LANGUAGE AND COGNITIVE PROCESSES, 2001, *16* (1), 35–63

Comprehension of long distance number agreement in probable Alzheimer's disease

Amit Almor* and Maryellen C. MacDonald*
Department of Psychology, University of Southern California, Los Angeles, USA

Daniel Kempler*
Department of Otolaryngology, and School of Gerontology, University of Southern California, Los Angeles, USA

Elaine S. Andersen*
Department of Linguistics, University of Southern California, Los Angeles, USA

Lorraine K. Tyler
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

* Also affiliated with the Program in Neural, Informational, and Behavioral Sciences at the University of Southern California.

Requests for reprints should be addressed to Amit Almor, Hedco Neuroscience Building, University of Southern California, Los Angeles, CA 90089-2520.
Email: almor@gizmo.usc.edu.

This work was supported by a grant from the National Institutes of Health to Maryellen MacDonald (NIH RO1 AG11773). We wish to acknowledge the following for referring Alzheimer patients for this research: Dr Victor Henderson, Professor of Neurology and Director of the Clinical Core of the Alzheimer's Disease Research Center at the University of Southern California (NIH AG05142); Dr Helena Chui, McCarron Professor of Neurology at the University of Southern California and Co-Director, Alzheimer's Disease Diagnostic and Treatment Center, Rancho Los Amigos Medical Center (DHS 94-20356), and Dr Lee Willis, USC Department of Neurology and Geriatric Neurobehavior Center, Rancho Los Amigos Medical Center. We are grateful to Lori Altmann, Mariela Gil, Laila Lalami, Karen Marblestone, Sarah Schuster and Karen Stevens for their efforts in collecting these data, and we would like to thank the participants and their families for taking part in this project.

http://www.tandf.co.uk/journals/pp/01690965.html DOI: 10.1080/01690960042000094

Two cross-modal naming experiments examined the role of working memory in processing sentences and discourses of various lengths. In Experiment 1, 10 memory impaired patients with probable Alzheimer's disease (AD) and 10 healthy elderly control participants showed similar sensitivity to violations of subject-verb number agreement in a short sentence condition and similar degradation to this sensitivity in a long sentence condition. Performance in neither length condition correlated with performance on working memory tasks, suggesting that the processes involved in interpreting a grammatical dependency between adjacent and nonadjacent elements are different from those required in the working memory tasks. In Experiment 2, the same 10 AD patients were less sensitive than the 10 control participants to pronoun-antecedent number agreement violations in a short discourse condition, but neither group was affected by additional length. In this experiment, performance in both the short and long conditions correlated with working memory performance. These results show that grammatical and discourse dependencies pose different memory and processing demands, and that these differences are not simply due to differences in the amount of intervening material between dependent words. The results also suggest that while the working memory deficits characteristic of AD do not interfere with on-line grammatical processing within sentences, they do compromise on-line discourse processing across sentences.

INTRODUCTION

Patients with Alzheimer's disease (AD) have difficulty comprehending spoken language (Kempler, 1995; Tomoeda, Bayles, Boone, & Kaszniak, 1990). This difficulty is apparent in their conversational interaction as well as in their performance in many laboratory tasks (e.g., Almor, Kempler, MacDonald, Andersen, & Tyler, 1999; Emery, 1985; Kempler, Almor, Tyler, Andersen, & MacDonald, 1998b; Rochon, Waters, & Caplan, 1994; Tomoeda et al., 1990). The origins of this difficulty, however, are not well understood. One question that has received much attention in recent years is whether this difficulty is due to a linguistic impairment in either the syntactic or semantic systems, or whether it can be explained on the basis of nonlinguistic impairments, such as a working memory deficit (Just & Carpenter, 1992; Kempler, 1995; Kempler, Almor, & MacDonald, 1998a; Kempler et al., 1998b; Kempler, Andersen, & Henderson, 1995; Kempler, Curtiss, & Jackson, 1987; Waters & Caplan, 1997). Indeed, because language processing is a complex activity that depends on many linguistic and cognitive processes, an interruption to any of these processes could result in a comprehension impairment (Small, Andersen, & Kempler, 1997).

Evidence obtained from sentence-picture matching tasks in which patients have more difficulty understanding syntactically complex vs. simple sentences, has led several researchers to conclude that AD patients' comprehension is affected by a genuine syntactic deficit, which interferes with the initial interpretation of the sentence structure (e.g., Grober & Bang, 1995). Following Rochon et al. (1994) we will refer to this kind of deficit as an ‘‘interpretive’’ deficit. Other researchers have noted, however, that sentence-picture matching tasks not only require successful processing of words and syntactic structure, but also require holding the sentence meaning in mind while scanning the picture array and selecting a matching picture (e.g., Kempler et al., 1998a; Rochon et al., 1994). The observation that these tasks impose significant nonlinguistic (e.g., memory) requirements has led these researchers to question whether syntactic complexity effects in these tasks imply the existence of an interpretive deficit in AD. Instead, these researchers suggest, AD patients may have a memory deficit that does not interfere with initial syntactic or interpretive processing but only affects processing that happens after that (a ‘‘post-interpretive’’ deficit; Rochon et al., 1994). Thus, patients may be able to interpret the sentence meaning adequately as they hear it, but perform poorly on a sentence comprehension task because they cannot hold the meaning of the sentence in mind while performing the task.

In order to better clarify whether there is a genuine syntactic processing deficit in AD, and to better assess the role of working memory in comprehension in general, several researchers have designed and used language processing tasks that minimise nonlinguistic memory and attentional requirements. This research has employed ‘‘on-line’’ methods, such as cross-modal naming (Marslen-Wilson & Tyler, 1980), in which the processing of linguistic material is assessed without requiring any conscious judgements about the sentence structure or meaning (e.g., Almor et al., 1999; Kempler et al., 1998a; Nebes, Boller, & Holland, 1986). In these studies, patients and healthy control adults produce similar response patterns in processing basic grammatical dependencies between adjacent words, such as a verb and its following argument (e.g., Kempler et al., 1998b). These data contrast with patients' impaired performance in off-line tasks and are consistent with the view that comprehension deficits in AD reflect patients' difficulties with the nonlinguistic memory demands imposed by off-line tasks.

This simple account of on- vs. off-line data is complicated by other findings, however. AD patients do show impairments in on-line tasks when the stimuli include discourse dependencies between nonadjacent words, such as a pronoun and its antecedent (Almor et al., 1999). One possible conclusion from this research is that AD patients have a comprehension problem specific to processing discourse dependencies such as pronoun-antecedent relations, but no deficit when interpreting grammatical dependencies such as subject-verb agreement. However, an additional finding of these studies, that the on-line processing of discourse dependencies correlates with working memory performance (Almor et al., 1999) while the on-line processing of grammatical dependencies does not (Kempler et al., 1998a), suggests a different explanation for these data, based on the differing processing demands imposed by the two kinds of dependencies. In speech production, subject-verb number agreement is more tightly controlled and is less subject to semantic information than is antecedent pronoun agreement (Bock, Nicol, & Cutting, 1999). Grammatical dependencies therefore have a high degree of predictability in the sense that people know to expect an agreeing verb after the subject noun and may be quite surprised to encounter a verb that does not agree with the subject. In contrast, discourse dependencies are often optional. That is, it is difficult to predict whether a discourse entity will be referred to again, and disagreeing references can be successfully resolved by assigning them to new referents. Furthermore, grammatical dependencies tend to occur between nearby words whereas discourse dependencies tend to occur over long distances, often crossing phrase and sentence boundaries. Maintaining information that may or may not be important later for processing a long distance dependency (as in the case of discourse but not grammatical dependencies) could be especially taxing for AD patients, who typically have limited working memory capacities. Thus it is not clear whether language comprehension performance is determined by the nature of the dependency probed (i.e., grammatical vs. discourse), the distance between the dependent constituents, or simply the working memory demands of processing these types of linguistic structures.

