricted to processes involving the detection of violations, diagnosis or reanalysis, as assumed in much earlier work, but also reflects undisrupted, successful syntactic processing [13,15,16,35,44]. Second, our results are consistent with the suggestion of Kaan and colleagues that the P600 amplitude is sensitive to the number of syntactic integration operations that occur as a new word is processed [44]. In both the short and long *wh*-dependency conditions in the current study the processing of the verb involved one more syntactic operation than in the control conditions.

Third, the finding that the timing of the P600 was affected by the dependency length manipulation but the amplitude of the P600 was not affected appears to be at odds with the more general claim of Kaan and colleagues that the amplitude of the P600 is an index of syntactic integration difficulty. The timing differences in the P600 together with the complexity rating results from Experiment 1 indicate that the dependency-length manipulation successfully varied processing difficulty, and that length specifically affects some aspect of the completion of a *wh*-dependency. This argues against the suggestion that two-clause dependencies might be no more difficult from one-clause dependencies because they in fact consist of a pair of one-clause dependencies, a suggestion that could be derived from cross-language studies of long-distance dependencies in the linguistics literature [4–6,29,57,58].

We suggest that processes at the completion of a syntactic dependency are collectively sensitive to the length of the dependency, but that the P600 reflects one sub-process of integration that is not itself length-sensitive. Gibson ([31], p. 11) draws a distinction between the reactivation of the head of a syntactic dependency, a process that is claimed to be length sensitive, and the truly integrative processes that follow reactivation, consisting of syntactic thematic role assignment and compositional semantic interpretation. Gibson assumes that these processes consume a fixed amount of processing resources for each dependency, and thus predicts that they are length-insensitive. We therefore suggest that the interpretation of the P600 presented in Kaan et al. [44] and Fiebach et al. [16] be narrowed, such that it does not reflect all of the processes that Gibson includes under the heading of 'integration cost'. The P600 amplitude reflects the syntactic and semantic operations involved in confirming the compatibility of the filler and the verb for thematic role assignment, and compositionally interpreting the verb and its arguments, and is thus insensitive to the length of the dependency. The process of reactivating a filler is not reflected in the amplitude of the P600, and only indirectly affects the P600, by causing a delay in the onset of the P600 response in longer syntactic dependencies. This accounts for the finding that the dependency-length manipulation in our study affected the timing but not the amplitude of the P600 response.

This account of the functional status of the P600 may also explain why no P600 was elicited at the clause-final verb in German in the study by Fiebach et al. [16]. If the P600 amplitude reflects the number of thematic role assignments that are computed at the position of the verb, then this should be unaffected by word order variation in a verb-final clause in German, since the number of thematic role assignments that are confirmed at the verb position is the same for any ordering of the arguments of the verb. Fiebach et al. did observe a P600-like positivity at the pre-verbal NP in the comparison of object-initial and subject-initial *wh*-questions. They interpret this effect as a reflection of the creation of a gap immediately to the right of the nominative NP in the object-initial condition, with this implication that the P600 reflects gap creation rather than syntactic or semantic integration with a verb, contrary to the proposal by Kaan et al. However, another possibility is that the P600 in the German study reflects the marked word order in the object-initial conditions. Although German freely allows accusative NPs to be moved to the left of nominative NPs, the word order in which a nominative definite NP appears to the right of a sequence of adverbial PP phrases (see example (4) above) is highly marked in German. The P600 may reflect the fact that this word order is dispreferred, and thus be an instance of the standard P600 elicited by incongruous structures. This effect may be related to a previous finding of a P600-like response elicited by unambiguous yet non-canonical word order in German [75].

In sum, we suggest that the P600 that appears at verb positions in congruous sentences is specifically associated with the number of thematic role assignments completed at that position. It is possible that the 'integrational positivity' is a special case of a more general phenomenon, and that other types of structure-building operations in congruous sentences will also elicit a late positivity, but current evidence leaves this as an open issue.

