



ACADEMIC
PRESS

Journal of Memory and Language 47 (2002) 250–272

Journal of
Memory and
Language

www.academicpress.com

Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions

Christian J. Fiebach,^{a,*} Matthias Schlesewsky,^b and Angela D. Friederici^a

^a *Max Planck Institute of Cognitive Neuroscience, P.O. Box 500 355, Leipzig D-04303, Germany*

^b *Department of Linguistics, University of Potsdam, Germany*

Received 9 October 2000; revision received 23 August 2001

Abstract

Event-related brain potentials (ERPs) were recorded while participants processed case-unambiguous German subject and object WH-questions with either a long or a short distance between the WH-filler and its gap. A sustained left anterior negativity was observed for object questions with long filler-gap distance but not for short object questions. This negativity was modulated by individual differences in working memory capacity. No comparable negativity was elicited by WHETHER-questions which did not contain a filler-gap dependency. A positive-going ERP effect was observed for short and long object WH-questions at the position of the second noun phrase. We interpret the sustained negativity as reflecting working memory processes required for maintaining the dislocated object in memory. Processing costs associated with integrating the stored element into the phrase structure representation are indicated by the local positivity. These results support the notion of separable syntactic working memory and syntactic integration cost components as causes of processing difficulty in complex sentences. © 2002 Elsevier Science (USA). All rights reserved.

Keywords: Syntax; Working memory; Sentence processing; WH-questions; ERPs

Sentence comprehension critically depends upon the reconstruction of grammatical relations between the arguments and the predicate of a sentence by assigning a hierarchical structure to the input of words (e.g., Frazier & Fodor, 1978). Syntactic processing, generally referred to as parsing, appears to be relatively easy if the sequential order of the linguistic input corresponds to the unmarked word order of the language in question—for example, subject before verb before

object (SVO) in English or SOV in German. In this case, each new word can be attached into the syntactic representation of the sentence without much effort. Language, however, allows variations in the word order which render sentences more difficult to process. One such operation, the formation of relative clauses like [1a] or constituent questions like [1b] by movement of an argument of the verb to clause-initial position, is known as WH-movement (e.g., Chomsky, 1981; Haegeman, 1994). As the two examples show, WH-movement can result in sentences with a noncanonical word order in which the object (i.e., “who”) is located before the subject.

* Corresponding author. Fax: +49-03-41-99-40-113.
E-mail address: fiebach@cns.mpg.de (C.J. Fiebach).

[1a] The reporter who the senator attacked admitted the error.

[1b] Who did the reporter attack?

Linguistic theory (e.g., Chomsky, 1981) assumes that an element that has been moved to the clause-initial position leaves behind a phonologically empty category and that the moved element is coindexed with this so-called gap or trace. Psycholinguistically, this index allows the listener or reader to identify the moved element as a filler for the gap. There is behavioral evidence from cross-modal priming studies that the filler is reactivated when the parser reaches the original position of the moved element (i.e., its gap; cf. Nicol & Swinney, 1989). This finding provides general support for the psychological reality of gaps.

It is, however, less clear how syntactic features are treated when identified in a filler. In principle, there are two possibilities. Either the syntactic features are identified and then reactivated at the trace position as implicated by Nicol and Swinney (1989) or syntactic features are identified and kept active in working memory until the trace is encountered, as was suggested, e.g., by Frazier and colleagues (e.g., Clifton & Frazier, 1989; Frazier & Flores D'Arcais, 1989), Kluender and Kutas (1993a), or by Fanselow, Kliegl, and Schlesewsky (1999).

Several empirical results support the assumption that working memory mechanisms are involved when structurally complex sentences are processed. For example, it has been demonstrated that parsing is more difficult for individuals with low working memory capacity (e.g., Just & Carpenter, 1992; King & Just, 1991) and that maintaining an additional load in working memory during sentence processing modulates parsing processes (e.g., King & Just, 1991; Vos, Gunter, Schriefers, & Friederici, 2001a; Vos, Gunter, Kolk, & Mulder, 2001b). However, such data so far provide only very limited direct evidence on how structural configurations internal to the sentence elicit working memory costs. Furthermore, it is still an unsettled question precisely what kind of cognitive resources are required to correctly process sentences that are syntactically complex (cf. Caplan & Waters, 1995; Just & Carpenter, 1992). In this study, we used event-related brain potentials (ERPs; cf. Garnsey, Tanenhaus, & Chapman, 1989; Kutas & Van Petten, 1994) as a fine-grained measure to explore the processes underlying the parsing of grammatically well-formed sentences with non-

canonical word order. We investigated whether working memory resources are required for resolving sentence-internal dependencies between the arguments of a sentence when these are encountered out of their preferred order. More precisely, we explored whether the processing of an object-argument in clause-initial position leads to increased working memory costs. Furthermore, this study was designed to more precisely determine where in the sentence this working memory load appears and how long it has to be maintained. We approached these questions in a twofold way. On the one hand, we varied the distance over which syntactic information had to be maintained in memory by using a purely syntactic manipulation internal to the sentences. Furthermore, we examined the interaction of individual differences in working memory capacity with these sentence-internal working memory processes by separately analyzing ERPs recorded from individuals with low and high working memory capacity.

Processing syntactically complex sentences

Early on, it was reported that sentences with object relative clauses are more difficult to process than sentences containing an embedded subject relative clause (e.g., Fodor, Bever, & Garrett, 1974). This processing difficulty was also demonstrated in a number of studies using online measures (e.g., Ford, 1983; Holmes & O'Regan, 1981; King & Just, 1991; King & Kutas, 1995; Schriefers, Friederici, & Kühn, 1995). Both types of sentences, subject relatives and object relatives, involve WH-movement (cf. examples [2a] and [2b]). However, only in object-initial structures does WH-movement result in a noncanonical word order. The canonical order of verbal arguments is preserved in subject relatives. Although transitive subject and object relatives have often been claimed to differ in their syntactic complexity, from a linguistic perspective they do not differ in the complexity of the phrase structure representation that must be built up (as determined, for example, by the number of non terminal nodes in a phrase structure tree; see, e.g., Fodor et al., 1974, and Yngve, 1960, for discussions of such measures of complexity). However, subject and object relatives differ with respect to the distance between the moved element and its gap as the coindexation in the following examples shows.

[2a] The reporter who_i attacked the senator admitted the error.

[2b] The reporter who_i the senator attacked_i admitted the error.

Using fine-grained behavioral methods (i.e., on-line techniques such as self-paced reading), researchers were able to determine the locus of the processing difficulty for object relative clauses more precisely. Some studies attributed complexity effects to the relative clause region as such (e.g., Wanner & Maratsos, 1978; cf. also Fanselow et al., 1999, for similar evidence from WH-questions). Others demonstrated increased processing load especially at the relative clause verb (i.e., “attacked” in [2b]) and sometimes also at the following main clause verb (Ford, 1983; King & Just, 1991).

Theoretical accounts for the processing difficulty of object-initial structures are heterogeneous. One class of models postulates differences in the demands on working memory elicited by sentences of varying complexity. Wanner and Maratsos (1978), for example, assumed that the head noun of the relative clause (i.e., “the reporter” in [2a] and [2b]) must be maintained in memory over a longer distance in object relatives than in subject relatives. Gibson (1991) proposed that an increased processing load in object relatives is induced by the necessity to maintain in memory more noun phrases (NPs) which have not yet been assigned a thematic role by the verb. Other models propose that processing difficulty in object relatives is mainly driven by more complex computational parsing processes taking place at the relative clause verb. For example, Ford (1983) suggested that reactivation of the head noun at the relative clause verb is more demanding in object relatives because reactivation must occur over a greater distance (compare [2b] to [2a]).

A third type of theory, the Capacity Theory of sentence comprehension (Just & Carpenter, 1992), was stimulated by data from King and Just (1991), who reported that the complexity effect found for object relatives interacted with the participants’ individual working memory capacity. These authors observed that the processing difficulty at the relative clause and matrix clause verbs was greater for individuals with low working memory capacity. From these data, it was concluded that both local computational parsing processes and intermediate memory processes competitively draw upon the same pool of resources (see also Just & Car-

penter, 1992).¹ Further evidence for this view comes from an ERP investigation into the processing of sentences with a local syntactic ambiguity under conditions of concurrent working memory load (Vos et al., 2001a). In this study, it was observed that on-line processing of locally ambiguous relative clauses was modulated by individual differences in working memory capacity and by the external working memory load.

A fourth model, which also explicitly incorporates both working memory demands and computational load as causes of processing difficulty in complex sentences, is the *Syntactic Prediction Locality Theory* (SPLT; Gibson, 1998). This theory differentiates between an “integration cost component,” determining the amount of cognitive resources to be spent for integrating new linguistic input into a mental representation of the sentence, and a “memory cost component.” It postulates that “memory costs” during sentence processing stem from syntactic predictions about the lexical categories minimally required to transform the current input into a grammatical sentence which have to be maintained in working memory (Gibson, 1998). “Integration costs”, on the other hand, arise when new input is integrated into the current structural representation of the sentence. Integration costs are assumed to increase with the distance between the elements to be integrated, as determined by the number of intervening discourse referents (cf. Gibson, 1998).

A number of studies have demonstrated that working memory effects can be studied using ERPs (e.g., Ruchkin, Jormson, Canoune, & Ritter, 1990). More recently, this technique has also been used to investigate working memory during sentence processing. For example, it can be shown that individual working memory capacity (as determined by the reading span task; cf. Daneman & Carpenter, 1980) and concurrent verbal working memory load modulate parsing processes associated with the resolution of syntactic ambiguities in temporarily ambiguous sentences (e.g., Friederici,

¹ Note, however, that these results were criticized by Caplan and Waters (1999, pp. 80–81) because (a) King and Just (1991) did not demonstrate statistically that the complexity effect obtained was indeed localized to the region of the greatest processing load (i.e., no ANOVA with sentence position as repeated measurement factor was reported) and (b) all analyses were based on conditions in which participants had to memorize one or two sentence final words (i.e., load conditions).

