PROVERB AND IDIOM COMPREHENSION IN ALZHEIMER DISEASE

DANIEL KEMPLER, PH.D., DIANA VAN LANCKER, PH.D.,
AND STEPHEN READ, M.D.

Abstract. Twenty-nine patients diagnosed with Probable Alzheimer Disease were administered tests of word, familiar phrases (idioms and proverbs), and novel phrase comprehension. From the early stage of the disease, patients performed worse at understanding familiar phrases than single words or novel phrases. The results uphold common observations that AD patients have difficulty interpreting abstract meanings. Cognitive variables responsible for poor idiom/proverb comprehension and the clinical implications of this new protocol are discussed.

Key Words. Language, Proverb, Idiom, Alzheimer

Language deficits appear early in the course of Alzheimer Disease (AD), and worsen along with other cognitive impairments (Bayles and Kaszniak, 1987; Cummings and Benson, 1983; Obler and Albert, 1981). Although onset and pattern of language disturbance varies across individuals (Wechsler, 1977), anoma is usually an early symptom, seen in poor performances in word-list generation and confrontation naming (Bayles and Tomsa, 1983; Cummings et al., 1985). By the moderate stages, other aphasic symptoms appear, including paraphasias, impaired narrative writing, and comprehension deficits (Appel et al., 1982; Cummings et al., 1985). It is now well established that control over meaning deteriorates prior to ability to use grammatical forms, which is relatively preserved through the moderate stages of the disease (e.g. Bayles, 1979; Kempler, 1984; Kempler et al., 1978; Schwartz et al., 1979). Finally, reiterative disturbances (echolalia and palilalia) characterize the late stages (Cummings, 1982).

Proverb and idiom interpretation, although not yet, to our knowledge, systematically studied in this population, is a standard part of routine mental status testing (e.g. Strub and Black, 1985), and is often included in dementia screening evaluations. AD patients often produce concrete and otherwise inappropriate interpretations of proverbs such as "Rome wasn’t built in a day" and idioms such as "loud tie." An ability to interpret these samples of figurative language is thought to reflect quality of thinking, concept formation, and abstract verbal reasoning (Lezak, 1983). Concrete interpretations are assumed to be symptomatic of loss of abstract ability, and are frequently observed in patients with widespread brain disease of any etiology. In an informal manner, an examiner asks the patient to interpret a proverb or idiom verbally (e.g. Wechsler, 1955), in writing, or in a verbal multiple choice format (Gorham, 1956). The patient responds, and the examiner determines whether the interpretation is abstract ("It takes time to do things well"), semiabstract ("Don’t do things too fast"), or concrete ("It took a long time to build Rome") (Strub and Black, 1985, page 131). If the patient produces one of these responses, the results are relatively easy to interpret.

1 School of Gerontology, University of Southern California, University Park, Los Angeles, CA 90089-0191, USA.
2 Neuropsychology Program, Department of Psychiatry and Biobehavioral Science, UCLA School of Medicine, Los Angeles, CA 90024, USA.
3 Neurobehavioral Unit, West Los Angeles Veterans Administration Medical Center, Los Angeles, CA 90073, USA.
4 Correspondence should be sent to Dr. Daniel Kempler.
This type of testing, however useful, often presents problems in practical application. Verbal responses, particularly from AD patients, often do not fit neatly into one category. For instance, comments and tangential remarks, which are generally not clear attempts at answering the question, are common response types from AD patients. It is often unclear how much to attribute their inadequate responses to comprehension deficits or to production problems, which include paraphasias and word finding difficulties. Responses can be judged to be poor because the patient chose the wrong word, did not respond to the question asked, or because the response is uninterpretable, and not necessarily because of a purported abstraction deficit. Some formats, including verbal multiple choice paradigms, are difficult for AD patients because of poor attention and short-term memory limitations. In addition, for most formats, no reliable age-matched norms are available for comparison, nor are there comparative data from other clinical groups.

Recognizing the widespread use of proverb interpretation in clinical screening and its importance as a measure of cognitive intactness, we designed the Familiar and Novel Language Comprehension (FANL-C) test to evaluate formally the comprehension of proverbs and idioms, in comparison to newly generated sentences. The protocol has several advantages over typical proverb and idiom interpretation tasks. Since only a picture pointing response is required, results are objective, quantifiable, and verbal production abilities are not involved. Thus, deficient performance can be interpreted more clearly, because the properties of the selected stimuli presented for a comprehension task can be specified: Proverbs and idioms are unitary, fixed expressions associated with complex meanings, whereas novel sentences are a sequence of words combined according to productive grammatical rules. Another advantage of using a comprehension test is that performance of clinical groups can be compared to normal controls and the results can be compared directly with comprehension of non-idiomatic (novel) phrases to gauge the contribution of memory, attention, and grammatical abilities on proverb and idiom comprehension.