It is also important to consider how the P600 response elicited by congruous sentences relates to the P600 response elicited by ungrammatical or garden-path sentences in many previous studies. In both areas, we can distinguish those processes that affect the timing of the P600 response and those processes that affect the amplitude of the P600. We suggest that the P600 reflects structure-building processes in congruous sentences, and structure re-building and rechecking processes in garden path and ungrammatical sentences respectively. This suggests that the timing of the onset of the P600 should depend on the time required to identify and activate the elements that will be assembled, re-assembled, or re-checked. In the current study, we argued that the timing of the P600 reflects the time required to activate the filler *wh*-phrase that is integrated with the verb. Systematic variation in P600 timing has also been observed in studies of reanalysis in German [23,26]. When ambiguous verb-final clauses in German must be reanalyzed from the preferred subject-initial order to the dispreferred object-initial order, the positivity associated with this reanalysis occurs earlier in ambiguous relative clauses ('P345') than in ambiguous complement clauses ('P600'). Friederici et al. propose that the positivity is more short-lived in the relative clause reanalysis since one of the two overt NPs in these constructions, the relative pronoun, does not need to be restructured, whereas both NPs must be restructured in the complement clause reanalysis. This contrast may also account for why the positivity has an earlier onset in the relative clause reanalysis, since fewer NPs must be reactivated in order for reanalysis processes to begin.

Note that although the P600 appears to be elicited by structure-building operations in congruous and incongruous sentences alike, it is unlikely that a general account of the P600 can be given that relies entirely on the number of structure-building operations carried out at a given word. The finding that P600 amplitude is greater for ungrammatical sentences than for similar sentences that allow reanalysis [70] speaks against such an account, because it is likely that more structure-building operations are carried out in cases of successful reanalysis than in cases where reanalysis is impossible due to ungrammaticality. Therefore, the processes measured by the P600 must also reflect additional processes, such as structure-checking and diagnosis processes [26].

5.3. Intermediate verb: N400

In contrast to the P600 elicited at the completion of the *wh*-dependency, a different response was observed at the intermediate verb. In the long-distance *wh*-dependency conditions, where the dependency spanned two clauses, a late negativity was elicited at the intermediate verb, which showed the timing and the centro-parietal scalp distribution characteristic of an N400 response. No N400 was observed at this region in the short-distance conditions, since the *wh*-

dependency condition and its control were identical up to this region.

It is well established that all content words in a sentence elicit an N400 response. Variation in the amplitude of the N400 response has been observed in connection with either compositional interpretive processes in sentence contexts or lexical interpretive processes in word lists. In sentence contexts, the N400 is typically elicited in contexts that are syntactically well-formed but semantically anomalous [52,53] or semantically unexpected [51,84]. Systematic variation in N400 amplitudes is also observed in word lists, as a function of word frequency and of repetition [87], and by other factors that increase the difficulty of lexical retrieval [51]. Even in sentence contexts some N400 variation appears to be related to lexical processes such as priming, rather than to congruity or expectancy at the level of compositional interpretation [14,17,50,51].

In the current study the N400 was elicited at the second verb in a sentence that began like 'The lieutenant knew which accomplice the detective hoped that...', where the argument structure of the verb *hope* prevents the *wh*-phrase *which accomplice* from being interpreted as a direct object. We suggest that this N400 may be understood as an instance of the N400 effects associated with lexical predictive processes and priming. A sequence like *knew which accomplice the detective* may create a strong expectation for an upcoming transitive verb, relative to the less-constraining control condition that contains the sequence *knew that the detective*. Therefore, when the sentence complement verb *hope* is encountered, this is more strongly incompatible with the lexical expectation in the *wh*-dependency condition than in the control condition. We favor this account of the N400 effect over one in which the N400 reflects an 'argument structure violation' due to premature completion of the *wh*-dependency, due to the fact that there is good evidence from behavioral studies that speakers are able to use argument structure information to defer completion of an infelicitous syntactic dependency [2,73], and also due to the fact that studies that have presented unambiguous argument structure violations have elicited both an N400 and a P600 ERP response [22].