Steinhauer, Mecklinger, & Meyer, 1998; Vos et al., 2001a), as well as processes of reanalysis in sentences with syntactic violations (Vos et al., 2001b). These data provide strong evidence for an involvement of working memory resources during parsing. However, they do not pertain directly to the question on which we focused in the present study, namely, how sentence-internal structural configurations induce working memory costs. First, the processes investigated in these studies were elicited mostly by sentences containing either a temporary ambiguity or a syntactic violation. In the present study, however, the focus was on the processing of unambiguous, well-formed sentences. Second, sentence-internal working memory costs were generally investigated by administering a secondary task, i.e., an external working memory load. An involvement of working memory during parsing, thus, was inferred mainly from interactions of established processing effects with the external load factor (cf. also King & Just, 1991). Direct evidence for working memory processes during syntactic analysis of well-formed and unambiguous sentences (i.e., without a secondary task) is only sparse (i.e., King & Kutas, 1995; Kluender et al., 1998) and rests mainly on post hoc explanations based on scalp topography.

The only study that could demonstrate individual differences in effects thought to be related to the maintenance of syntactic information in working memory was an ERP study by King and Kutas (1995). In this study it was shown that performance (i.e., good comprehenders vs poor comprehenders) modulated a sustained left anterior negativity elicited by object relatives. However, this negativity was confounded by differences in word class between the two elements to be compared because it was elicited in the early region of the object relative clause (i.e., at “the senator” in [2b] as compared to “harshly attacked” in the subject relative [2a]). Thus, in our view, it remains to be demonstrated how exactly the working memory mechanisms for storing syntactic information are implemented with respect to temporal organization, functional characteristics, and topographical localization. In the present study, we manipulated sentence-internal working memory costs by varying the distance over which syntactic information had to be maintained in working memory. We thus provide experimental evidence for working memory costs going beyond previous evidence which rested mainly on evidence from individual differences and sentence-external working memory load manipulations.

WH-questions as a test case for investigating working memory during parsing

An influential model of how filler-gap dependencies are processed is the *Active Filler Strategy* (AFS; Clifton & Frazier, 1989; Frazier & Flores D’Arcais, 1989; for alternative accounts, see Crocker, 1994; De Vincenzi, 1996). AFS was introduced to describe how the subject-object ambiguity of the WH-filler in sentences like [2a] and [2b] can be resolved. Thus, AFS does not make predictions regarding the working memory load induced by maintaining the filler in working memory. Nevertheless, it explicitly assumes that working memory processes are necessary to establish the dependency between the dislocated element and its gap, and it further assumes that the WH-filler is actively kept in working memory until the gap is identified.

When processing filler-gap dependencies, the filler can be integrated when the subcategorizing verb allows the projection of the gap. In English, this is reflected in increased processing times at the verb (e.g., Ford, 1983; King & Just, 1991) and in P600 effects in event-related brain potentials at the same location in WH-questions (Kaan, Harris, Gibson, & Holcomb, 2000). Note that this is the position immediately before the linguistically postulated gap. German differs from English in (at least) two respects. First, as verbal arguments in German are often overtly case-marked (e.g., in masculine nominative and accusative relative pronouns or interrogative pronouns), there is often no temporary ambiguity in filler-gap constructions. Therefore, processing such dependencies in German frequently does not require strategies for ambiguity resolution. Second, due to the SOV word order of German, the object gap in object-initial sentences is encountered before the subcategorizing verb. It is not unlikely, thus, that the dependency between the filler and its gap can be established in German before the verb is encountered. For example, the Active Trace Strategy proposed by Crocker (1994) postulates that once the VP can be accessed (e.g., because the processing of the subject NP licenses its projection), the attachment site for the filler becomes available and the filler can be linked to the gap. The notion that these parsing processes can take place even before the verb is processed was empirically supported by a self-paced reading study on NP attachment in German verb-final clauses (Bader & Lasser, 1994), as well as by a reading time study reported by Fanselow et al. (1999). In

the latter study, subject and object WH-questions with overtly case-marked interrogative pronouns were compared (cf. [3a] and [3b]). Although case-marking immediately indicated the grammatical role (i.e., “object”) of the WH-filler in [3b], increased reading times for object-initial structures were obtained between the filler and the second NP.

[3a] Es ist egal, wer vermutlich zufälligerweise den Mann erkannte.

It does not matter, who_{NOM} presumably accidentally the_{ACC} man recognized.

[3b] Es ist egal, wen vermutlich zufälligerweise der Mann erkannte.

It does not matter, who_{ACC} presumably accidentally the_{NOM} man recognized.

These data suggested that the WH-filler cannot be integrated immediately because its attachment site is not available and, therefore, must be maintained in working memory until the gap becomes available. In line with the predictions derived from Crocker’s (1994) Active Trace Strategy, the data further suggest that the parser does not have to delay the establishment of the filler–gap dependency until the verb is encountered because the reading time difference disappears before the verb is encountered. This conclusion is also supported by a recent cross-modal priming experiment reported by Muckel and Pechmann (2000). In this study, antecedent reactivation was observed at a trace position prior to the verb in German ergative sentences.

It should be mentioned here that not all parsing models share the assumption of fillers and gaps. A strong alternative account to this approach is the Direct Association Hypothesis (e.g., Pickering & Barry, 1991), which postulates that when processing unbounded dependencies, a clause-initial object is not linked to its canonical gap site. Rather, it is directly associated with its subcategorizing verb. Thus, this theoretical approach predicts increased processing load due to processes of direct association at the verb. This prediction, however, does not take into account that case-marking information of the argument NPs in German might trigger certain integration processes even before the clause-final verb is encountered (cf. Bader & Lasser, 1994). Accordingly, the present experiment can also be viewed as a test of the predictions made by the Direct Association Hypothesis. In the context of Direct Association, it is not specified whether the clause-initial object must be maintained in working memory or whether it is reactivated at the verb in

order to be able to associate it with the verb. We assume that the notion of maintenance of the filler in working memory in principle is compatible with Direct Association.

The present study

In this study, we investigated ERPs elicited during the processing of indirect German WH-questions with either a subject or an object moved to clause-initial position (cf. examples [4a] and [4b]). Working memory operations of temporarily storing unintegrated structural information between the filler and its gap should be detectable in multiword ERPs (i.e., in ERPs spanning a whole clause; cf. King & Kutas, 1995). In order to explore the difficulty of local integration processes, local ERPs elicited at the individual words or phrases of the sentences were also analyzed.

[4a] Karl fragt sich, wer_i __j den Doktor verständigt hat.

[4b] Karl fragt sich, wen_i der Doktor __j verständigt hat.

In addition to investigating the effects of canonical vs noncanonical word order by contrasting subject and object WH-questions, we included two further tests of the syntactic working memory hypothesis. First, the length of the distance between the filler and its gap was varied by inserting either one or two nonargument prepositional phrases between the WH-filler and the second NP. In terms of Direct Association, this manipulation increased the distance between the clause-initial argument and its subcategorizing verb while under a gap-filling assumption, the distance between filler and gap in the object WH-questions was increased. The filler–gap distance, thus, was determined on the basis of the number of words to be processed before the gap is reached, but not with respect to the structural complexity of the inserted material. If, as we assume, the distance between the filler and its gap is bridged by maintaining the filler in working memory, the length of the filler–gap distance should influence the amount of resources that must be spent in order to successfully link the filler with its gap. Second, the influence of individual differences in working memory capacity on the parsing of these WH-questions was investigated.

In addition to subject and object WH-questions, indirect questions with a subordinated “WHETHER”-clause were also investigated.

These differ from WH-questions in that they are not derived by WH-movement and, therefore, do not involve filler–gap dependencies. As the clause-initial element “whether” does not carry syntactic information indicating that it must be linked to a gap later on, no syntactic working memory load should be induced. On the other hand, it is unlikely that there is no working memory cost elicited at all when processing these structures. For example, from SPLT (Gibson, 1998) it can be derived that the parser predicts a subject and a verb when encountering a clause-initial “whether.” Thus, these sentences should cause greater working memory costs than subject-initial WH-questions. A complete set of stimulus sentences is displayed in Table 1.

On the basis of the theoretical assumptions outlined in the previous sections, we derived the following predictions:

1. Object WH-questions should show a reflection of the greater working memory demands induced by maintaining the dislocated object NP activated until it can be linked to its gap. In analogy to other ERP investigations of slow potentials elicited by syntactic complexity manipulations (King & Kutas, 1995; Kluender et al., 1998) or retention processes during verbal working memory tasks (e.g., Ruchkin et al., 1990), we expected a frontally distributed negativity in the

multiword ERPs between the filler and the gap for object WH-questions, relative to subject WH-questions.

2. We predicted a length-dependency of memory-related ERP effects. Working memory effects should be especially pronounced if the filler and the gap are separated by a longer distance because it is more costly to maintain syntactic information activated in working memory for a longer duration. If one assumes that information in working memory is subject to decay over time, more and more resources are required to keep the information available. Thus, if a sustained ERP effect is indeed related to the amount of working memory resources required for keeping information available, this effect should increase in amplitude over the course of the sentence.

3. Differential effects in the slow waves were predicted for individuals with high working memory capacity as opposed to those with low capacity. It is commonly held that low capacity readers have fewer resources available (cf. Just & Carpenter, 1992). Therefore, these individuals must use a greater portion of the available resources for successfully maintaining the filler in working memory. This prediction is based on several previous studies which have demonstrated interactions between individual working memory capacity and sentence processing (e.g., King &

Table 1
Sample set of sentence material with word-for-word translations to English

<i>Sentence type</i>
<i>Short subject WH-question</i> Thomas fragt sich, wer am Dienstag den Doktor verständigt hat. Thomas asks himself, who _(NOM) on Tuesday the _(ACC) doctor called has.
<i>Short object WH-question</i> Thomas fragt sich, wen am Dienstag der Doktor verständigt hat. Thomas asks himself, who _(ACC) on Tuesday the _(NOM) doctor called has.
<i>Short WHETHER-question</i> Thomas fragt sich, ob am Dienstag der Doktor geweint hat. Thomas asks himself, whether on Tuesday the _(NOM) doctor cried has.
<i>Long subject WH-question</i> Thomas fragt sich, wer am Dienstag nachmittag nach dem Unfall den Doktor verständigt hat. Thomas asks himself, who _(ACC) on Tuesday afternoon after the accident the _(ACC) doctor called has.
<i>Long object WH-question</i> Thomas fragt sich, wen am Dienstag nachmittag nach dem Unfall der Doktor verständigt hat. Thomas asks himself, who _(ACC) on Tuesday afternoon after the accident the _(NOM) doctor called has.
<i>Long WHETHER-question</i> Thomas fragt sich, ob am Dienstag nachmittag nach dem Unfall der Doktor geweint hat. Thomas asks himself, whether on Tuesday afternoon after the accident the _(NOM) doctor cried has.