Resolving these issues is not only important for a precise characterisation of the language comprehension deficits of AD patients, but also has significant implications for theories of working memory and language processing more generally. Just and Carpenter (1992) have argued that there is a single pool of verbal working memory resources from which all linguistic processes draw (see also Carpenter, Miyake, & Just, 1994; Just, Carpenter, & Keller, 1996; Miyake, Carpenter, & Just, 1994). Because both grammatical and discourse dependencies can span considerable stretches of intervening material, and it is the need to simultaneously store and process information which, in this theory, largely determines how much working memory is used, the distance between dependent words should affect working memory utilisation in processing both grammatical and discourse dependencies. Furthermore, increasing distance between dependent words should especially affect processing for individuals with a working memory deficit, such as patients with AD.

Just and Carpenter's single resource theory has been critiqued by Waters and Caplan, who, in a series of papers (e.g., Caplan & Waters, 1999; Rochon et al., 1994; Waters & Caplan, 1996, 1997; Waters, Caplan, & Rochon, 1995), argue that there is not one but at least two resource pools involved in language comprehension. In Waters and Caplan's theory, one pool of working memory resources serves the interpretive processes involved in extracting meaning from linguistic input including syntactic processing, and a separate pool of working memory resources serves the post-interpretive processes that are involved in further acting upon the generated representation (e.g., choosing a picture in a sentence-picture matching task, or answering a question). Accordingly, Waters and Caplan distinguish between effects of syntactic complexity, which they attribute to interpretive processing, and effects of the number of propositions in an utterance, which they attribute to post-interpretive processing. When they independently manipulated syntactic complexity and number of propositions in a sentence picture matching task, Waters et al. (1995) found that both AD patients and age-matched normal controls were equally affected by an increase in syntactic complexity but that patients showed a selective impairment in response to an increase in the number of propositions. Waters et al. (1995) concluded that patients with AD have an impairment in the post-interpretive working memory system but not in the interpretive working memory resources that are involved in syntactic processing. Thus, in contrast to Just and Carpenter, Waters and Caplan's theory would not necessarily predict a similar effect of intervening material on AD patients' processing of all kinds of dependencies. Instead, Waters and Caplan's theory predicts that if a certain task shows a selective impairment for AD patients in processing intervening material, it must be because the task requires post-interpretive processing, and therefore, this effect should be associated with the number of propositions added by the intervening material. In other words, any impairment associated with intervening material that is observed in AD patients but not in age-matched healthy controls should be related to an increase in the number of propositions in the intervening material but not to its syntactic complexity.

A third approach draws on an analysis of task demands. As noted above, a number of researchers have previously observed that on- and off-line tasks impose differing processing requirements. Kempler et al. (1998a), who found correlations between working memory performance and off-line task performance (in a sentence picture matching task and grammaticality judgement task) but not on-line performance (in a cross-modal naming task), suggested that one explanation for these results could lie in an analysis of task demands: performance in the off-line tasks and the working memory tasks was correlated because both tasks required substantial activation of semantic information (i.e., word meanings). In contrast, performance in the on-line task that Kempler et al. used did not require substantial activation of semantic information and thus was not correlated with working memory performance. A similar analysis can be extended to task demands imposed by different on-line tasks, and even to demands imposed by different linguistic processes. For example, linguistic working memory tasks typically require participants to report a series of words after a distraction interval; these words are kept active through some mix of phonological and semantic information (the exact mix could depend on both the task and the individual participant's strategies). We would expect on-line language comprehension tasks to correlate with working memory tasks to the extent that these tasks and the linguistic materials in them impose similar demands on the maintenance of phonological and semantic information. For example, the processing of some grammatical dependencies, like grammatical subject verb agreement, would appear to impose lesser demands on at least semantic representations than does the processing of some discourse dependencies, like co-referentiality between pronouns and their antecedents. This is because the semantic representation of a noun is largely irrelevant to the computation of grammatical subject-verb number agreement—it doesn't matter whether a particular noun phrase is ''the boy'' or ''the man'', what matters is whether it is singular or plural. By contrast, a very rich semantic representation of nouns may be required to determine whether ''the boy'' or ''the man'' is the most plausible antecedent of ''he'' in a discourse; phonological or morphological information is insufficient. Moreover, because, in English, pronouns in and of themselves are more grammatically ambiguous than agreement markers, the interpretation and processing of pronouns may be inherently more dependent on semantic information than the interpretation and processing of agreement markers. On this view, correlations between working memory tasks and on-line processing of pronominal reference in discourse should be stronger than between working memory tasks and processing of grammatical number agreement.

The research described in this paper was undertaken to investigate the role of working memory in processing dependencies of different types and different lengths by examining the effect of intervening material on AD patients' and healthy normal controls' (NCs) on-line processing of grammatical and discourse number agreement. We independently manipulated the nature of a number of agreement dependency (grammatical vs. discourse) and the distance between dependent elements to test the effects of each of these factors. We also looked for associations between performance in the different comprehension conditions and performance on working memory tasks.

Grammatical vs. discourse number agreement

Linguistic number, namely the morphological marking of words for number information, is a rudimentary linguistic mechanism that is common to most human languages (Croft, 1990; Greenberg, 1966). In English, both nouns and verbs are marked for number such that the grammatical subject and the matrix verb of each sentence must agree in number (e.g., ''The boy was running'' is grammatical but ''The boy were running'' is not). This subject-verb number agreement is reflected in almost every sentence spoken and heard by English speakers and is thought by linguistics to be entrenched in the grammar of the English language. Research on human sentence processing has also demonstrated that on-line sentence comprehension is sensitive to subject-verb agreement dependencies (Osterhout & Mobley, 1995) and is slowed down by subject-verb number agreement violations (e.g., Nicol, Forster, & Veres, 1997; Pearlmutter, Garnsey, & Bock, 1999). Furthermore, previous research with AD patients also found that, in spite of their many comprehension impairments, AD patients' sensitivity in cross-modal naming tasks to subject-verb number agreement violations between an adjacent subject and verb does not differ from that of age-matched healthy controls (Kempler et al., 1998b). Finally, there is substantial evidence that intervening material between the subject and the verb interferes with the processing of subject-verb number agreement in young healthy adults. In production, people tend to make more agreement errors when they produce intervening material between the subject and the verb (Bock & Eberhard, 1993; Eberhard, 1997), and in comprehension, people are less sensitive to agreement violations as the text distance between the subject and the verb increases (Nicol et al., 1997).

Number agreement is reflected not only in the grammatical dependencies between the elements of a single sentence but also in the discourse dependencies between elements of different sentences. For example, English requires both gender and number agreement between anaphoric pronouns and their antecedents (e.g., In ''The boy was running. He seemed happy'', *the boy* and *he* are likely to be interpreted as co-referential, whereas in ''The boys were running. He seemed happy'', *the boys* and *he* cannot be interpreted as co-referential). As with subject-verb number agreement, on-line sentence comprehension in healthy young adults has been shown to be sensitive to number agreement between pronouns and their antecedents (e.g., McDonald & MacWhinney, 1995). However, unlike subject-verb number agreement, previous research found the processing of number agreement between pronouns and their antecedents to be impaired in patients with AD (Almor et al., 1999). Thus, number agreement provides an ideal opportunity to assess the effects of intervening material on both a grammatical dependency that is known to be processed effectively by AD patients, and a discourse dependency that is known to be processed with difficulty by these patients.

Experiment 1 tests the effect of intervening material on the processing of subject-verb number agreement in AD patients and in normal elderly adults. Experiment 2 tests the effect of intervening material on the processing of pronoun-antecedent number agreement in the same groups of AD patients and elderly adults. Performance on both tasks is compared with independent measures of working memory. The research presented here thus extends our and others' previous research in that it directly compares the on-line processing of a grammatical dependency and a discourse dependency with varying amounts of intervening material.