6. Conclusion

The aim of this study was to use the fine-grained temporal information provided by ERPs to clarify the locus of the well-known preference for shorter syntactic dependencies. Previous behavioral studies and Experiment 1 of the current study indicate that longer syntactic dependencies incur a greater processing load, but it remains unclear what the source of this length-sensitivity is. In the case of filler-gap dependencies, the effect of dependency-length could be due to an increasing cost of holding a *wh*-phrase in working memory, or due to increased cost for reactivating the *wh*-phrase at the conclusion of the dependency, or due to greater cost of syntactic and semantic integration when the *wh*-phrase is combined with the verb that assigns it a thematic role.

By comparing single-word and multiple-word ERP responses elicited by short-distance and long-distance *wh*-dependencies, the current study was able to more narrowly pinpoint the source of the dependency-length effect. Holding an incomplete *wh*-dependency in working memory gave rise to a sustained anterior negativity, as shown in previous studies [16,47]. Although the peak amplitude of the anterior negativity was greater in the long-distance *wh*-dependency conditions, analyses of non-cumulative multi-word ERPs and the fact that amplitudes differed already in the first clause of the long-distance *wh*-dependency suggest that these results do not provide evidence for a length-sensitive storage process. Completion of the *wh*-dependency elicited a P600 response, as also seen in earlier studies [16,44]. The amplitude of the P600 was not affected by dependency-length, but the timing of this response was delayed in the long-distance *wh*-dependency condition. We interpret this finding as evidence that the P600 reflects a length insensitive process of syntactic thematic role assignment and compositional semantic interpretation, and that the source of dependency-length effects is the reactivation of the *wh*-phrase that is required prior to syntactic and semantic integration. Although these findings lead us to a more narrow conception of the role of the P600 in the formation of well-formed syntactic dependencies than proposed in previous studies [44], these conclusions about the source of length-sensitivity in processing syntactic dependencies conform very well to the predictions of one version of Gibson's model of linguistic complexity ([31], p. 65ff.; [32]), and thus show the value of using ERP measures to inform cognitive models of linguistic processes.

Acknowledgments

Preparation of this paper was supported by grants from the James S. McDonnell Foundation (CNS99-31T), the National Science Foundation (BCS-0196004, SBR-9977628), the Human Frontier Science Program (RGY-0134), and by a Semester Research Award from the University of Maryland. We are also grateful to Danny Dilks, Robert Ellis, Baris Kabak, and Kaia Wong, who provided a great deal of help in the early stages of this project, and to Daniel Garcia-Pedrosa and Ellen Lau for assistance with data collection and analysis.

References

[1] S. Aoshima, C. Phillips, A. Weinberg, Processing filler-gap dependencies in a head final language, Journal of Memory and Language 51 (2004) 23–54.