Just, 1991; Friederici et al., 1998; Münte, Schiltz, & Kutas, 1998; Vos et al., 2001a,b).

4. In addition to the slow potential effects predicted for multiword ERPs, we expected to see a reflection of the syntactic integration of the moved constituent into the phrase structure at the point where the dependency between the filler and the corresponding gap can be established. An ERP component that reflects increased integration difficulty is the P600 (Kaan et al., 2000). Apart from its function as an indicator of reanalysis processes in syntactic violations (e.g., Friederici, Pfeifer, & Hahne, 1993) and garden-path sentences (Osterhout & Holcomb, 1992), it has recently been proposed as a marker of the difficulty of syntactic integration processes and as reflecting the amount of resources required for such processes (Kaan et al., 2000). Accordingly, we expected a positivity in local ERPs for object as compared to subject WH-questions at the point in the sentence where the information stored in working memory can be integrated into the phrase structure. This position in the sentence should furthermore be characterized by the disappearance of sustained ERP effects reflecting maintenance in working memory. As already discussed in the previous section, in German this can take place either at the noun phrase preceding the gap or at the verb following the gap. A Direct Association approach (cf. Pickering & Barry, 1991) would predict increased integration costs specifically at the verb.

Methods

Materials

Forty sets of six indirect questions like those displayed in Table 1 were constructed. All sentences were made up (a) of a matrix clause that always consisted of a first name and the verbal phrase “fragt sich” (i.e., “asks himself”) and (b) of a subordinate question. The four embedded WH-questions contained a subject, an object, and the past participial form of a transitive verb followed by the auxiliary verb “hat.” Either subject or object was moved to the clause-initial position in the form of the masculine nominative or accusative interrogative pronoun (i.e., the WH-filler). The distance between the WH-filler and the second noun phrase (second NP) was varied by inserting either one (short sentence conditions; i.e., “am Dienstag/on Tuesday”) or two (long sentence

conditions; i.e., “am Dienstag nachmittag | nach dem Unfall/on Tuesday afternoon | after the accident”) prepositional phrases (PPs) after the WH-filler. WH-questions always followed the schema [[Matrix clause], [[WH-filler] [1 vs 2 PPs] [2nd NP] [Verb] [hat/AUX]]].

In order to keep the WHETHER-items identical in length to the WH-questions, they were constructed with only one noun phrase, i.e., the subject, and an intransitive verb. WHETHER-questions were also included in versions with one and two prepositional phrases, respectively. The complete set of material used in this experiment is available at www.ideallibrary.com.

For the behavioral comprehension task, probe assertions restating facts of the corresponding items were constructed (e.g., “The doctor was called after the accident.” or “The doctor called somebody on Tuesday.”). Critical information from these assertions, like prepositional phrases, the noun phrase, the verb, or thematic role assignment, was exchanged in half of the assertions in order to form incorrect probes.

Experimental procedure

Before the experiment, a short training block was conducted. Experimental sentences were presented in dark blue letters at the center of a light gray computer screen in a serial phrase-wise mode including punctuation. This mode of presentation is generally used in our lab and in other groups investigating sentence processing in German, mainly because the definite determiners can be misinterpreted as relative pronouns if presented in isolation. The phrase with the greatest number of letters covered 10.5 cm on the screen; viewed at a distance of 90 cm this resulted in a maximum visual angle of 3.34° horizontally. Each sentence presentation consisted of a series of either 7 (short conditions) or 10 (long conditions) frames (i.e., “Karl | fragt sich, wer | am Dienstag | [nachmittag | nach | dem Unfall] | den Doktor | gerufen | hat.”). Frames were presented for either 600 ms for words presented in isolation or 700 ms for noun phrases, prepositional phrases, and for the verb of the embedded question. Each trial was started by the participant. After each sentence, a blank screen was shown for 800 ms, followed by the probe assertion, which was presented in one frame. The participants’ task was to judge whether or not the probe assertion was correct in relation to the critical item. The main purpose of the comprehension task was to

control for correct processing of the stimuli. Participants were instructed to focus on giving correct answers. Speed of response was considered to be secondary as the performance data primarily served to exclude incorrectly answered trials from ERP analyses. Furthermore, participants were asked to blink only while answering the comprehension task and before starting the next trial.

In addition to the 240 experimental items, another 320 sentences (complex conjuncts of two main clauses and one subclause) were included as fillers. The experiment was run in two sessions of seven blocks with 40 items each. Block lengths varied but were no longer than 10 min. The sequence of items and fillers was pseudo-randomized.

Participants

Twenty-two paid volunteers (12 females; right-handed native speakers with normal or corrected to normal vision; mean age 23.9 years; 20 to 28 years) participated in the experiment. All participants were native speakers of German and right-handed as determined using the Edinburgh handedness inventory (Oldfield, 1971). On the basis of the reading span task (Daneman & Carpenter, 1980), a test that is commonly used in psycholinguistic studies to investigate individual differences in working memory capacity (e.g., King & Just, 1991; Just & Carpenter, 1992; Münte et al., 1998; Vos et al., 2001a,b), participants were selected such that they formed one group with a relatively high working memory capacity (i.e., greater than 3.5 on a scale from 2 to 6) and one with a relatively low capacity (i.e., below 3.5). The value of 3.5 was chosen as a cutoff because in our lab, this reading span score was observed to be the most frequent and thus was taken to represent an average working memory capacity. Ten participants were assigned to the high-span group (mean reading span = 4.78; $SD = 0.8$) and nine participants to the low-span group (mean reading span = 2.7; $SD = 0.26$). Three participants with reading span scores of 3.5 were not assigned to one of the two groups.

Recording procedures

A continuous electroencephalogram was recorded from 51 AgAgCl-type scalp electrodes selected from the extended 10–20 system (cf. Sharbrough et al., 1991) which were referenced to

the left mastoid. For detection of eye movements and blink artifacts, a bipolar electrooculogram (EOG) was recorded from two electrodes placed at the outer canthi of the left and the right eye and from two electrodes placed above and below the right eye. Impedances of all electrodes were kept below 2 k Ω . Signals were amplified with a lowpass filter of 30 Hz and sampled at a frequency of 250 Hz.

Data analysis

Trials with EOG artifacts were identified by automatic and manual screening. The mean number of trials with artifacts was 2.8% in local ERPs (i.e., ERPs for single words or phrases) and 7.47 and 9.08% in multiword ERPs elicited by short and long sentences, respectively. As all EEG data sets showed a slow shift to the negative, a detrending algorithm was used to correct for a common linear component. In time windows of 20 s, a linear trend was estimated by minimizing absolute errors and, subsequently, subtracted from the original signal. As Footnote 2 under Results shows, detrending did not affect relative differences between the conditions. After preprocessing, local average ERPs at specific words or phrases were calculated with time windows of 0 to 1000 ms relative to word onset. Multiword average ERPs starting with the question words and spanning the complete embedded question (i.e., 0 to 3700 ms in short questions and 0 to 5600 ms in long questions) were calculated. For all ERP averages, a baseline of 200 ms prior to the onset of the embedded question or, for local averages, prior to word onset, was used. Trials containing artifacts in these time windows and incorrectly answered trials were excluded from averaging. The minimum number of averaged trials per participant in multiword ERPs of long questions, in which the greatest number of artifact-contaminated trials had to be rejected, was 22 in one condition; however, the mean number of averaged trials in long sentences was 31.1 for subject WH-questions, 30.1 for object questions and 31.9 for WHETHER-questions. Averages were re-referenced to the linked mastoids for display and further analyses. For purposes of visualization, ERP waves were smoothed using a low-pass filter with a cutoff frequency of 7 Hz.

For statistical analyses, four regions of interest (i.e., ROIs) encompassing five lateral electrodes each (left anterior AF3, F3, F5, FC3, FC5; right anterior AF4, F4, F6, FC4, FC6; left posterior

CP3, CP5, P3, P5, PO3; right posterior CP4, CP6, P4, P6, PO4) were introduced into analyses of variance (ANOVAs) with the factors “hemisphere” (left vs right) and “scalp extension” (anterior vs posterior). Effects with central distribution were analyzed on midline electrodes with four ROIs (FPZ/AFZ; FZ/FCZ; CZ/CPZ; PZ/POZ) using “scalp extension” as a factor. All measured electrodes were used for displaying the scalp distributions of ERP effects with interpolated topographic potential maps (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987).

ANOVAs were calculated with the within-subjects factors “sentence length” (short vs long) and “sentence type” (subject-WH vs object-WH vs WHETHER), as well as with topographical factors as described above. Time windows for statistical analyses will be reported with the individual analyses under Results. Generally, all six conditions were introduced into the statistical analyses. If ANOVAs were performed on a selection of experimental conditions, this is indicated under Results. Corrected p values (Greenhouse & Geisser, 1959) will be reported automatically together with original degrees of freedom in the case of effects with more than one degree of freedom in the numerator. Reliable effects in behavioral or ERP data were analyzed in a second analysis with the additional between-subjects factor “reading span group.” If main effects of the three-staged factor “sentence type” became significant, planned comparisons were calculated with an adjusted significance threshold set to $\alpha_{\text{planned}} = .033$ (determined on the basis of a modified Bonferroni correction; e.g., Keppel, 1991). For the interpretation of the polarity of reliable ERP effects in the WH-questions, the subject WH-condition was used as baseline relative to which the polarity of reliable effects elicited by object questions was determined.

ANOVAs for the behavioral data were performed analogously, using the factors “sentence length,” “sentence type,” and “reading span group.” Reaction times for correctly answered trials were aggregated within each participant before group averages were calculated. Trials with response times differing by more than 2.5 standard deviations from the overall mean of the respective participant were treated as outliers and not included in the analysis. Error rates were also aggregated within each participant before being used for group statistics. In addition to the classical ANOVAs with participants as random effect ($F1$),

we also report item analyses ($F2$) for behavioral data (cf. Clark, 1973).

Results

Behavioral performance

Answers to long questions were significantly slower than those to short questions [$F1(1, 21) = 13.31, p = .001$; $F2(1, 39) = 36.16, p < .0001$; cf. Table 2 for behavioral data].