EXPERIMENT 1: SUBJECT-VERB AGREEMENT AND LENGTH

Experiment 1 employed a cross-modal naming paradigm (Marslen-Wilson, Tyler, & Koster, 1993; Tyler & Marslen-Wilson, 1977) to assess AD patients' sensitivity to violations of intra-sentential number agreement and to test the effect of intervening linguistic material on this sensitivity. The cross-modal naming paradigm was chosen because it provides an assessment of participants' on-line processing, that is, it assesses comprehension as it is occurring, and does not impose extensive memory demands beyond what is required for comprehension itself. Previous research has shown that mild to moderate AD patients can perform a cross-modal naming task, and that in many cases their sensitivity to grammatical violations as gauged by this task is comparable to that of healthy controls' (Kempler et al., 1998a, 1998b; Nebes et al., 1986).

In cross-modal naming tasks, participants listen to auditory sentence fragments and then read aloud as quickly as possible a visually presented word which appears on a computer screen at the offset of the auditory sentence fragment. The visual target is either a good (grammatical, interpretable) continuation of the sentence fragment or not. Naming latencies for visual targets in the cross-modal task provide a measure of the degree to which participants are sensitive to the appropriateness and grammaticality of the target word as a continuation for the auditory fragment. If participants are processing the sentence, then they are slower to read anomalous continuations compared to felicitous ones. In this experiment, visual targets consisted of verbs that were grammatical continuations for the auditory fragment and verbs that were not grammatical because of incompatible number (“The boy WAS” vs. “The boy WERE”). Because subject-verb agreement is a basic grammatical constraint that is encountered by all speakers and listeners with high frequency, and because previous research found intact basic grammatical processing in patients with AD, we expected both normal elderly participants and AD patients to show an effect of verb grammaticality—all participants should name grammatical verb targets faster than ungrammatical ones. We also manipulated the number of intervening words between the subject and the verb such that in the short condition, the visual target verb appeared immediately after the subject noun, and in the long condition, the target appeared following a relative clause that modified the subject noun. If the presence of intervening words generally affects people's ability to detect subject-verb agreement violations, then both groups should exhibit reduced sensitivity to verb grammaticality in the long condition compared to the short condition. However, if patients are more affected by intervening material than are normal controls, then there should be a three-way interaction between population, length, and grammaticality, such that the difference between the grammaticality effects in the short and long conditions is larger for the patients than for the healthy controls.

Because working memory deficits are widespread in AD and have been suggested as a possible source of language comprehension deficits, we will also correlate language comprehension performance with independent measures of working memory. If the introduction of intervening material during processing of number agreement taxes working memory resources, then grammaticality effects in the long condition should correlate with performance on working memory tasks, but grammaticality effects in the short condition should correlate with working memory performance more weakly or not at all.

Method

Participants. Ten participants diagnose with probable Alzheimer's disease and ten healthy age-matched controls participated in this study. The Alzheimer participants were referred by the University of Southern California Alzheimer's Disease Research Center and the Alzheimer's Disease Diagnostic and Treatment Center at Rancho Los Amigos Medical Center. All AD participants met the NINCDS–ADRDA criteria for probable Alzheimer's disease (McKhann et al., 1984). Results of neurological, laboratory (including computed tomography (CT) or magnetic resonance (MR) scan), and neuropsychological assessment failed to suggest other causes of dementia. All participants were native speakers of Standard American English. Participant information is shown in Table 1. The AD participants were mildly to moderately demented as gauged by the Mini-Mental State (Folstein, Folstein, & McHugh, 1975). The normal elderly control (NC) participants were slightly younger than the patients

TABLE 1
Experiment 1. Participant information

	MMSE Mean (SD, range)	*Age* Mean (SD, range)	*Education* Mean (SD, range)
AD (N = 10)	21 (3.8, 17–28)	82 (4.3, 76–89)	15 (2.5, 11–20)
NC (N = 10)	29.4 (1.1, 28–30)	78 (4.2, 70–84)	17 (1.9, 14–22)

and had completed slightly more years of education, though these differences did not quite reach statistical significance: age, $t(18) = 2$, $p < .07$; years of education, $t(18) = 1.97$, $p < .07$. However, because there were these small differences, we included the factors of age and education in all analyses that include a group factor.

Materials and procedure

Cross-modal naming. Twenty auditory stimuli were constructed by combining an introductory sentence and a sentence fragment that ended after the grammatical subject, before the main verb (see Table 2 for a sample item in all conditions). In half of the items, the subject of the sentence fragment was a plural noun phrase and in the other half singular. The verbs *was* and *were* were presented as visual targets. Each verb thus constituted a grammatical continuation to half of the items and an ungrammatical continuation to the other half. Each sentence fragment appeared in two conditions, short and long. In the short condition, the subject of the fragment was a three-word noun phrase of the form *determiner-adjective-noun*, and the auditory fragment ended at the offset of the noun. In the long condition, the subject noun phrase was followed by a 10–15-word relative clause. All of the relative clauses were subject relatives, that is, the subject of the sentence fragment was also the subject

TABLE 2
Sample item in Experiment 1 (visual targets in capital letters)

Short condition
Some people always lose their keys.
The old lady WAS/WERE

Long condition
Some people always lose their keys.
The old lady, who searched very carefully through every one of those old trash cans behind the stores, WAS/WERE

of the relative clause. The last word of the relative clause was always a noun that was marked for a different number than the subject noun. Appendix A shows the complete list of items. The visual targets were always presented at the offset of the auditory fragment. Response times were measured from the onset of the visual target to voice initiation that was detected by the microphone. Thus in the short condition the visual target verb was presented immediately following the head noun that carried the number information, and in the long condition it was presented following the relative clause. The experiment therefore had three factors—population (AD, NC), grammaticality (grammatical, ungrammatical targets), and length (short, long).

We tested participants in four separate sessions, a minimum of one week apart for the AD patients and three weeks apart for the control participants. The stimuli presented in each sessions were counterbalanced so that a participant never encountered more than one version of the same item in the same session. Stimuli were pseudo-randomly mixed with fillers from different experiments using different sentence structures and different visual targets (which included other verbs, nouns, and adjectives). In some filler items the visual target was presented at the middle of the sentence and in others it was presented at the end of the sentence. In each session, 25% of the items were from this experiment and 75% were fillers (items from other experiments, including the items for Experiment 2 below). In half of all items the visual target was a good continuation of the auditory context, and in the other half it was not. Ten practice items began each session.

Stimuli were digitally recorded using MacRecorder and a Macintosh computer in a sound attenuated booth with a constant mouth-to-microphone distance of 8 inches. A standard filler word was recorded at the end of each fragment instead of the target word (to avoid coarticulation cues) and then digitally excised.

Each participant was tested in a quiet room, sitting in front of a Macintosh Classic computer. The experiment was controlled by the PsyScope software package (Cohen, MacWhinney, Flatt, & Provost, 1993). Auditory stimuli were presented via a high quality loudspeaker, with volume adjusted to ensure that each participant could hear the stimuli adequately. Participants were asked to read the visual targets aloud as quickly and accurately as possible (Marslen-Wilson et al., 1993). In order to assure that participants were paying attention to the stimuli (and not just reading the words aloud, see Tyler & Marslen-Wilson, 1977), we included an additional comprehension task: after naming each target word, control participants were asked whether the word was a good or bad continuation of the sentence. Pilot testing with AD participants indicated that this secondary task was so distracting that they could not produce accurate and fast responses to both the primary (word reading) and secondary tasks. Therefore, an alternative directed attention task was developed for the AD participants: they were asked yes/no comprehension questions for 20% of the context sentences. This secondary task successfully directed their attention to the stimuli and did not interfere with their ability to perform the primary task. For both groups, responses to the secondary task were not scored, but merely served to encourage the participants to integrate the context fragment and the visual target. It should be noted that even though subjects from the two populations engaged in a slightly different secondary task, this difference was necessary in order to ensure that their performance of the main task would be comparable. We have now used this combination of secondary tasks in a number of studies that found no difference between the performance of AD patients and normal controls in the main task (Kempler et al., 1998a, 1998b; also see the results of the present experiment below). In light of those findings it is highly unlikely that the small difference in the secondary task used with the two populations introduced any meaningful confounds to the results of the cross-modal naming.