[2] J. Boland, M.K. Tanenhaus, S.M. Garnsey, G.N. Carlson, Verb argument structure in parsing and interpretation: evidence from *wh*-questions, Journal of Memory and Language 34 (1995) 774–806.
[3] C.M. Brown, P. Hagoort, On the electrophysiology of language comprehension: implications for the human language system, in: M. Crocker, M. Pickering, C. Clifton Jr. (Eds.), Architectures and Mechanisms for Language Processing, Cambridge University Press, Cambridge, UK, 1999, pp. 213–237.
[4] B. Bruening, Two types of *wh*-scope marking in Passamaquoddy, Natural Language and Linguistic Theory 22 (2004) 229–305.
[5] N. Chomsky, Conditions on transformations, in: S. Anderson, P. Kiparsky (Eds.), A Festschrift for Morris Halle, Holt, Rinehart and Winston, New York, 1973, pp. 232–286.
[6] S. Chung, The Design of Agreement: Evidence from Chamorro, University of Chicago Press, Chicago, 1998.
[7] G. Cinque, Types of A′ dependencies, MIT Press, Cambridge, MA, 1990.
[8] E.A. Cowper, Constraints on sentence complexity: a model for syntactic processing. Ph.D. thesis, Brown University, Providence, RI 1976.
[9] S. Crain, J.D. Fodor, How can grammars help parsers? in: D. Dowty, L. Kartunnen, A.M. Zwicky (Eds.), Natural Language Parsing: Psycholinguistic, Computational, and Theoretical Perspectives, Cambridge University Press, Cambridge, UK, 1985, pp. 94–128.
[10] S. Crain, J.D. Fodor, Competence and performance in child language, in: E. Dromi (Ed.), Language and Cognition: A Developmental Perspective, Ablex, Norwood, N.J., 1993, pp. 141–171.
[11] M. Crocker, On the nature of the principle-based sentence processor, in: C. Clifton, L. Frazier, K. Rayner (Eds.), Perspectives on Sentence Processing, Erlbaum, Hillsdale, NJ, 1994, pp. 245–265.
[12] G. Fanselow, R. Kliegl, M. Schlesewsky, Processing difficulty and principles of grammar, in: S. Kemper, R. Kliegl (Eds.), Constraints on Language, Kluwer Academic, Dordrecht, 1999, pp. 171–201.
[13] S. Featherston, M. Gross, T.F. Münte, H. Clahsen, Brain potentials in the processing of complex sentences: an ERP study of control and raising constructions, Journal of Psycholinguistic Research 29 (2000) 141–154.
[14] K.D. Federmeier, M. Kutas, A rose by any other name: long-term memory structure and sentence processing, Journal of Memory and Language 41 (1999) 469–495.
[15] C. Felser, H. Clahsen, T.F. Münte, Storage and integration in the processing of fillergap dependencies: an ERP study of topicalization and *wh*-movement in German, Brain and Language 87 (2003) 345–354.
[16] C. Fiebach, M. Schlesewsky, A. Friederici, Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German *wh*-questions, Journal of Memory and Language 47 (2002) 250–272.
[17] I. Fischler, P.A. Bloom, D.G. Childers, S.E. Roucos, N.W. Perry, Brain potentials related to stages of sentence verification, Psychophysiology 20 (1983) 400–409.
[18] J.D. Fodor, Parsing strategies and constraints on transformations, Linguistic Inquiry 9 (1978) 427–473.
[19] L. Frazier, Syntactic processing: evidence from Dutch, Natural Language and Linguistic Theory 5 (1987) 519–560.
[20] L. Frazier, G. Flores d'Arcais, Filler-driven parsing: a study of gap filling in Dutch, Journal of Memory and Language 28 (1989) 331–344.
[21] A.D. Friederici, Towards a neural basis for auditory sentence processing, Trends in Cognitive Sciences 6 (2002) 78–84.
[22] A.D. Friederici, S. Frisch, Verb argument structure processing: the role of verbspecific and argument-specific information, Journal of Memory and Language 43 (2000) 476–507.
[23] A.D. Friederici, A. Mecklinger, Syntactic parsing as revealed by brain responses: first-pass and second-pass processes, Journal of Psycholinguistic Research 25 (1996) 157–176.