Descriptively, response times were longest for WHETHER-questions and fastest for subject WH-questions. The subject analysis did not support this observation statistically. Neither the main effect of sentence type nor the interaction of sentence type and sentence length became significant. In the item analysis, however, there was a reliable main effect of sentence type [$F2(2, 78) = 3.33, p < .05$], which was due to increased response times for WHETHER-questions as opposed to subject questions [$F2(1, 39) = 5.84, p < .033$] and, in tendency, also to object questions [$F2(1, 39) = 3.44, p = .07$]. The number of correct responses was greater in WHETHER- than in WH-questions [$F1(2, 42) = 14.28, p < .0001$; $F2(2, 78) = 36.91, p < .0001$]. A marginally significant interaction between sentence length and sentence type [$F1(2, 42) = 2.58, p = .08$; $F2(2, 78) = 3.02, p = .06$] suggested that only in long WH-questions was there a tendency toward making more errors in object questions than in subject WH-questions [$F1(1, 21) = 4.04, p = .05$; $F2(2, 78) = 13.55, p < .0001$].

Participants with low reading span tended to make more errors than those from the high-span group [$F(1, 17) = 3.65, p = .07$; cf. Table 3]. No interactions involving span group reached significance in the error rates (all $F < 1$). Furthermore, reaction times did not differ reliably between high-span and low-span individuals ($F < 1$).

Table 2

Reaction times, error rates, and corresponding standard errors ($n = 22$)

Sentence type	RT (in ms)	Errors (in %)
Short subject WH	1845.1 (± 68.2)	21.6 (± 2.0)
Short object WH	1896.5 (± 72.1)	21.3 (± 3.0)
Short WHETHER	1900.0 (± 57.6)	9.3 (± 1.7)
Long subject WH	1973.0 (± 94.2)	18.5 (± 1.9)
Long object WH	1980.0 (± 83.1)	23.3 (± 2.7)
Long WHETHER	2010.8 (± 66.5)	12.2 (± 2.4)

Table 3

Reaction times, error rates, and corresponding standard errors for individuals with high and low reading span

Sentence type	Low-span group ($n = 9$)		High-span group ($n = 10$)	
	RT (in ms)	Errors (in %)	RT (in ms)	Errors (in %)
Short subject WH	1861.3 (± 87.2)	25.0 (± 3.2)	1819.5 (± 114.1)	19.3 (± 3.0)
Short subject WH	1955.0 (± 93.3)	26.9 (± 4.4)	1794.5 (± 92.0)	18.8 (± 4.8)
Short WHETHER	1922.3 (± 83.56)	11.7 (± 2.7)	1918.8 (± 88.0)	7.0 (± 1.4)
Long subject WH	2002.1 (± 102.6)	20.8 (± 3.3)	1888.9 (± 114.7)	17.0 (± 2.3)
Long object WH	2026.1 (± 100.0)	29.2 (± 4.4)	1943.7 (± 121.7)	21.5 (± 3.5)
Long WHETHER	2025.9 (± 90.3)	16.1 (± 4.9)	2049.4 (± 90.2)	8.3 (± 2.0)

Event-related potentials

The section reporting electrophysiological results is grouped into three subsections corresponding to the ERPs examined. First, potentials elicited by the question word at the initial position of the subordinate question are reported. Following this, analyses of multiword ERPs starting from the onset of the question word are described for long and short questions separately. In the third subsection, local ERPs elicited at the second noun phrase and at the verb of the embedded questions will be reported.

Local ERPs to question words

No ERP effects of sentence type were present at the question words. Visual inspection revealed that object WH-questions showed a very subtle positivity relative to subject WH-questions between 600 and 750 ms. This positivity was analyzed in ANOVAs with both lateralized and midline ROIs. In both analyses, neither the main effects of sentence type or sentence length nor any interaction of sentence type with topographical factors or sentence length was reliable even in time windows as small as 50 ms (all $F < 1.5$). For this reason, no local ERPs are shown for these words.

Multiword ERPs to embedded questions

Long question conditions. Due to the different number of words of the embedded questions in short and long sentence conditions, multiword ERPs elicited by short and long questions were analyzed separately. Long object WH-questions elicited a sustained negativity relative to long subject questions (cf. Figs. 1A, B, and 2A). Visual inspection of the grand average ERPs revealed that this negativity started at about 400 ms after the onset of the first prepositional phrase (i.e.,

“am Dienstag”) and was present at most electrodes until the last phrase of the PP region (i.e., until “dem Unfall” in the examples given in Table 1; cf. shaded area in Fig. 2A). To analyze this negativity statistically, an average was calculated for the time during which this negativity was clearly observable (i.e., from 1000 to 3400 ms after the onset of the question word). ERPs elicited by WHETHER-questions in the same time window were also negative-going. However, this effect appeared to be much less pronounced (cf. Fig. 1A and C). A laterality analysis using the time window described above showed a significant main effect of sentence type [$F(2, 42) = 4.46, p < .05$]. Planned comparisons revealed that this main effect was due to (a) significantly more negative potentials in the object condition as compared to the subject WH-questions [$F(1, 21) = 6.64, p < .033$],² and (b) the fact that the multiword ERP elicited by the long WHETHER-question was between the two WH-questions, but closer to the object question, thus also differing significantly from subject WH-questions [$F(1, 21) = 5.71, p < .033$].

The topographical potential maps of the differences between object and subject WH-questions and between WHETHER- and subject WH-questions in the selected time window (cf. Fig. 1B and C) reveal that the sustained negativity for object questions was maximal at left-anterior electrode positions while the negativity for WHETHER-questions did not exhibit a clear topographical distribution. This observation was supported by a significant interaction of sentence

² An analysis of variance that compared the difference between the ERPs for long subject and long object WH-questions in the detrended signal with the corresponding difference calculated from the ERPs before detrending showed that there were no reliable differences between the effects due to detrending (all $F < 1.2$).

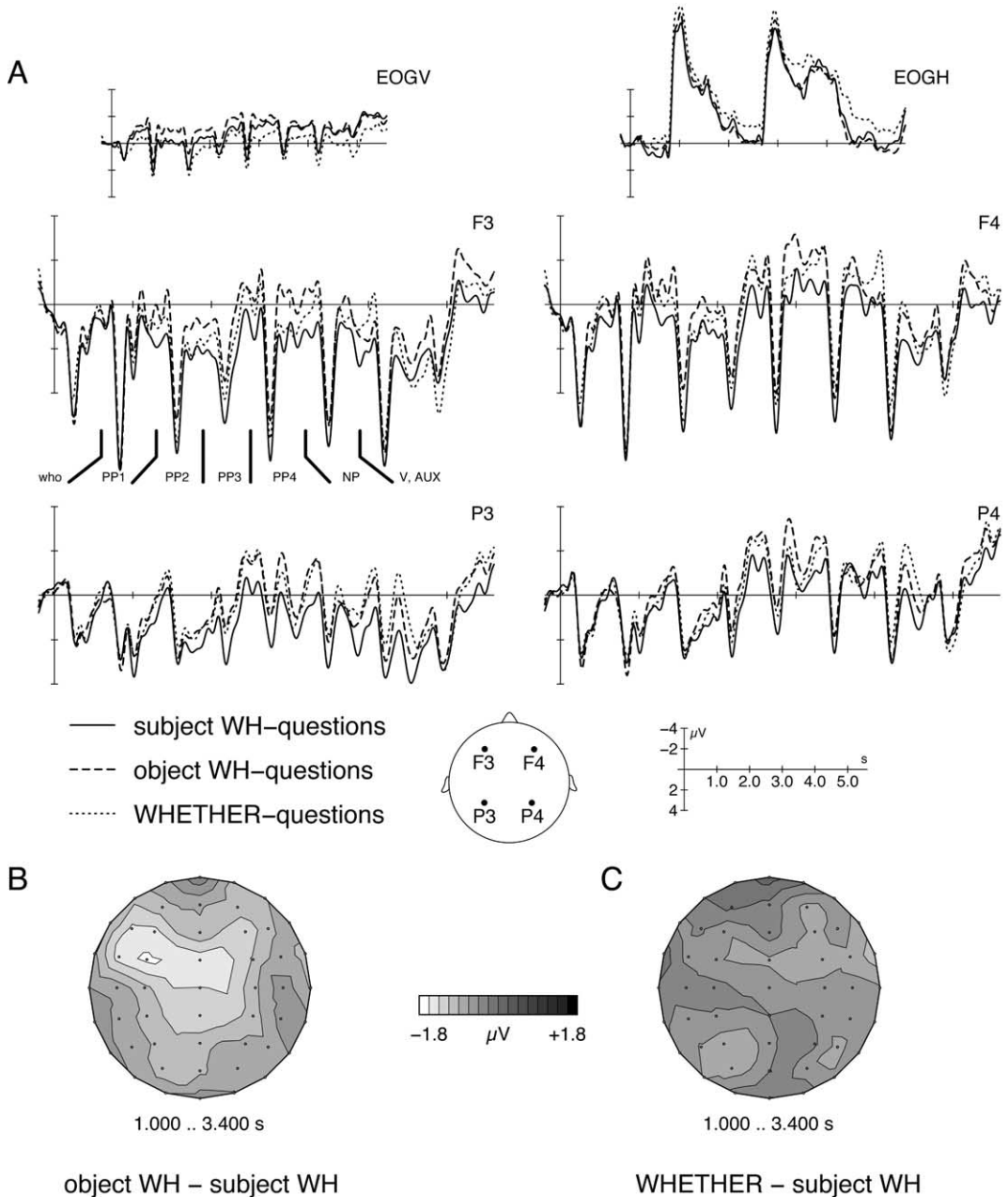


Fig. 1. Multiword ERPs averaged from the onset of the WH-filler for long subject and long object WH-questions, as well as for long WHETHER-questions ($n = 22$). Four selected electrodes representing the lateral regions of interest are displayed, as well as vertical and horizontal EOGs (A). The topographical potential maps display the distribution of the sustained negativity for object as compared to subject WH-questions (B) and for WHETHER-questions as compared to subject WH-questions (C) in the time window used for the statistical analysis (i.e., 1000–3400 ms). Lighter shading indicates more negative potential differences.

type with hemisphere and scalp extension [$F(2, 42) = 3.69, p < .05$]. The interaction was followed up to show a main effect of sentence type

over left-anterior electrode sites [$F(2, 42) = 5.13, p < .05$] that was due to a reliable difference between subject and object WH-questions as re-