Working memory and mental status. Previous research with AD patients (e.g., Waters & Caplan, 1997), as well as our pilot work, suggested that working memory tasks that are standardly used with young participants, such as Daneman and Carpenter's (1980) span tasks, are too difficult for AD patients and yield insufficient variance for meaningful statistical analysis. Therefore, we developed two new working memory tasks, month ordering and digit ordering, to assess linguistic working memory. Like listening span tasks, our ordering tasks were designed to require simultaneous storage and manipulation of information. Tasks that contain both storage and processing components have been found to be good correlates of language processing abilities, in contrast to tasks with minimal processing components, such as simple forward digit span (Daneman & Carpenter, 1980). In the month ordering task, participants were required to put into calendar sequence an increasingly long set of months presented out of calendar order. For example, in a given trial, a participant may have heard: "June, September, February", and should have responded: "February, June, September". In the digit ordering task, participants were required to put into numerical order an increasingly long set of digits presented out of order. For example, when presented with the list of digits: "nine, three, five", the target response would have been: "three, five, nine". These tasks create similar working memory and processing demands to those involved in auditory language comprehension, in that the participant must process a speech signal (the list of months or digits), maintain a representation of the input in mind, convert it to a semantic representation (i.e., establish that the phonological string "June" means the month of June), and use this representation to re-order the input and formulate a response. Note that due to the presentation of nonconsecutive digits and months, participants could not perform the task by simply comparing the items to a list of digits and months in long-term memory. The administration procedure for these tasks was adapted from Daneman and Carpenter (1980). There were four trials presented at each span level, starting with two-item trials at the first level, and ending with trials containing six items at the fifth and last level. Testing was stopped after a participant incorrectly sequenced two trials at a span level. Participants could score from 0 to 20, which was the overall number of correct sequences. Our previous research (Almor et al., 1999; Kempler et al., 1998a, 1998b) has shown that although AD patients perform poorly on these tasks relative to elderly controls, they can generally perform the task with short lists and thus seem to preserve at least some knowledge regarding the correct order of digits and months.

To evaluate the effect of overall dementia severity and to allow us to factor out the effect of dementia severity in subsequent analyses, we administered the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975).

Results

Naming latencies were analysed after eliminating all machine and reading errors (7.7% of responses) and latencies beyond 3 standard deviations from a participant's mean naming latency in each session (affecting a further 1.4% of the data.) AD patients were overall slower than NCs (patients' mean RT was 1676 ms whereas NCs' was 1028 ms), but this difference was not statistically significant, $t(18) = 1.1, p < .3$. To minimise irrelevant effects due to variations in the speed of individual participants across sessions and establish a common basis for evaluating responses of different participants from different populations, naming latencies were normalised by transforming them into z-scores. The z-transformation was based on each participant's mean and standard deviation for trials from this experiment in each session. A z-transformation was chosen so as to minimise differences in variance between the different groups and ensure that the deletion of error trials would not lead to an artificial change in the means (without the z-transformation, error trial removal from "slow sessions" would result in the incorrect shortening of the mean RT of the conditions to which the error trials belonged).[1] Filler trials were not included in calculation of the mean and standard deviation for the transformation. The normalised mean naming latencies are shown in Figure 1.

A 2 × 2 × 2 ANCOVA, Population × Grammaticality × Length with covariates age and education, found a main effect of grammaticality such that naming ungrammatical verbs was slower than naming grammatical ones, $F_1(1, 18) = 63.54$, $p < .001$, $F_2(1, 19) = 63.2$, $p < .001$,[2] and also an interaction between grammaticality and length such that the grammaticality effect was almost twice as large for the short items as for the long items, $F_1(1, 18) = 8.27$, $p < .01$, $F_2(1, 19) = 6.65$, $p < .02$. There were no effects of population, and as is evident in Figure 1, both groups were equally sensitive to the grammaticality of the visual target and were similarly affected by length. There was also a main effect of length, with longer RTs in the short condition than in the long condition, $F_1(1, 18) = 5.62$, $p < .03$, $F_2(1, 19) = 4.34$, $p < .051$, which may reflect participants' increased readiness for a visual target of any kind to appear later in the sentence, as in the long condition, rather than immediately after the grammatical subject, as in the short condition. No other main or interaction effects were significant.

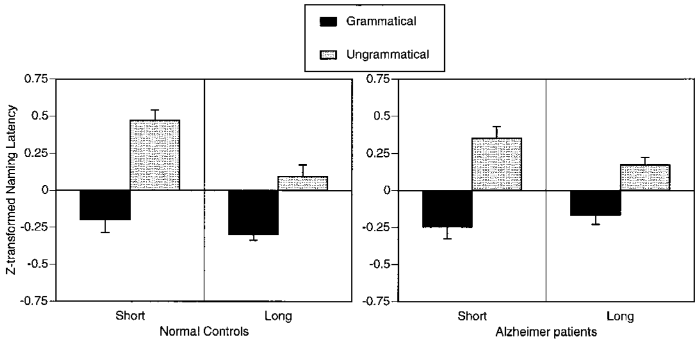

Figure 1. Verb naming latency in Experiment 1.

[1] A z-transformation was chosen because it provides an effective means for equating the variance in different sessions and for subjects from both groups. The ANCOVAs and ANOVAs reported for this and the following experiment were also performed after a log transformation of the scores. The results of the analyses based on these log transformed scores were identical to the results of the analyses reported here for both experiments with the exception that the interaction of Length and Grammaticality found in Experiment 1 was only significant at a $p < .055$ level with the log transformed scores. Because the log transformed scores still failed to yield homogeneous variance for the two populations and because they do not solve the problem introduced by the removal of error trials, we only report the results from the analyses of the z-transformed scores.

[2] The item analyses did not include the age and education covariates.

Dementia severity and working memory. AD patients scored lower than NCs on the MMSE, 21 vs. 29.4, $t(18) = 5.29$, $p < .001$, the month ordering task, 6 vs. 13.9, $t(18) = 4.98$, $p < .001$, and the digit ordering task, 9.3 vs. 17.9, $t(18) = 5.03$, $p < .001$. There was a correlation between MMSE and month ordering while controlling for age and education, Pearson $r(16) = .52$, $p < .02$, and between MMSE and digit ordering, $r(16) = .57$, $p < .02$. There was also a strong correlation between performance on the month ordering and digit ordering tasks, $r(16) = .88$, $p < .001$, even after the effect of general dementia, gauged by MMSE, was factored out, $r(15) = .83$, $p < .001$. To obtain a single working memory measurement for subsequent analyses, a composite ordering score was calculated by averaging each participants' month and digit ordering scores. AD patients' mean composite ordering score was 7.65, and NCs' 15.9, $t(18) = 5.4$, $p < .001$.

To assess whether sensitivity to the grammaticality of the visual target verbs was linked to dimentia severity and/or working memory performance, we calculated the correlation between performance in the on-line task and scores on the MMSE and working memory tasks. Because we wanted to test for the relation between on-line performance, working memory, and MMSE in general, we included the data from all participants, AD patients and NCs, in these analyses. For each participant, we calculated a "short grammaticality sensitivity score" and a "long grammaticality sensitivity score" by subtracting the participant's mean normalised grammatical target naming latency from his/her mean normalised ungrammatical target naming latency separately for the short and long conditions. We also calculated for each participant a "length score" by subtracting the participant's long grammaticality sensitivity score from his/her short grammaticality sensitivity score. This length score provides a measure of the reduction in grammaticality sensitivity due to sentence length.