[24] A.D. Friederici, E. Pfeifer, A. Hahne, Event-related brain potentials during natural speech processing: effects of semantic, morphological, and syntactic violations, Cognitive Brain Research 1 (1993) 183–192.
[25] A.D. Friederici, A. Hahne, A. Mecklinger, The temporal structure of syntactic parsing: early vs. late effects elicited by syntactic anomalies, Journal of Experimental Psychology. Learning, Memory, and Cognition 22 (1996) 1219–1248.
[26] A.D. Friederici, A. Mecklinger, K.M. Spencer, K. Steinhauer, E. Donchin, Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials, Cognitive Brain Research 11 (2001) 305–323.
[27] S. Frisch, M. Schlesewsky, D. Saddy, A. Alpermann, The P600 as an indicator of syntactic ambiguity, Cognition 85 (2002) B83–B92.
[28] S.M. Garnsey, M.K. Tanenhaus, R.M. Chapman, Evoked potentials and the study of sentence comprehension, Journal of Psycholinguistic Research 18 (1989) 51–60.
[29] C. Georgopoulos, Variables in Palauan syntax, Natural Language and Linguistic Theory 3 (1985) 59–94.
[30] E. Gibson, A Computational Theory of Human Linguistic Processing: Memory Limitation and Processing Breakdown. PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 1991.
[31] E. Gibson, Linguistic complexity: locality of syntactic dependencies, Cognition 68 (1998) 1–76.
[32] E. Gibson, Dependency locality theory: a distance-based theory of sentence processing difficulty, in: A. Marantz, Y. Miyashita, W. O'Neil (Eds.), Image, Language, Brain: Papers from the first Mind Articulation Project Symposium, MIT Press, Cambridge, MA, 2000, pp. 95–126.
[33] E. Gibson, G. Hickok, Sentence processing with empty categories, Language and Cognitive Processes 8 (1993) 147–161.
[34] E. Gibson, G. Hickok, C.T. Schütze, Processing empty categories: a parallel approach, Journal of Psycholinguistic Research 23 (1994) 381–405.
[35] A. Gouvea, Syntactic Complexity: Cross-linguistic Differences and ERP Evidence. PhD dissertation, University of Maryland, College Park, 2003.
[36] S. Greenhouse, S. Geisser, On methods in the analysis of profile data, Psychometrika 24 (1959) 95–112.
[37] P. Hagoort, C.M. Brown, Brain responses to lexical ambiguity resolution and parsing, in: C. Clifton Jr., L. Frazier, K. Rayner (Eds.), Perspectives on Sentence Processing, Erlbaum, Hillsdale, NJ, 1994, pp. 45–80.
[38] P. Hagoort, C.M. Brown, J. Groothusen, The syntactic positive shift (SPS) as an ERP measure of syntactic processing, Language and Cognitive Processes 8 (1993) 439–484.
[39] P. Hagoort, C.M. Brown, L. Osterhout, The neurocognition of syntactic processing, in: C.M. Brown, P. Hagoort (Eds.), The Neurocognition of Language, Oxford University Press, Oxford, UK, 1999, pp. 273–316.
[40] K. Hakuta, Grammatical description versus configurational arrangement in language acquisition: the case of relative clauses in Japanese, Cognition 9 (1981) 197–236.
[41] R. Hudson, 1996. The difficulty of so-called 'self-embedded' structures. Manuscript, University College, London. http://www.phon.ucl.ac.uk/home/dick/home.htm.
[42] R. Jackendoff, P. Culicover, A reconsideration of dative movement, Foundations of Language 7 (1971) 392–412.
[43] E. Kaan, Investigating the effects of distance and number interference on processing subject-verb agreement: an ERP study, Journal of Psycholinguistic Research 31 (2002) 165–193.
[44] E. Kaan, A. Harris, E. Gibson, P. Holcomb, The P600 as an index of syntactic integration difficulty, Language and Cognitive Processes 15 (2000) 159–201.
[45] J. Kimball, Seven principles of surface structure parsing in natural language, Cognition 2 (1973) 15–47.
[46] J. King, M. Just, Individual differences in syntactic processing: the role of working memory, Journal of Memory and Language 30 (1991) 580–602.