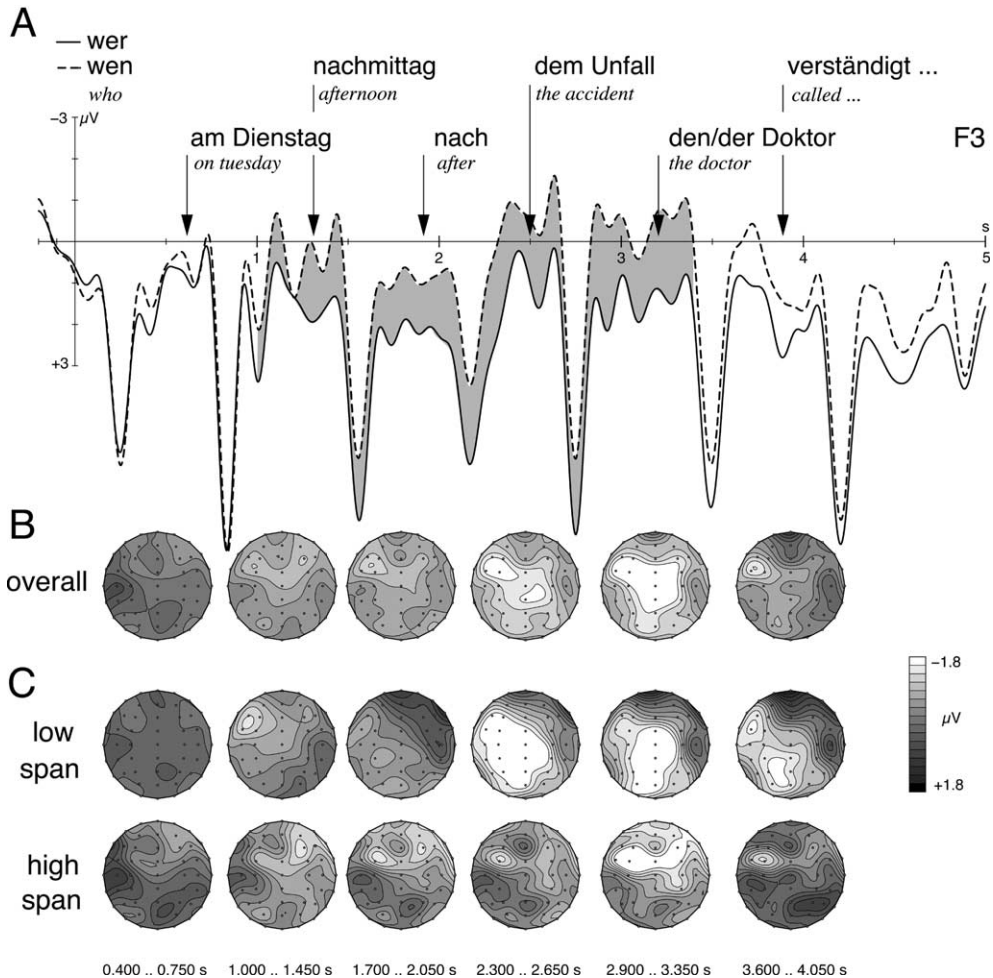


Fig. 2. (A) Multiword ERPs for long subject (solid line) and long object WH-questions (dashed line) from one left-anterior ERP channel (F3). The shaded area indicates the time window used for statistical analyses. Topographical potential maps display the time course and distribution of the sustained negativity for object WH-questions as compared to subject WH-questions for the whole group (B; $n = 22$), as well as for the two reading span groups (C) with low working memory capacity ($n = 9$) and with high working memory capacity ($n = 10$). Lighter shading indicates more negative potential differences.

vealed by planned comparisons [$F(1, 21) = 7.1$, $p < .033$]. No comparable effects of sentence type were found for the three other ROIs. In the left-anterior ROI, ERPs to WHETHER-questions did not differ from ERPs elicited by subject WH-questions [$F(1, 21) = 2.1$, $p > .15$].

In order to describe the time course of the sustained left-anterior negativity for long object WH-questions more precisely, a more detailed ANOVA was calculated using only the two WH-conditions. Mean potentials for six time windows, i.e., from the WH-filler to the second NP, were extracted from the multiword ERP and analyzed

separately (cf. Fig. 2B). The time windows started at 400 ms after the onset of the phrase or word and lasted until the onset of the next P200 component (i.e., for either 350 or 450 ms, depending on the length of the presentation of the respective word or phrase). Table 4 displays the time windows and the respective phrases covered. From Fig. 2A and B, it is evident that the sustained negativity became more and more pronounced and more and more broadly distributed until the last phrase of the PP region was processed. At the second noun phrase, however, the negativity abruptly changed its topography and only a small

Table 4

Time window analysis of the time course of the sustained negativity between long object and long subject WH-questions

	wer/wen who (ms)	am Dienstag on Tuesday (ms)	nachmittag afternoon (ms)	nach after (ms)	dem Unfall the accident (ms)	den/der Doktor the doctor (ms)
milliseconds	400–750	1000–1450	1700–2050	2300–2650	2900–3350	3600–4050
<i>Main effect</i>	—	*	—	**	***	—
<i>Left-anterior ROI</i>	—	*	*	**	***	—
<i>Right-anterior ROI</i>	—	—	—	—	**	—
<i>Left-posterior ROI</i>	—	—	—	**	**	—
<i>Right posterior ROI</i>	—	—	—	**	***	—

Note. The first row displays the statistical results for the main effect of sentence type. The following rows show the effects of sentence type in the different regions of interest. —, not significant (with $F < 1.5$), * $p < .05$, ** $p < .01$, *** $p < .001$.

focus over left-anterior electrodes remained. This time course was confirmed by a significant interaction of the factors time window and sentence type [$F(5, 105) = 4.00$, $p < .005$]. The first row of Table 4 contains results for the main effect of sentence type, analyzed in each time window separately. The increasing strength of the negativity is clearly reflected in these statistics. Furthermore, the factor time window also entered a four-way interaction with hemisphere, scalp extension, and sentence type [$F(5, 105) = 2.95$, $p < .05$]. This interaction appears to indicate that the continuous increase in amplitude until the second NP was present most clearly over left-anterior electrodes. However, Fig. 2B also shows that the topographical distribution became relatively broad in some time windows. This finding is reflected in the time window analyses for the other three ROIs displayed in Table 4, which shows that at the point where the negativity was strongest, this effect was also reliable in other ROIs.

In order to investigate to what degree the sustained negativity in long WH-questions was dependent upon individual working memory capacity, we repeated the above analyses with the additional between-subjects factor span group. Scalp topography maps for the two groups are shown in Fig. 2C. They demonstrate that the sustained negativity was stronger, more broadly distributed, and present earlier for individuals with low working memory capacity. ERPs to WHETHER-questions did not differ between the two reading span groups. In the ANOVA involving the long time window (i.e., 1000 to 3400 ms), two interactions of sentence type and span group with topographical factors were observed. First, a reliable three-way interaction with hemisphere [$F(2, 34) = 3.79$, $p < .05$] was resolved to show a significant interaction of hemisphere

with sentence type in the low-span group [$F(2, 16) = 4.72$, $p < .05$] but not in the high-span group ($F < 1$). This result was due to a lateralization of the sentence type effect to the left hemisphere [$F(2, 16) = 4.76$, $p < .05$; $F < 1$ for the right hemisphere] in the low-span group only (cf. Fig. 2C). Planned comparisons revealed that ERPs to object WH-questions were significantly more negative than ERPs to subject WH-questions in this group [$F(1, 8) = 6.53$, $p < .033$] while there was no difference between WHETHER-questions and subject WH-questions [$F < 1$]. The second interaction involved the factors scalp extension, sentence type, and span group [$F(2, 34) = 4.17$, $p < .05$]. This effect was caused by a marginally significant scalp extension by sentence type interaction [$F(2, 18) = 3.14$, $p = .07$] in the high-span group. Although Fig. 2C reveals that the negativity for long-object as compared to subject WH-questions manifested itself primarily over anterior electrode sites in the high-span group, this interaction could not be further resolved. Statistically, the time course of the sustained negativity did not differ between the two groups.

Short question conditions. Short question conditions were analyzed in a time window that was defined in analogy to that used for long questions. It lasted from 400 ms after the onset of the prepositional phrase to 900 ms after the onset of the prepositional phrase, resulting in a window of 1000 to 1500 ms relative to the onset of the question word. No significant main effect of sentence type or interactions involving this factor were observed in the specified time window (all $F < 2$; cf. Fig. 3).

However, visual inspections of the ERPs to short WH-questions suggested a small and transient negativity for short object questions between