MMSE scores did not correlate with short grammaticality sensitivity scores $r(16) = .19$, $p < .45$, long grammaticality sensitivity scores, $r(16) = .31$, $p < .2$, or length scores, $r(16) = -.07$, $p < .8$. Factoring out the effect of working memory by controlling for the composite working memory score did not enhance these correlations: with short grammaticality sensitivity scores, $r(15) = -.05$, $p < .9$; with long grammaticality sensitivity scores, $r(15) = .26$, $p < .32$; or with length scores, $r(15) = -.26$, $p < .31$. Similarly, the composite working memory scores did not correlate with short grammaticality sensitivity scores, $r(16) = .4$, $p < .1$, long grammaticality sensitivity scores, $r(16) = .18$, $p < .47$, or length scores, $r(16) = .26$, $p < .3$. The working memory scores did not correlate with any on-line scores even after MMSE was factored out: with short grammaticality sensitivity scores, $r(15) = .37$, $p < .15$; with long grammaticality sensitivity scores, $r(15) = .01, p < .97$; and with length scores, $r(15) = .35, p < .16$. Thus, performance in the on-line comprehension task, with or without intervening material, did not correlate with performance in the working memory tasks or MMSE.

Discussion

The use of long distance dependencies within sentences did not elicit any indication of impairment in patients' sentence processing. NCs and AD patients were equally sensitive to subject-verb agreement in the cross-modal naming task. Furthermore, for both populations, intervening material between the head noun and the verb had a similar adverse effect on this sensitivity. We can thus conclude that previous findings of unimpaired performance in the cross-modal naming task (e.g., Kempler et al., 1998a, 1998b; Nebes et al., 1986) did not stem from using only short conditions in those studies; AD patients can process grammatical structures very well even in longer, more complex sentences than previously tested. Participants' sensitivity to subject-verb number agreement, with or without intervening material, was not linked to working memory performance, nor was the reduction in this sensitivity that was caused by intervening material. Therefore, the decreased sensitivity in the long condition was apparently not related to the working memory performance assessed by the digit and month ordering tasks. One likely explanation for the strong grammaticality effect in the short than in the long condition (for both groups) is that participants' expectation for a verb was stronger in the short condition than in the long condition because verbs appear more often immediately after the subject than after a relative clause. This may have caused participants to respond faster to appropriate verbs and slower to inappropriate verbs in the short condition than in the long condition (even though their overall expectation for a visual target of any kind was stronger in the long condition as is indicated by the overall shorter RTs in the long condition). Another possible explanation for the reduced grammaticality sensitivity in the long condition is that the last noun in the intervening material in the long condition always disagreed with the number of the subject, and may have thus created an interference effect much like in studies with young healthy adults (Nicol et al., 1997; Pearlmutter et al., 1999). Regardless of the interpretation for the interaction of length and grammaticality in this experiment, the lack of population differences and lack of association with working memory suggest that intervening material in intra-sentential grammatical dependencies does not place extra demands on the type of working memory resources that are tested here and diminished in AD.

EXPERIMENT 2: PRONOUN-ANTECEDENT AGREEMENT

Although Experiment 1 found no evidence that AD patients' grammatical processing is disproportionately affected by intervening material between dependent constituents within sentences, the longer amount of intervening material that typically separates related elements in discourse could still underlie the impairment that AD patients show in discourse processing. Also, it may be that the lack of correlations between on-line processing in Experiment 1 and working memory performance indicates an insufficient power in our various tests. Experiment 2 tested the effect of intervening material on the processing of a discourse dependency, number agreement between pronouns and their antecedents, and assessed the role of working memory in this processing.

Method

Participants. The same participants as in Experiment 1 were tested in this experiment (see Table 1) in the same four sessions.

Materials and procedure. This experiment employed the same cross-modal naming paradigm as in the previous experiment. Stimuli consisted of 20 auditory discourse fragments followed by a visual target pronoun that in half of the trials constituted an appropriate co-referential continuation to the discourse fragment and in the other half did not (see Table 3 for a sample item). The discourse fragments included an initial sentence, which always introduced two entities, one plural, and one singular, and concluded with a final incomplete sentence fragment. The sentence fragment mentioned one of the entities again and ended right before the other entity was likely to be mentioned. In the short condition, items consisted of only the initial sentence and the final fragment. In these items, the target pronouns were separated from their antecedents by 10 to 15 words; note that this "short" condition contained roughly the same number of

TABLE 3
Sample item in Experiment 2 (visual targets in capital letters)

Short condition
The children loved the silly clown at the party.
During the performance, the clown threw candy to HIM/THEM.
Long condition
The children loved the silly clown at the party.
The show was very funny.
During the performance, the clown threw candy to HIM/THEM.

intervening words as in the "long" condition of Experiment 1. Items in the long condition also included a middle sentence, five to seven words long, which continued the same topic but did not mention either entity from the first sentence. A complete list of items appears in Appendix B. Because AD patients are impaired in their ability to maintain number information for pronoun reference resolution (Almor et al., 1999), we expected them to exhibit reduced sensitivity to pronoun appropriateness, compared to NCs. However, we were also interested in whether this sensitivity would be affected by length and whether any such length effects would be associated with working memory scores. Indeed, because working memory measures have previously been shown to correlate with the ability to interpret pronouns, particularly in long discourses (Daneman & Carpenter, 1980), we expected that in this experiment too, performance on the working memory measures would correlate with pronoun processing.

Overall, like Experiment 1, this experiment had three factors—population (AD, NC), appropriateness (appropriate, inappropriate), and length (short, long). Auditory stimuli were recorded as in Experiment 1. Participants were tested over the same four sessions as in Experiment 1 such that 25% of the items in each session were from this experiment and 75% were filler items (including the items from Experiment 1 above).

Results

As in the previous experiment, naming latencies were analysed after eliminating all machine and reading errors (6.8% of responses) and latencies beyond 3 standard deviations from a participant's mean naming latency in each session (further 1.8% of the data). In this experiment, AD patients were again slower than NCs (patients' mean RT was 958 ms and NCs' was 836 ms), but, again, this difference was not statistically significant, $t(18) = .8$, $p < .4$. As in the previous experiment, data were normalised by transforming them into z-scores based on each participant's mean and standard deviation calculated for the scores from this experiment separately in each session. The resulting pronoun naming latencies are shown in Figure 2.

A $2 \times 2 \times 2$ ANCOVA, Population $\times$ Appropriateness $\times$ Length with Age and Education as covariates, found a main effect of appropriateness with faster naming of appropriate pronouns than inappropriate ones, $F_1(1, 18) = 112.0$, $p < .001$, $F_2(1, 19) = 60.71$, $p < .001$, and an interaction between population and appropriateness such that NCs showed an appropriateness effect almost three times larger than did the patients, $F_1(1, 18) = 27.98$, $p < .001$, $F_2(1, 19) = 24.33$, $p < .001$. This interaction shows that the on-line task is sensitive enough to capture population differences. The subject analysis also detected a marginal main effect of

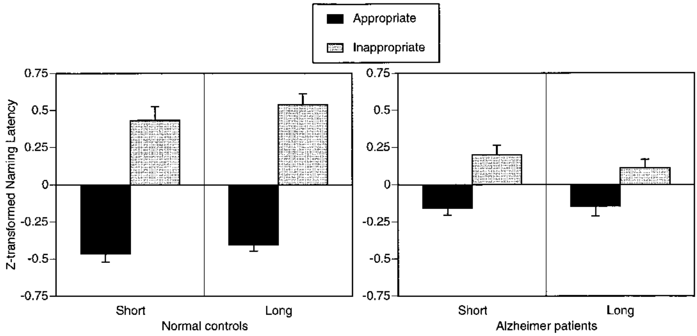

Figure 2. Pronoun naming latency in Experiment 2.