[47] J. King, M. Kutas, Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading, Journal of Cognitive Neuroscience 7 (1995) 376–395.

[48] R. Kluender, M. Kutas, Bridging the gap: evidence from ERPs on the processing of unbounded dependencies, Journal of Cognitive Neuroscience 5 (1993) 196–214.

[49] R. Kluender, M. Kutas, Subjacency as a processing phenomenon, Language and Cognitive Processes 8 (1993) 573–633.

[50] J. Kounios, P.J. Holcomb, Structure and process in semantic memory: evidence from event-related brain potentials and reaction times, Journal of Experimental Psychology. General 121 (1992) 459–479.

[51] M. Kutas, K.D. Federmeier, Electrophysiology reveals semantic memory use in language comprehension, Trends in Cognitive Sciences 4 (2000) 463–470.

[52] M. Kutas, S.A. Hillyard, Reading senseless sentences: brain potentials reflect semantic anomaly, Science 207 (1980) 203–205.

[53] M. Kutas, S.A. Hillyard, Brain potentials during reading reflect word expectancy and semantic association, Nature 307 (1984) 161–163.

[54] M. Kutas, C. van Petten, Psycholinguistics electrified: event-related brain potential investigations, in: M.A. Gernsbacher (Ed.), Handbook of Psycholinguistics, Academic Press, New York, 1994, pp. 83–143.

[55] R. Lewis, 1993. An Architecturally-based Theory of Human Sentence Processing. PhD thesis, Carnegie Mellon University, Pittsburgh, PA.

[56] B. MacWhinney, The competition model, in: B. MacWinney (Ed.), Mechanisms of Language Acquisition, Erlbaum, Hillsdale, NJ, 1987, pp. 249–308.

[57] J. McCloskey, Quantifier float and *wh*-movement in an Irish English, Linguistic Inquiry 31 (2000) 57–84.

[58] J. McCloskey, The morphosyntax of *wh*-extraction in Irish, Journal of Linguistics 37 (2001) 67–100.

[59] R. McKinnon, L. Osterhout, Constraints on movement phenomena in sentence processing: evidence from event-related potentials, Language and Cognitive Processes 11 (1996) 495–523.

[60] G. McKoon, R. Ratcliff, G. Ward, Testing theories of language comprehension: an investigation on the on-line lexical decision task, Journal of Experimental Psychology. Learning, Memory, and Cognition 20 (1994) 1219–1228.

[61] G.A. Miller, N. Chomsky, Finitary models of language users, in: R.D. Luce, R.R. Galanter, E. Bush (Eds.), Handbook of Mathematical Psychology, vol. 2, Wiley, New York, 1963, pp. 419–491.

[62] H. Müller, J. King, M. Kutas, Event-related brain potentials to relative clause processing in spoken sentences, Cognitive Brain Research 5 (1997) 193–203.

[63] T.F. Münte, M. Matzke, S. Johannes, Brain activity associated with syntactic incongruity in words and pseudo-words, Journal of Cognitive Neuroscience 9 (1997) 318–329.

[64] T.F. Münte, A. Szentkuti, B. Wieringa, M. Matzke, S. Johannes, Human brain potentials to reading syntactic errors in sentences of different complexity, Neuroscience Letters 235 (1997) 105–108.

[65] H. Neville, J. Nicol, A. Barss, K.I. Forster, M.I. Garrett, Syntactically-based sentence processing classes: evidence from event-related brain potentials, Journal of Cognitive Neuroscience 3 (1991) 151–165.

[66] J. Nicol, D. Swinney, The role of structure in coreference assignment during sentence comprehension, Journal of Psycholinguistic Research 18 (1989) 5–19.

[67] J. Nicol, J. Fodor, D. Swinney, Using cross-modal lexical decision tasks to investigate sentence processing, Journal of Experimental Psychology. Learning, Memory, and Cognition 20 (1994) 1229–1238.

[68] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh Inventory, Neuropsychologia 9 (1971) 97–113.