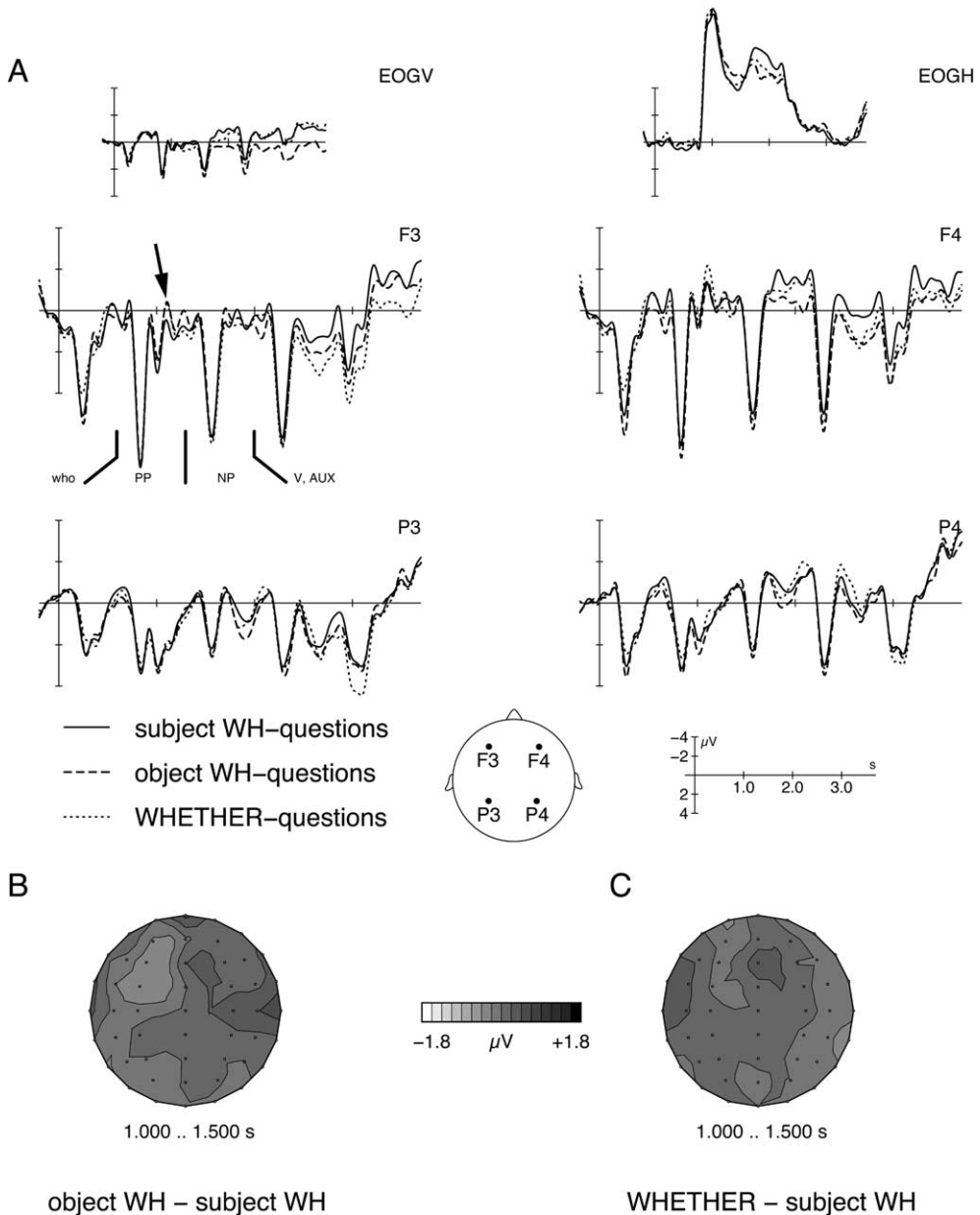


Fig. 3. Multiword ERPs averaged from the onset of the WH-filler for short subject and short object WH-questions, as well as for short WHETHER-questions ($n = 22$). Four selected electrodes representing the lateral regions of interest are displayed, as well as vertical and horizontal EOGs (A). The arrow indicates a transient negativity for object WH-questions. The topographical potential maps display the distribution of the sustained ERPs for object as compared to subject WH-questions (B) and for WHETHER-questions as compared to subject WH-questions (C) in the time window used for the statistical analysis (i.e., 1000–1500 ms). Lighter shading indicates more negative potential differences.

the WH-filler and the gap, namely, at the prepositional phrase (as indicated by an arrow in Fig. 3). As we did not want to preclude the possibility of detecting more transient effects by using long

time windows, an additional ANOVA with a smaller time window was calculated (i.e., 700 to 800 ms after the onset of the PP). At this position in the sentence, participants could not be aware of

the length of the filler–gap distance of the current sentence as the first PP was identical in all six conditions. Furthermore, the same negative peak was also present in ERPs elicited by long object WH-questions (cf. Figs. 1A and 2A). For this reason, short and long questions were both entered into the analysis. As expected, there was no main effect of sentence length ($F < 1$). However, a marginally significant main effect of sentence type [$F(2, 42) = 2.66$, $p < .1$] could be resolved to demonstrate a negativity for object WH-questions, as compared to subject WH-questions. This effect, however, did not reach the significance threshold required for planned comparisons [$F(1, 21) = 4.22$, $p = .05$].

Local ERPs at the second noun phrase and the verb

At the second noun phrase position, object WH-questions elicited more positive-going ERPs than subject questions (cf. Fig. 3). This effect was examined with local ERPs averaged from the onset of the noun phrase. At this position, a positivity in the time window between 400 and 700 ms was observed for object as opposed to subject WH-questions for both short and long

questions (cf. Fig. 4A and B). This positive-going ERP effect peaked between 500 and 550 ms and was maximal over central and parietal midline electrodes. As can also be seen in Fig. 4, ERPs elicited by WHETHER-questions were also slightly more positive-going than those elicited by subject questions. An ANOVA using the four midline ROIs in the time window of 400 to 700 ms yielded a reliable main effect of sentence type [$F(2, 42) = 4.33$, $p < .05$] with a significant difference, according to planned comparisons, between subject and object WH-questions [$F(1, 21) = 6.39$, $p < .033$]. WHETHER-questions did not differ from subject WH-questions ($F < 1$) and tended to be less positive-going than object WH-questions [$F(1, 21) = 3.76$, $p = .07$]. The amplitude of the positivity did not differ statistically between the four ROIs (all interactions involving scalp extension were $F < 2$).

The interaction between sentence type and sentence length was not significant ($F < .5$), thus supporting the observation that the relative positivity observed for object WH-questions was present in both short (Fig. 4A) and long (Fig. 4B) WH-questions and did not differ reliably in strength. A main effect of sentence length

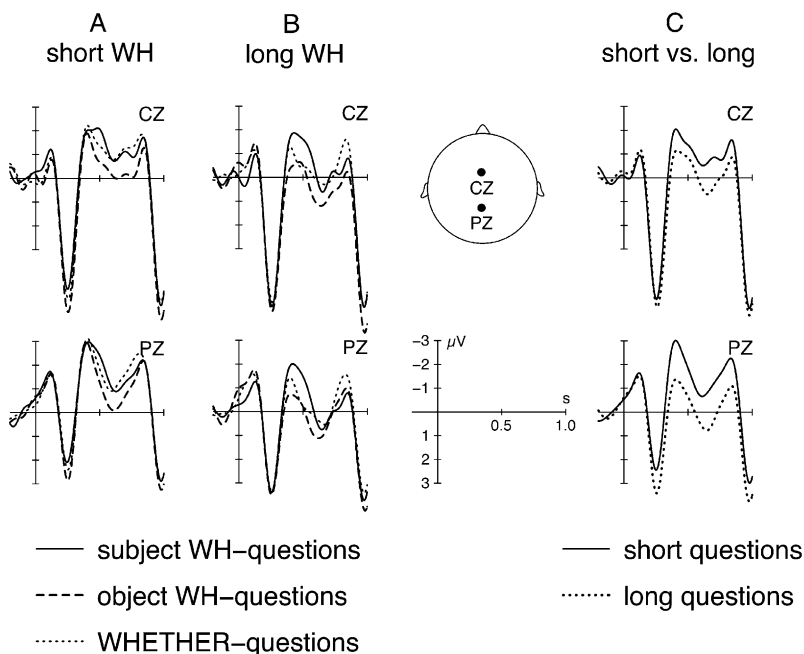


Fig. 4. Local ERPs elicited by subject and object WH-questions, as well as WHETHER-questions at the second noun phrase for short sentences (A) and for long sentence conditions (B) at electrodes CZ and PZ. (C) ERPs for short and long questions, averaged across sentence types ($n = 22$).

[$F(1, 21) = 8.17, p < .01$], however, reflects the finding that across all three sentence types, the absolute amplitude of ERPs elicited by long questions was more positive-going than the amplitude of ERPs elicited by short questions in the time window of 400–700 ms (cf. Fig. 4C). This finding might be due to the presence of a pronounced negativity on the element prior to the second NP, which was used as baseline. However, a positivity of this kind was not induced at the other sentences positions at which a sustained negativity was present in the baseline. At the words and phrases constituting the prolonged PP region in long question conditions (excluding the first phrase which was already reported above), there were no ERP differences in local ERPs (all $F < 1.5$) although there was a pronounced sustained negativity present in this region. There were no differences between individuals with high and low reading span with respect to the positivity elicited at the second noun phrase (all interactions involving span group $F < 1.5$).

Contrary to ERPs elicited at the second NP, there was no positivity for object or WHETHER-questions in the corresponding time window at the verb (cf. Fig. 5). At this word, there were no main effects of sentence length and sentence type (both $F < .5$). However, there was a reliable interaction between scalp extension (i.e., anterior–posterior)

and sentence type [$F(6, 126) = 9.3, p < .001$]. This interaction could be resolved to show a significant sentence type effect only in the most posterior region of interest [i.e., at electrodes PZ and POZ; $F(2, 42) = 6.97, p < .005$; all other ROIs $F < 2$]. The sentence type effect in the posterior ROI was due mainly to a negativity for WHETHER-questions as compared to subject WH-questions [$F(1, 21) = 9.54, p < .01$; mean voltage difference = 1.24 μV]. There was also a posterior negativity for object WH-questions, which, however, was smaller in mean amplitude difference (i.e., 0.62 μV) and did not reach the significance threshold used for planned comparisons [$F(1, 21) = 4.49, p = .046$].

Discussion

In the present study, ERPs were recorded while participants processed well-formed and unambiguous German subject and object WH-questions with either a short or a long distance between the WH-filler and its gap. Furthermore, WHETHER-questions were processed that did not contain filler–gap dependencies. To summarize, we observed the following pattern of effects. While local ERPs to the question words did not differ between sentence types, WH-movement of the object NP into the clause-initial position induced a sustained left-anterior negativity as compared to WH-movement of the subject in long WH-questions. Short object WH-questions elicited a transient negativity that lasted for about 100 ms. Accordingly, prediction 1 could be confirmed. The sustained negativity for long object questions started at the first prepositional phrase in the sentence, i.e., at the position after the WH-filler, and lasted until the subject noun phrase was encountered. In support of prediction 2, it was observed that the amplitude of the negativity increased with increasing distance from the filler. Furthermore, as expected in prediction 3, the negativity was modulated by individual working memory capacity. It was stronger and more broadly distributed in individuals with low working memory capacity than in high-span readers. Although the WHETHER-questions also elicited a negativity relative to subject WH-questions, this effect was small in amplitude and did not show a clear topographical pattern.

In line with prediction 4, a late positivity for object WH-questions was obtained in local ERPs at the second noun phrase position between 400

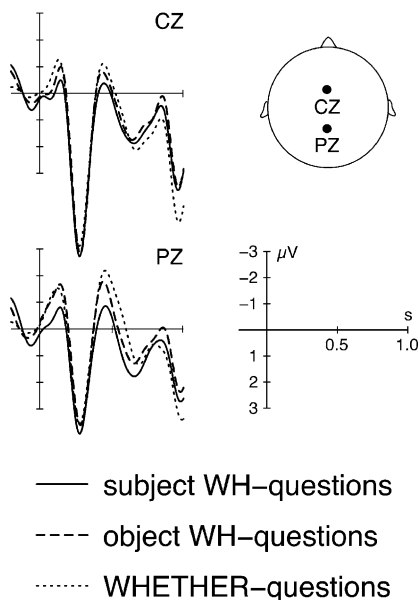


Fig. 5. Local ERPs elicited at the verb for electrodes CZ and PZ ($n = 22$).

and 700 ms, independent of the distance between the filler and the gap. Generally, the positivity at the second NP was greater in amplitude for long questions than for short questions when ERPs were averaged across sentence types. No local ERP effects were present in the region where the sustained negativity was observed. At the verb, a negative ERP effect was observed for WHETHER-questions and, in tendency, also for object WH-questions.