population, $F_1(1, 16) = 4.41$, $p < .06$, although this was not significant in the item analysis, $F_2 < 1$. No other main or interaction effects were significant. Thus, in this experiment, although AD patients were sensitive to the appropriateness of the visual target, their sensitivity was impaired. Note that it is unlikely that this difference between the groups is an artifact of the z-transformation. Although, in principle, the higher variability in patients' responses could have led to an artificially reduced sensitivity effect for the patients with the z-transformed scores, the fact that we did not observe any group difference in Experiment 1 shows that this was not the case. If the z-transformation was the reason for the differences in sensitivity between the two populations in this experiment then we would have expected to find similar differences in group sensitivity in Experiment 1, which we did not.[3]

The presence of intervening material did not have an effect on performance in either group. To make sure that this lack of length effect was not due to any group difference, we carried out a Length × Appropriateness analysis separately for the AD patients and for the NCs. Both analyses found a main effect of appropriateness but no effect of length or a length by appropriateness interaction, thus confirming the results of the previous analysis. The finding that, in contrast to the previous experiment, there was no main effect of length in this experiment may be explained by the fact that the visual target in both the short and the long conditions in this experiment appeared at the same position within the final sentence fragment—participants had no reason to expect the visual target in the long condition more than the short condition or vice versa.

[3]More credence is lent to the interpretation of this difference as indicating a true population effect by the fact the same effect was also detected in the analysis of the log transformed scores mentioned in Footnote 1.

Dementia severity and working memory. As in the previous experiment, we conducted a series of correlational analyses on performance in the on-line task and scores on the MMSE and working memory tasks. We calculated for each participant a "short appropriateness sensitivity score" and a "long appropriateness sensitivity score" by subtracting the participant's mean normalised appropriate target naming latency from his/her mean normalised inappropriate target naming latency separately for the short and long conditions. Again, we calculated for each participant a "length score" by subtracting the participant's long appropriateness sensitivity score from his/her short appropriateness sensitivity score.

MMSE scores did not correlate with short appropriateness sensitivity scores, $r(16) = .21$, $p < .4$, marginally correlated with long appropriateness sensitivity scores, $r(16) = .46$, $p < .06$, and did not correlate with length scores, $r(16) = -.24$, $p < .34$. Factoring out the effect of working memory by controlling for the composite working memory score totally removed the marginally significant correlation between MMSE and long appropriateness sensitivity scores, $r(15) = .15$, $p < .57$, and there was still no correlation between MMSE and short appropriateness sensitivity scores, $r(15) = -.12$, $p < .66$, or length scores, $r(15) = -.22$, $p < .4$. However, in this experiment, the composite working memory scores correlated with short appropriateness sensitivity scores, $r(16) = .53$, $p < .03$, and with long appropriateness sensitivity scores, $r(16) = .64$, $p < .004$, but not with length scores, $r(16) = -.11$, $p < .67$. These working memory score correlations were virtually unaffected by factoring out the effect of MMSE. The composite working memory scores still correlated with short appropriateness sensitivity scores, $r(15) = .5$, $p < .04$, and with long appropriateness sensitivity scores, $r(15) = .53$, $p < .03$, but not with length scores, $r(15) = .03$, $p < .91$. The fact that in this experiment, performance in the on-line task, in both long and short conditions, correlated with our measures of working memory even after the effect of general cognitive functioning was factored out shows that the ordering tasks provide a measure that is powerful enough to allow the detection of correlations with our on-line comprehension measures. Finally, the fact that even in this experiment, in which overall performance in the short and long conditions correlated with working memory, length scores did not correlate with working memory suggests that the addition of this amount of intervening material only poses a negligible burden on the working memory resources of the sort that are tapped by the ordering tasks.

Discussion

The group by appropriateness interaction found in this experiment indicates that the AD patients were impaired in their ability to differentiate between number-appropriate and number-inappropriate pronouns. The presence of intervening material, however, did not seem to affect performance of the AD patients or the controls—each group was equally sensitive to appropriateness in the short and long conditions. Although performance on both conditions of this task appeared strongly related to working memory performance, processing intervening material (the length effect) was not linked to working memory. Dementia severity does not appear to be an important factor in processing these types of discourses, given that the only significant MMSE correlation disappeared when the effect of working memory was factored out.

These results contrast with those of Experiment 1, in which performance of AD patients and controls did not differ. The comparison between the long condition of Experiment 1 and the short condition of Experiment 2 is particularly interesting, because these two conditions contained very similar numbers of intervening words between the dependent items. AD patients were no worse than controls in the long grammatical dependency of Experiment 1, but in Experiment 2, AD patients were less sensitive to discourse number agreement compared to the controls. This result underscores the importance of the kind of dependency being comprehended; the pattern of results cannot be attributed simply to the number of words intervening between dependent items. While length seems to modulate processing difficulty in some circumstances, the nature and strength of its effect is in turn modulated by the nature of the dependency. One factor that may explain the population differences in discourse is discourse focus. In all the items in this experiment the visual target referred to the focused discourse entity, but the fact that the nonfocused entity appeared in the subject position of the target sentence may have been interpreted by subjects as an indication of focus shift (Grosz, Joshi, & Weinstein, 1995). The shift in focus may have eliminated any length effects and may have been the reason for AD patients' poor performance; their working memory impairment may interfere especially with transitions in discourse focus. The particular role discourse focus may play in AD patients' impaired performance is an interesting issue that merits further research.

GENERAL DISCUSSION

Experiment 1 demonstrated that, in the cross-modal task, AD patients were able to process grammatical dependencies just as well as control participants in both short and long sentences, although performance for both groups was significantly worse in the long conditions. The finding that AD patients' performance is comparable to NCs' despite significant impairments in the working memory tasks suggests that working memory as we have measured it is largely irrelevant to the processing of subject-verb number agreement. Experiment 2 showed significant language processing deficits in the same group of AD patients (compared to controls) when they were required to process a discourse dependency, pronoun-antecedent number agreement, in both short and long conditions. Performance with these discourse dependencies was strongly correlated with measures of working memory capacity.

The findings presented here suggest that the on-line processing of grammatical dependencies is different from the on-line processing of discourse dependencies in three important ways. First, it appears that regardless of the presence of linguistic material intervening between dependent words, patients with AD are not impaired in the on-line processing of the subject verb agreement grammatical dependency, but are impaired in the on-line processing of the pronoun antecedent discourse dependency. Importantly, this difference holds even when there is a similar number of words separating the subject from the verb and the antecedent from the pronoun (as in the long condition of Experiment 1 and the short condition of Experiment 2). Second, although intervening material between elements of a linguistic dependency affects comprehension for both normal elderly and patients with AD, it only affects the processing of the subject-verb number agreement grammatical dependency. Both patients' and healthy elderly's on-line processing of the grammatical dependency was compromised by the presence of intervening material between the subject and the verb, but their on-line processing of the discourse dependency was not affected by additional intervening material. Third, the on-line processing of the pronoun-antecedent number agreement discourse dependency but not the on-line processing of the subject-verb number agreement grammatical dependency is linked to working memory performance as gauged by the ordering tasks.

Implications for working memory

Our results are not compatible with Just and Carpenter's theory of verbal working memory and its role in language comprehension (Just & Carpenter, 1992; Just et al., 1996). While our finding of reduced sensitivity for subject-verb number agreement when the subject and the verb are separated by intervening material is predicted by Just and Carpenter's single verbal working memory resource theory, the fact that there was no difference between memory impaired patients and NCs is not. According to Just and Carpenter's theory, the patients should be affected more by intervening material than NCs because of their overall reduced verbal working memory resources. The lack of difference between patients and NCs cannot be simply attributed to a lack of power in the cross-modal naming task because this same task used with the same participants was in fact powerful enough to capture population differences in Experiment 2. Also incompatible with the Just and Carpenter theory is our failure to find any correlation between working memory performance and the reduction in grammatical sensitivity that was associated with intervening material. Although, as a null result, this failure to find correlations should be interpreted with caution, the fact that the same working memory measures correlated with sensitivity to pronoun appropriateness in Experiment 2 supports the notion that the effect of intervening material on processing subject-verb number agreement is truly independent of the resources demanded by working memory tasks. Thus, our results are not compatible with a single resource theory such as Just and Carpenter's. In order for Just and Carpenter's theory to account for our results, it would have to be extended to associate very different resource demands with the processing of grammatical dependencies than with the processing of discourse dependencies. It is not clear that, with such an extension, Just and Carpenter's theory would not become practically indistinguishable from multiple resource theories (e.g., Caplan & Waters, 1999).