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

[70] L. Osterhout, P.J. Holcomb, D.A. Swinney, Brain potentials elicited by garden path sentences: evidence of the application of verb information during parsing, Journal of Experimental Psychology. Learning, Memory, and Cognition 20 (1994) 768–803.

[71] D. Pesetsky, *wh*-in-situ, movement, and unselective binding, in: E. Reuland, A.G.B. ter Meulen (Eds.), The Representation of (In)definiteness, MIT Press, Cambridge, MA, 1987, pp. 89–129.

[72] M.J. Pickering, G.D. Barry, Sentence processing without empty categories, Language and Cognitive Processes 6 (1991) 229–259.

[73] M.J. Pickering, M.J. Traxler, Strategies for processing unbounded dependencies: lexical information and verb-argument assignment, Journal of Experimental Psychology. Learning, Memory, and Cognition 27 (2001) 1401–1410.

[74] B.L. Pritchett, Subjacency in a principle-based parser, in: R.C. Berwick (Ed.), Principle-based Parsing: Computation and Psycholinguistics, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991, pp. 301–345.

[75] F. Rösler, T. Pechmann, J. Streb, B. Röder, E. Henninghausen, 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 (1998) 150–176.

[76] D.S. Ruchkin, R. Johnson, H. Canoune, W. Ritter, Shortterm memory storage and retention: an event related brain potential study, Electroencephalography and Clinical Neurophysiology 76 (1990) 419–439.

[77] I.A. Sag, J.D. Fodor, Extraction without traces, in: R. Aranovich, W. Byrne, S. Preuss, M. Senturia (Eds.), Proceedings of the 13th Annual Meeting of the West Coast Conference on Formal Linguistics, CSLI, Stanford, CA, 1994, pp. 365–384.

[78] Schulte im S. Walde, 1998. Automatic semantic classification of verbs based on their alternation behavior. Diplomarbeit, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

[79] M.E. Smith, E. Halgren, Event-related potentials during lexical decision: effects of repetition, word frequency, pronounceability and concreteness, Electrencephalography and Clinical Neurophysiology 40 (1987) 417–421.

[80] L.A. Stowe, Parsing *wh*-constructions: evidence for on-line gap location, Language and Cognitive Processes 1 (1986) 227–245.

[81] R.S. Sussman, J.C. Sedivy, The time-course of processing syntactic dependencies: evidence from eye-movements, Language and Cognitive Processes 18 (2003) 143–163.

[82] D. Swinney, L. Osterhout, Inference generation during auditory language comprehension, in: A. Graesser, G. Bower (Eds.), The Psychology of Learning and Motivation, vol. 25, Academic Press, New York, 1990, pp. 17–33.

[83] M.J. Traxler, M.J. Pickering, Plausibility and the processing of unbounded dependencies: an eye-tracking study, Journal of Memory and Language 35 (1996) 454–475.

[84] J.J.A. van Berkum, P. Hagoort, C.M. Brown, Semantic integration in sentences and discourse: evidence from the N400, Journal of Cognitive Neuroscience 11 (1999) 657–671.

[85] C. van Petten, A comparison of lexical and sentence-level context effects in event-related potentials, Language and Cognitive Processes 8 (1993) 485–531.

[86] C. van Petten, M. Kutas, Interactions between sentence context and word frequency in event-related brain potentials, Memory and Cognition 18 (1990) 380–393.

[87] C. van Petten, M. Kutas, R. Kluender, M. Mitchiner, H. McIsaac, Fractionating the word-repetition effect with event-related potentials, Journal of Cognitive Neuroscience 3 (1991) 131–150.

[88] E. Wanner, M. Maratsos, An ATN approach to comprehension, in: M. Halle, J. Bresnan, G.A. Miller (Eds.), Linguistic Theory and Psychological Reality, MIT Press, Cambridge, MA, 1978, pp. 119–161.

[89] V. Yngve, A model and an hypothesis for language structure, Proceedings of the American Philosophical Society 104 (1960) 444–466.