The ERP effects obtained for WH-questions were in line with the behavioral data which demonstrated that slightly more errors were made for long object WH-questions than for long subject questions and that individuals with low working memory capacity tended to make more errors than those with high capacity. In the following, ERP results for different regions of the sentences will be discussed separately and implications for theoretical models will be discussed.

Question words

The fact that there were no differences between ERPs elicited by subject and object WH-fillers at the position of the question word is in line with previous behavioral data. For example, corresponding results have been obtained in the self-paced reading study described above (Fanselow et al., 1999). In this study, reading time increases for object WH-pronouns were only marginally significant. There were also no ERP differences between WH-questions and WHETHER-questions, although these conditions differed at this position with respect to the presence or absence of an argument. Thus, it appears that neither the first steps of structure building in unambiguous WH-questions nor the fact that there is no argument at the clause-initial position in WHETHER-questions differentially affected processing costs reflected in ERPs.

Note that this finding differs from results obtained for case-marked relative pronouns. In relative clauses, the object relative pronoun was shown to elicit more positive-going ERPs than the subject relative pronoun (Friederici et al., 1998). However, our results are in line with previous ERP data from grammatically correct questions containing moved and unmoved complementizers (e.g., Kaan et al., 2000; Kluender & Kutas, 1993a). In these studies, no effects were reported for the question words, but only for the words following the complementizers or WH-fillers (Kluender & Kutas, 1993a) or on the verb of the

embedded clause (Kaan et al., 2000). However, there has been one report of a negativity for WH-fillers as opposed to nonmoved complementizers (Kluender & Kutas, 1993b). This effect was interpreted as a modulation in the amplitude of the N400 component due to lexical semantic differences between the conditions (Kluender & Kutas, 1993b).

Filler-gap dependency

The finding of a sustained left-anterior negativity is in line with previous data (e.g., King & Kutas, 1995; Kluender et al., 1998). However, the present result that this negativity is dependent upon the length of the filler-gap distance is an important novel finding which provides strong evidence for the assumption that sentence-internal structural dependencies are indeed costly in terms of working memory resources. Although this notion has been proposed previously, the length manipulation included in the present study provides the most direct evidence so far in support of this assumption, strengthened by the finding of an interaction with individual working memory capacity. Thus, our results are in agreement with parsing strategies that assume maintenance of the filler in working memory (such as, e.g., the Active Filler Strategy; Frazier & Flores D'Arcais, 1989). The observation that the sustained left-anterior negativity interacted with individual working memory capacity is compatible with the view that the cognitive resources which are required for maintaining syntactic information available in working memory are limited in capacity (Just & Carpenter, 1992).

This conclusion is further supported by the finding that participants with low working memory capacity performed worse than participants with a high working memory capacity. As the ERPs reported are based on correctly answered trials, it can be concluded that for the successful processing of filler-gap dependencies, individuals with low capacity must invest relatively more working memory resources than individuals with high capacity. As these working memory processes compete with other processes for a more limited total amount of resources in individuals with a low working memory capacity, these participants probably reached a critical limitation of resource usage more often and therefore made more errors than high-span readers.

In short-object WH-questions, on the other hand, the filler-gap distance was so short that

there was no need to spend a comparable amount of resources in order to make the linking of filler and gap possible. Although there was a transient negativity elicited in short object questions, there was no need to maintain the WH-filler in working memory over a longer duration. Transient demands on working memory during parsing, thus, can be coped with rather effortlessly while maintenance over longer regions of a sentence strongly increases working memory costs. The appearance of a local left-anterior negativity at the word following the object WH-filler in short WH-questions is consistent with previous data from English WH-questions, which showed similar effects relative to questions without WH-movement (cf. Kluender & Kutas, 1993a; but see Kaan et al., 2000, and McKinnon & Osterhout, 1996). Our data suggest that the reason that in some studies a sustained negativity was found (e.g., King & Kutas, 1995; Kluender et al., 1998) and in others was not (e.g., Kluender & Kutas, 1993a) might be related to the distance over which syntactic information had to be maintained in memory. However, on the basis of the present results, the question arises whether the sustained negativity is driven by the mere number of words that must be processed until the gap is encountered or, alternatively, by the structural complexity of the intervening material.

Our interpretation of the sustained left-anterior negativity as a reflection of maintenance in working memory is supported also by the topographical similarity of this ERP component to frontal slow wave potentials reported in ERP studies of working memory. For example, Ruchkin et al. (1990) reported a frontal negative ERP wave during the retention interval in a verbal memory task which increased with load. This negativity was also lateralized to the left hemisphere. These authors argued that the frontal negative wave in their study was associated with retention processes rather than acquisition operations because it was absent in control trials that consisted of a search task without retention demands.

The left lateralization of the sustained negativity suggests that at least part of the neural generators of this ERP component might overlap with brain areas underlying local left-anterior negativities (i.e., LAN effects) elicited by syntactic violations (e.g., Coulson, King, & Kutas, 1998; Friederici et al., 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991). However, there are two important points that make it hard to unequiv-

cally interpret the present result as reflecting a genuinely syntactic working memory component. First, local LAN effects generally associated with syntax have a different time course and are elicited by syntactic violations. Kluender et al. (1998) demonstrated that two LAN effects, a local syntactic LAN and a global slow negative potential, can be experimentally dissociated. Thus, local and global left-anterior negativities cannot easily be taken to reflect the same cognitive processes. Second, other working memory processes which are not syntactic in nature also have been shown to elicit left-anterior negativities. For example, Münte et al. (1998) demonstrated that a conceptual-semantic load during sentence processing resulted in a very similar ERP effect. However, it is important to stress that the negativity observed in the present study cannot be explained by assuming differences in the semantic load that had to be maintained in memory. In this respect, our data clearly differ from the Münte et al. (1998) study. The memory load eliciting the sustained negativity in the present study was caused by structural dependencies internal to the sentence, not by variations in the conceptual order of the described events between two clauses. Furthermore, in our study, differences between working memory span groups were in the opposite direction from those reported by Münte et al. (1998). This finding might be due to differences in the content that had to be maintained in working memory (i.e., semantic vs syntactic information).

The working memory effect observed in the present study could be isolated without adding concurrent load to the verbal working memory system while sentence processing was accomplished (as was done in many other studies of working memory and language; e.g., King & Just, 1991; Vos et al., 2001a,b). Instead, all effects reported here were obtained from syntactic manipulations internal to the sentences. Note that the topographical distribution of the present sustained (global) negativity indexing syntactic memory load differed markedly from a negativity induced by a concurrent verbal memory load during sentence processing in the Vos et al. (2001a) study. In the latter study, a transient (local) negativity was reported for the high load condition (i.e., three words as opposed to one word) which was present in both reading span groups, albeit with different scalp distributions: This negativity was broadly distributed for the high-reading-span group while it was maximal at posterior electrodes in the low-span group (Vos

et al., 2001a). In contrast, the sustained negativity observed in the present study revealed a different topography (left anterior) and clear differences in amplitude between high- and low-span readers. Thus, it is not very plausible to assume that the sustained ERP effect observed here reflects classical verbal memory processes, such as for instance, articulatory rehearsal. Instead, we suggest that the working memory processes observed here are internal to the parsing mechanism, as proposed, e.g., by Caplan and Waters (1995, 1999).³

Finally, although the last two paragraphs stressed the left-anterior focus of the sustained negativity, it must be noted that this distribution varied over time and between the two reading span groups. Especially in the low-span group, the negativity was very broadly distributed at some positions in the sentence and was also present over posterior electrodes. Nevertheless, we refrained from discussing these differences in scalp topography as being due to the activation of different cognitive processes (and, thus, as reflecting different ERP components) for several reasons. First, the time course of the sustained negativity does not allow us to interpret the data as being due to a series of local ERP components present at successive words. The negativity is present over the whole prepositional phrase region, as described under Results. Second, examination of local ERPs at the words or phrases presented in the PP region in long WH-questions did not reveal local ERP differences between conditions when baselines were aligned. Our interpretation of the more widespread scalp topography in low-span readers assumes that these readers used more resources for achieving the same performance as high-span readers. The result also suggests that this increased resource usage might be supported by more distributed brain regions. However, due to the relatively poor spatial resolution of ERPs, this proposal cannot be further specified.

³ It is important to note that the sustained negativity was not elicited by the fact that the prepositional phrases encountered between the question word and the noun phrase have been scrambled from within the verb phrase to their actual position on the surface. As we have argued elsewhere (Friederici, Schlesewsky, & Fiebach, in press; cf. also Rösler, Pechmann, Streb, Röder, & Hennighausen, 1998), scrambling operations elicit transient LAN effects on the moved elements, indicating that scrambled phrases induce local syntactic violations. Such effects were not observed in the present experiment.

Second noun phrase

At the second noun phrase position, the sustained negativity was almost not observable any longer. In addition, a broadly distributed positivity was elicited for object as opposed to subject WH-questions at the second NP for both short and long object WH-questions. This ERP effect was similar to the positive ERP components elicited by syntactic anomaly (e.g., Friederici et al., 1993; Osterhout & Holcomb, 1992). In line with a recent interpretation of late positive ERP effects in syntactically well-formed sentences which was proposed by Kaan et al. (2000), we take this positivity to reflect the difficulty of local integration processes associated with this position in the sentence. As the late positivity did not interact with individual working memory capacity, we conclude that the local integration processes indexed by this ERP component are independent of the working memory resources which are required for temporarily storing the WH-filler.