Although our findings may at first appear compatible with the general idea of separate grammatical and discourse processing, they cannot be easily accommodated by Waters and Caplan's modular framework (Caplan & Waters, 1999; Waters & Caplan, 1996, 1997; Waters et al, 1995). Waters and Caplan argued that patients with AD have an impairment in the post-interpretive working memory system but not in the interpretive working memory resources that are involved in syntactic processing. The present findings, however, do not support this hypothesis. Patients in our studies performed normally when tested on grammatical relations, even when we introduced a relative clause that added both syntactic complexity and an additional proposition. According to Waters and Caplan, insofar as adding propositions to a sentence increases the post-interpretive load (which puts demands on "post-interpretive" working memory), the lack of a number of propositions effect is an indication that the comprehension task we used, cross-modal naming, is not sensitive to the post-interpretive processing of sentences. If this task measured post-interpretive processing, we should have observed a selective impairment in the AD patients in the long conditions of Experiment 1 because this condition had an additional proposition. However, if this task does not measure post-interpretive processing, there should also have been no population differences in either experiment. The fact that Experiment 2 showed a significant difference between the populations appears to undermine Waters and Caplan's contention that AD patients are impaired only in their post-interpretive processing. In addition, the fact that performance in Experiment 2 was linked to working memory performance runs counter to Waters and Caplan's claim that working memory as gauged by tasks such as the ones we used does not affect interpretive processing. One might imagine an extension of the Waters and Caplan account in which interpretive processes utilise the output of post-interpretive processes from prior sentences, which could explain why AD patients show an impairment in interpretive processing of a sentence that includes references to previous discourse, as in Experiment 2. Because Waters and Caplan's approach focuses on intra-sentential phenomena, however, it is difficult to assess the implication of such an extension. Thus, while our results are in line with the general notion that grammatical and discourse processing are distinct, they are not compatible with Waters and Caplan's specific version of this general theory.

The traditional modular perspective, however, is not the only theoretical perspective that is compatible with our findings. The contrast between patients' intact grammatical performance in Experiment 1 and their impaired discourse performance in Experiment 2 can be interpreted in terms of the processing demands imposed by grammatical and discourse dependencies, and relating these processing demands to statistical facts about the language. Subject-verb number agreement is an extremely high frequency dependency that is a necessary and predictable part of every English sentence. Although not all sentences in English show overt subject-verb agreement, it is likely that subject-verb agreement is computed even when not overtly coded in the sentence (Chomsky, 1981). We suggest that the frequency and regularity of this agreement pattern is reflected in good comprehension by AD patients, even as they become impaired in comprehending rarer and more difficult structures during the course of the disease. One computation that suffers in AD is pronoun antecedent number agreement, which is both less frequent than subject-verb agreement (almost all English sentences have a subject and a main verb, but not all English sentences have a pronoun) and less predictable (whereas a comprehender knows that a verb must arrive at some point after a subject-noun phrase, comprehenders cannot predict with certainty that a discourse entity will be referred to again with a pronoun). On this view, the extent to which AD patients will exhibit impairments in linguistic processing varies with the frequency and predictability of the individual dependency; it is not necessarily the case that all discourse dependencies will yield an impairment or that all grammatical dependencies will be spared.

This explanation of performance on our comprehension tasks complements the proposal provided by Bates, Harris, Marchman, Wulfeck, and Kritchevsky (1995) to account for their AD production data. Using an elicited production task, Bates et al. found that AD patients produced the same range of syntactic constructions as NCs, but that the distribution of constructions within that range was different; the patients produced fewer of the infrequent constructions (e.g., passives) than NCs. Bates et al. interpreted this finding as indicating that while there is no purely syntactic deficit in language production in AD, the patients' impairments had a larger effect on infrequent constructions than frequent ones. Together, the two studies show that in both language production and language comprehension, frequency and predictability constrain AD patients' performance.

Clearly, the data presented in this paper are insufficient for disentangling this statistical view from the more traditional modular view of Caplan and Waters (1999). This is because the grammatical and discourse dependencies that were tested here represent two ends of a range of frequency and predictability: the grammatical dependencies were very frequent and highly predictable while the discourse dependencies were relatively infrequent and unpredictable. Disentangling the statistical and modular views would require testing the processing of grammatical dependencies that are less frequent and less predictable than subject-verb agreement as well as the processing of discourse dependencies that are more predictable than inter-sentential pronoun antecedent agreement. Further experimentation is also required in order to determine how frequency and predictability should be measured and in particular whether the frequency of subject-verb agreement should be based on the number of times the grammatical dependency occurs or the number of time the dependency occurs for each specific verb form.

The analysis we provided here can be extended more generally to the relationship between comprehension performance and working memory measures. We have suggested that the working memory task is more similar in task demands to the pronoun processing task than to the interpretation of grammatical number agreement, in that both pronoun processing and working memory tasks require activation of semantic information (word meaning) to a greater degree than grammatical processing. Following MacDonald and Christiansen (in press) and Christiansen and MacDonald (1999), we further suggest that working memory tasks have often been accorded a special status in the literature, as they are often seen to be the pure reflection of a "working memory capacity", and not subject to any task analysis. This view has lead to misleading interpretations of correlations between working memory and comprehension tasks, in that researchers have sometimes made the mistake of inferring a causal relationship from this correlation and concluding that working memory is "used" in the comprehension task (see MacDonald & Christiansen, in press, for review). Our alternative examines the task demands imposed by working memory tasks and interprets correlations between working memory and comprehension tasks as indicative of the similar task demands imposed by the two tasks. We thus consider the ordering tasks not as assessing an independent underlying working memory resource, but as providing a good indicator for which linguistic tasks, off-line and on-line, would be hard for AD patients.

Manuscript received October 1998
Revised manuscript received April 2000

REFERENCES

Almor, A., Kempler, D., MacDonald, M.C., Andersen, E.S., & Tyler, L.K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's Disease. *Brain and Language*, *67*(3), 202–227.

Bates, E., Harris, C., Marchman, V., Wulfeck, B. & Kritchevsky, M. (1995). Production of complex syntax in normal aging and Alzheimer's Disease. *Language and Cognitive Processes*, *10*(5), 487–539.

Bock, K., & Eberhard, K.M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*(1), 57–99.

Bock, K., Nicol, J., & Cutting, J.C. (1999). The ties that bind: Creating number agreement in speech. *Journal of Memory and Language*, *40*(3), 330–346.

Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–126.

Carpenter, P., Miyake, A., & Just, M.A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, pp. 1075–1122. San Diego, CA: Academic Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Forris.

Christiansen, M.H., & MacDonald, M.C. (1999). Fractionating working memory: Even in pebbles, it's still a soupstone. *Brain and Behavioral Sciences*, *22*, 97–98.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope. Software developed at Carnegie Mellon University, Department of Psychology.

Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Eberhard, K.M. (1997). The marked effect of number on subject-verb agreement. *Journal of Memory and Language*, *36*(2), 147–164.

Emery, O.B. (1985). Language and aging. *Experimental Aging Research*, *11*(1), 3–60.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Greenberg, J.H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In J.H. Greenberg (Ed.), *Universals of grammar*, 2nd edn, pp. 73–113. Cambridge, MA: MIT Press.