Furthermore, the relative strength of the late positivity at the second noun phrase did not differ between short and long object WH-questions. Thus, it appears that the number of integrations that take place or the difficulty of these computational operations determines the amplitude of the positivity, rather than the distance over which this integration takes place (i.e., its locality; cf. Gibson, 1998). However, the amplitude of ERPs elicited at this position differed between short and long questions, independent of sentence type. The increased positivity for long questions might reflect generally increased processing costs when two verbal arguments are integrated over a greater distance. In fact, the distance between the dislocated argument (i.e., the WH-filler) and the second NP is identical in subject and object WH-questions (cf. Table 1). Thus, despite the independence of the late positivity from the filler-gap distance, there seems to be some support for SPLT's locality assumption in the present data. However, this interpretation must be treated with caution, as it cannot be completely ruled out that the presence of a negativity at the preceding word, from which the baseline was taken, caused this positivity. On the other hand, this possibility is relatively unlikely as local ERPs calculated at the sentence constituents preceding this position suffered from the same problem but did not show local ERP differences. Another possibility is that the linear

position of the second NP in the sentences, which differed between long and short questions, is the cause of this amplitude difference between short and long conditions.

Given that at the following verb, no ERP effect was observed which might have suggested increased syntactic integration costs, we conclude from the present results that the dependency between the filler and its gap can be established incrementally already at the position of the second NP, that is, before the verb is encountered. Thus, our data strongly support the proposal that some syntactic integration processes can be triggered by case-marking information, without necessarily requiring the information provided by the verb (cf. Bader & Lasser, 1994; Crocker, 1994; Muckel & Pechmann, 2000). Furthermore, the present results do not support the strong prediction of the Direct Association Hypothesis that the clause-initial object WH-filler is associated directly with the verb (cf. Pickering and Barry, 1991). At the verb position, no ERP effects which could be interpreted as reflecting direct association were observed. However, the observation of such effects at the second NP position might be consistent with a modified version of the Direct Association Hypothesis, which would allow direct association between two arguments if these are encountered before the verb and, unlike in English, unambiguously case-marked. Under the plausible assumption that the parser can profit from overt case-marking information, the present results could be taken to suggest that the clause-initial object filler can be directly associated with the subject NP and, thereby, would allow one to preliminarily establish thematic relations. At the verb, this preliminary analysis would merely have to be checked.

The finding of differences between WHETHER-questions and WH-questions in an N400-like ERP component at the verb position was not predicted in the present study. One possible explanation for this effect is the lexical difference between the verbs in the two question types. While the verbs in the WH-questions were transitive, those used for constructing WHETHER-questions were not. However, the amplitude difference might also be due to another cause, namely, to the fact that the verbs of the WH-questions were presented twice as often as those of the WHETHER-questions. Thus, the difference in N400 amplitude might be a mere reflection of differences in the frequency of presentation.

Theoretical implications

What implications do the findings of (a) a sustained left-anterior negativity in the region of the filler-gap dependency and (b) a late positivity at the second NP position in German object WH-questions have for models of parsing? First, we can conclude that it is a valid assumption to differentiate between syntactic memory costs and syntactic integration costs as two aspects contributing to processing difficulty in syntactically complex sentences (cf. Gibson, 1998). The high temporal resolution of the ERP method enabled us to tease apart these two independent mechanisms. Second, the basic assumption of active maintenance of the WH-filler until the gap is located, which was proposed in models such as the Active Filler Hypothesis (Clifton & Frazier, 1989; Frazier & Flores D'Arcais, 1989; also see Fanselow et al., 1999), receives strong support from our data. Syntactic memory costs can account for processing difficulty observable over extended regions of sentences. With respect to WH-movement, our data suggest that the mechanism that allows the establishment of the filler-gap dependency during parsing of sequential linguistic input is a working memory mechanism. Importantly, this mechanism is required even when case-marking of the filler overtly indicates that the filler must be linked to the object position of the VP.

The question of what exactly is maintained in syntactic working memory cannot be answered on the basis of the existing data. From models such as the Active Filler Strategy (Clifton & Frazier, 1989; Frazier & Flores D'Arcais, 1989), the assumption can be derived that what is held in memory is the filler itself or some of its syntactic features. Syntactic Prediction Locality Theory (Gibson, 1998), however, posits that on the basis of the available input a prediction is made about what constituents are minimally required to form a grammatical clause. Memory costs in this model are due to the maintenance of a prediction which is more complex for object WH-questions than for subject WH-questions. Both approaches make comparable predictions with respect to working memory load and are equally plausible given the present data.

We cannot make claims as to whether the two aspects of processing difficulty (i.e., integration costs and memory costs) draw on the same pool of resources (as suggested by Just & Carpenter, 1992) or on separate resources dedicated to different aspects of working memory during sentence

processing (as is implied, e.g., by Caplan & Waters, 1999; Waters & Caplan, 1996). However, we notice that the two ERP components indicative of the two aspects of syntactic complexity (i.e., the late positivity at the second NP and the sustained left-anterior negativity) responded differently to the two working memory variables which were manipulated in this study. This dissociation might be taken to suggest that working memory processes during sentence comprehension and syntactic integration processes draw upon different processing resources. As the sustained negativity and the late positivity occurred on-line (i.e., during the processing of the sentences), the processes reflected by these ERP effects both must be attributed to working memory resources dedicated to the assignment of syntactic structure and to its use in deriving sentence meaning from the input (i.e., to “interpretive processing resources” as opposed to “postinterpretive resources” involved in using the prepositional content of sentences to accomplish behavioral tasks; cf. Caplan & Waters, 1999). The present dissociation with respect to individual working memory capacity might suggest a further subdivision within the domain of “interpretive processing resources.” For example, computational processes such as local parsing operations and more sustained storage or maintenance processes might require different cognitive or neural resources.

Conclusion

The establishment of a filler–gap dependency is not a mere representational assumption but a psychologically valid process that unfolds in time. Due to the sequential nature of the linguistic input, syntactic working memory resources play a critical role in the reconstruction of the filler–gap dependency during parsing. In the present experiment, maintenance of the moved WH-filler in syntactic working memory over an increased distance was reflected by a sustained left-anterior negativity. Freeing memory from this load and integrating the dislocated element into the phrase structure representation were indexed by a local ERP component with positive polarity at the second noun phrase. The two mechanisms are independent as they elicited ERP components with different time scales, topographies, and polarities. Furthermore, they were differentially affected by a manipulation of the length of the WH-movement as well as by individual differences

in working memory capacity. These results clearly support the notion of two separable aspects underlying the processing of complex sentences: syntactic memory costs and syntactic integration costs.

Acknowledgments

The authors thank Stefan Frisch and Ina Bornkessel for their support during data acquisition, data analysis, and manuscript preparation. We are also grateful to Colin Phillips and three anonymous reviewers for helpful comments on earlier versions of the manuscript. This research was supported by the Leibniz Science Price awarded to A.D.F. by the German Research Foundation and by Grant FI 848/1 awarded to C.J.F. by the German Research Foundation.

References

- Bader, M., & Lasser, I. (1994). German verb-final clauses and sentence processing: evidence for immediate attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 225–242). Hillsdale, NJ: Erlbaum.
- Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, *12*, 637–649.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 11–126.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.
- Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Dordrecht: Kluwer Academic.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected—event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21–58.
- Crocker, M. (1994). On the nature of the principled-based sentence processor. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 245–265). Hillsdale, NJ: Erlbaum.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

- De Vincenzi, M. (1996). Syntactic analysis in sentence comprehension: effects of dependency types and grammatical constraints. *Journal of Psycholinguistic Research*, 25, 117–133.
- Fanselow, G., Kliegl, R., & Schlesewsky, M. (1999). Processing difficulty and principles of grammar. In S. Kemper & R. Kliegl (Eds.), *Constraints on language* (pp. 171–201). Dordrecht: Kluwer Academic.
- Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language*. New York: McGraw-Hill.
- Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 203–218.
- Frazier, L., & Flores D'Arcais, G. B. (1989). Filler driven parsing: a study of gap filling in Dutch. *Journal of Memory and Language*, 28, 331–344.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition*, 6, 291–325.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic morphological and syntactic violations. *Cognitive Brain Research*, 1, 183–192.
- Friederici, A. D., Schlesewsky, M., & Fiebach, C. J. (in press). WH movement vs. scrambling: the brain makes a difference. In S. Karimi & T. Langendoen (Eds.), *Word order and scrambling*. Cambridge, MA: Blackwell.
- Friederici, A. D., Steinhauer, K., Mecklinger, A., & Meyer, M. (1998). Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses. *Biological Psychology*, 47, 193–221.
- Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistics Research*, 18, 51–60.
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Haegeman, L. (1994). *Introduction to government and binding* (2nd ed.). Oxford, UK: Blackwell.
- Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior*, 20, 417–430.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of jsyntactic integration difficulty. *Language and Cognitive Processes*, 15, 159–201.
- Keppel, G. (1991). *Design and analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580–602.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 1, 376–395.
- Kluender, R., & Kutas, M. (1993a). Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196–214.
- Kluender, R., & Kutas, M. (1993b). Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8, 573–633.
- Kluender, R., Münte, T., Cowles, H. W., Szentkuti, A., Walenski, M., & Wieringa, B. (1998). *Brain potentials to English and German questions*. Poster presented at the Annual Meeting of the Cognitive Neuroscience Society.
- Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.
- McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. *Language and Cognitive Processes*, 11, 495–523.
- Muckel, S., & Pechmann, T. (2000). *Does the parser search for traces?* Poster presented at the 13th CUNY Conference on Human Sentence Processing.
- Münte, T., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, 395, 71–73.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5–19.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785–804.
- Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, 66, 75–81.
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229–259.

- Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a language with varying word order: word-by-word variations of processing demands are revealed by event-related brain potentials. *Journal of Memory and Language*, 38, 150–176.
- Ruchkin, D. S., Jormson, R., Jr., Canoune, H., & Ritter, W. (1990). Short-term memory storage and retention: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology*, 76, 419–439.
- Schriefers, H., Friederici, A. D., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language*, 34, 499–520.
- Sharbrough, F., Chatrian, G.-E., Lesser, R. P., Luders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society. Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8, 200–202.
- Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001b). Working memory constraints on syntactic processing: an electrophysiological investigation. *Psychophysiology*, 38, 41–63.
- Vos, S. H., Gunter, T. C., Schriefers, H., & Friederici, A. D. (2001a). Syntactic parsing and working memory: the effects of syntactic complexity, reading span, and concurrent load. *Language and Cognitive Processes*, 16, 65–103.
- Wanner, E., & Maratsos, M. (1978). An ATM approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.
- Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review*, 103, 761–772.
- Yngve, V. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society*, 104, 444–466.