Grober, E., Bang, S. (1995). Sentence comprehension in Alzheimer's disease. *Developmental Neuropsychology*, *11*(1), 95–107.

Grosz, B.J., Joshi, A.K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–226.

Just, M.A., & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *98*(1), 122–149.

Just, M.A., Carpenter, P.A., & Keller, T.A. (1996). The capacity theory of comprehension: new frontiers of evidence and arguments. *Psychological Review*, *103*(4), 773–780.

Kempler, D. (1995). Language changes in dementia of the Alzheimer type. In L. Lubinski (Ed.), *Dementia and communication: Research and clinical implications*, pp. 98–114. San Diego: Singular Publishing Group.

Kempler, D., Almor, A., & MacDonald, M. (1998a). Teasing apart the contributions of memory and language impairments in Alzheimer's disease: An on-line study of sentence comprehension. *American Journal of Speech-Language Pathology*, *7*(1), 61–67.

Kempler, D., Almor, A., Tyler, L.K., Andersen, E.S., & MacDonald, M. (1998b). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language*, *64*(3), 297–316.

Kempler, D., Anderson, E.S., & Henderson, V.W. (1995). Linguistic and attentional contributions to anomia in Alzheimer's disease. *Neuropsychology, Neuropsychiatry, and Behavioral Neurology*, *8*(1), 33–37.

Kempler, K.D., Curtiss, S., & Jackson, C. (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech and Hearing Research*, *30*(3), 343–350.

MacDonald, M.C., & Christiansen, M.H. (in press). Individual differences without working memory: A reply to Just & Carpenter and Waters & Caplan. *Psychological Review*.

Marslen-Wilson, W., & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Marslen-Wilson, W.D., Tyler, L.K., & Koster, C. (1993). Integrative processes in utterance resolution. *Journal of Memory and Language*, *32*(5), 647–666.

McDonald, J.L., & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*(4), 543–566.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

Miyake, A., Carpenter, P., & Just, M. (1994). A capacity approach to syntactic comprehension disorder: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*(6), 671–717.

Nebes, R.D., Boller, F., & Holland, A. (1986). Use of semantic context by patients with Alzheimer's disease. *Psychology and Aging*, *1*(3), 261–269.

Nicol, J.L., Forster, K.I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587.

Osterhout, L., & Mobley, L.A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773.

Pearlmutter, N.J., Garnsey, S.M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.

Rochon, E., Waters, G.S., & Caplan, D. (1994). Sentence comprehension in patients with Alzheimer's disease. *Brain and Language*, *46*(2), 329–349.

Small, J.A., Andersen, E.S., & Kempler, D.K. (1997). Effects of working memory capacity on understanding rate-altered speech. *Aging, Neuropsychology, and Cognition*, *4*(2), 126–139.

Tomoeda, C.K., Bayles, K.A., Boone, D.R. & Kaszniak, A.W. (1990). Speech rate and syntactic complexity effects on the auditory comprehension of Alzheimer patients. *Journal of Communication Disorders*, *23*(2), 151–161.

Tyler, L.K., & Marslen-Wilson, W.D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(6), 683–692.
Waters, G.S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review*, *103*(4), 761–772.
Waters, G.S., & Caplan, D. (1997). Working memory and on-line sentence comprehension in patients with Alzheimer's disease. *Journal of Psycholinguistic Research*, *26*(4), 377–400.
Waters, G.S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, *12*(1), 1–30.

APPENDIX A

List of auditory stimuli used in Experiment 1. The text in parentheses was only included in the long condition. All stimuli were followed by the visual targets WAS and WERE.

1. Artists can be very strange. A famous painter, (who won awards in several recent modern art competitions in many cities)
2. Some people always lose their keys. The old lady, (who searched very carefully through every one of those old trash cans behind the stores)
3. Busy people don't have time to return messages. The top manager, (who left on a long trip through Europe by way of several Asian capitals)
4. Women always look after everyone. The young mother, (who drove to the suburbs of Chicago every morning through heavy traffic jams)
5. The employees were pleased with their new jobs. The friendliest worker, (who spent every day in the narrow park down by the flower gardens)
6. The students are doing well. The young girl, (who improved greatly every day because of the excellent teaching and good books)
7. Some drivers can be reckless. The wild teenager, (who drove like a maniac through the quiet neighbourhood on several occasions)
8. The students learn things very quickly. The smartest child, (who graduated last year in June during the longest heat spell in years)
9. Some people have all the luck. The old man, (who hunted to find a bargain in the smallest, rural towns and villages)
10. The magicians have been practicing for weeks. The newest apprentice, (who learned magic in grade school between studying other subjects)
11. Some people like gardens. The new homeowners, (who lived in the greater metropolitan area in a large rambling house in terrible disrepair)
12. The knitting club was having fun. The new members, (who joined the club last year in the weeks just before Christmas)
13. Most people run errands in a hurry. The busiest customers, (who stopped on the way to work just after the store had opened)
14. Some people are creative. The old men, (who invented a lot of interesting things during and after the Second World War)
15. The family owns a lot of land. The oldest children, (who grew many different kinds of vegetables from seed every summer)
16. Many laborers are working on the building. The best plumbers, (who helped to maintain a large and expensive hotel on the other side of town)
17. Many housewives donate to charities. The generous women, (who had just cleaned out closets and garages in preparation for big sale)
18. The children were very attentive. The best students, (who paid attention to everything in the class and the review session)

19. Parents worry about their children. The nervous fathers, (who went to the hospital's special class about childhood illness)
20. Everyone loves to perform. The young actors, (who wanted to be in a popular musical or dramatic play on Broadway)

APPENDIX B

List of auditory stimuli used in Experiment 2. The text in parentheses was only included in the long condition. All stimuli were followed by the visual targets HIM and THEM.

1. The football players took the lucky boy to the big game. (The stadium was completely packed). At half-time the boy waved to
2. The children loved the silly clown at the party. (The show was very funny.) During the performance the clown threw candy to
3. The sales ladies hated the manager of the store. (The business was poorly run.) Before long, the manager would not even pay
4. The judge gave the criminals life in prison. (The courtroom was very noisy.) After the sentencing, the criminals yelled at
5. The young boy took the seniors through the garden. (The air was cool and crisp.) After the tour, the seniors tipped
6. The young man sent the elderly couple on a vacation cruise. (The holiday was very relaxing.) During the trip, the couple wrote to
7. A group of students interviewed the famous professor at the cafeteria. (The meeting had to be cut short.) Later that day, the professor called
8. The school principal asked the students to stand in straight lines. (The field trip was already behind schedule.) As usual the students ignored
9. The bank manager begged the robbers to return the money. (The situation was nerve wracking.) On the way out, the robbers just laughed at
10. The musicians watched the conductor closely during the concert. (The auditorium was nearly full.) After the performance, the conductor praised
11. The townspeople asked the mayor for financial support. (The city was in a panic.) Without delay, the mayor helped
12. The priest welcomed the visitors to the chapel (The church was beautifully decorated.) After the service, the visitors thanked
13. The patients gave the retiring nurse a farewell party. (The food was low fat and delicious.) After dessert, the nurse hugged
14. The father read the children a bedtime story. (The plot was about a magic castle.) Afterwards the children kissed
15. The coach showed the swimmers around the new gym. (An indoor pool had just been built.) As a joke, the swimmers splashed
16. Mr. and Mrs. Smith called the mechanic early in the morning. (The car would not start.) Later that day the mechanic helped
17. The boss told the janitors to clean the floors. (The hallway was especially dirty.) Reluctantly, the janitors obeyed
18. The travellers listened to the safari guide on the long bus trip. (At times the scenery was boring.) During the dull parts, the guide entertained
19. The referee yelled at the hockey players on the ice rink. (The game was very loud.) At times the players couldn't hear
20. The demanding guests nagged the maid for a freshly washed towel. (The washing machine was broken.) Naturally, the maid avoided