**Materials and Method**

Three tasks were designed to assess comprehension of single words, familiar phrases\(^4\) (idioms and proverbs), and novel language. Ten nouns, 10 familiar stimuli, and 10 novel stimuli were selected (see Appendix). Single word comprehension was tested to ascertain that the subject could do the task and to establish a baseline. Familiar stimuli met the following criteria: 1) each was a common, overlearned expression having a stereotyped and non-literal meaning and 2) each stimulus could be depicted in a simple line drawing (see Fig. 1). Novel phrases were then created to match the familiar phrases in length, word frequency, and grammatical structure (see Fig. 2). For all 3 tasks, the examiner spoke each stimulus aloud and provided the subject with a response sheet consisting of 4 line drawings. The subjects were encouraged to guess if they were reluctant to choose a single picture; in nearly all cases, a response was elicited.

For the familiar phrases, 3 foils (wrong answers) were designed, such that 1 contained a referential representation of 1 word in the stimulus (i.e. a concrete response), the second was a meaning opposite or related to the idiomatic meaning (a related response), and the third was an irrelevant choice. No literal foils (e.g. a man eating from a woman's hand for “she's got him eating out of her hand”) were included because literal interpretations are, in one sense, correct. As the example illustrates, literal

---

\(^4\) We use the term "familiar phrases" to refer generically to overlearned, holistic expressions, including proverbs, idioms, and social interaction formulae (e.g. greetings). Some of the items are full sentences while others are phrases, but all involve nonliteral interpretation, and all will be referred to as "phrases" throughout this paper.
interpretations of ambiguous familiar phrases are often bizarre or less likely to occur than the figurative meaning. It has been shown that given the choice, listeners reject literal meanings in favor of figurative meanings, based on real world plausibility rather than ability to perceive abstract meanings (Gibbs, 1980; Van Lancker and Canter, 1981). The inclusion of a literal interpretation as a choice would make the results difficult to interpret, since it could not be unequivocally determined whether the literal response was "right" or "wrong." Moreover, we were interested in patients' responses on familiar phrases that do not have a picturable literal interpretation, which is the usual case (e.g. "She took a sudden turn for the worse" and "That's enough to drive a man to drink"). For the novel phrases, all 3 foils were grammatical misreadings such as adjective misassignment or subject-object reversal.

For each subtest, 2 practice items were given. Wrong answers on these items were corrected, correct answers were explained, and practice was repeated until the patient performed the task appropriately. The order of the tasks was always (1) words, (2) familiar phrases, and (3) novel phrases.

The meaning of familiar phrases is derived from associating the phrase as a whole with a single, often complex meaning. On the other hand, the meaning of a novel phrase relies on analysis of word meaning, word order, and grammatical relations. For the familiar phrases, this type of grammatical and single word analysis does not lead to a correct interpretation, but instead, results in a wrong answer. As mentioned above, the familiar and novel phrases were matched in length, grammatical structure, and word frequency. Therefore, the 2 phrase subtests provided a meaning/grammar contrast. That is, they differed only in whether the correct interpretation relied on decoding word meanings and syntactic relations (novels), or on inferring an abstract meaning where knowledge of individual words and grammatical relations was not useful for accurate comprehension (familiars).

Clinical subjects were 29 patients diagnosed with Probable Alzheimer Disease (10 female, 19 male) referred from the Neurobehavioral Unit of the West Los Angeles Veterans Administration Hospital, the UCSD Alzheimer Disease Research Center, and the UCLA Geriatric Outpatient Clinic. Consecutive referrals were taken for the study; subjects were not pre-selected on the basis of language dysfunction or any other criteria. The evaluations were done by clinicians experienced in the evaluation of dementias. Each patient had physical and neurologic examinations, neuropsychological evaluation, laboratory tests (including urinalysis, CBC, thyroid function tests, serologic tests for syphilis, calcium, folate, SMAC panel, and vitamin B12 levels), EEG, EKG, chest X-ray, and CT of the head. The diagnostic evaluation documented a range of cognitive dysfunctions and a progressive history typical of AD, alongside an absence of focal motor, sensory, cerebellar, or cranial nerve defects. Diagnostic criteria were in accord with the recent consensus (McKhann et al., 1984), and in most cases were
Figure 2. Response array for the novel phrase, “He’s sitting deep in the bubbles.”

more restrictive in requiring definite evidence of progressive cognitive dysfunction, and in excluding patients with atypical neurologic or psychiatric features. Five subjects were biopsied at the West Los Angeles VA Medical Center and all 5 were subsequently reclassified as Definite Alzheimer Disease.

Written informed consent was obtained from primary caregivers in all cases. All subjects were monolingual, native speakers of Standard American English (SAE), who were schooled in SAE and used SAE in their home and work. None had any known speech or language pathology prior to the diagnosis of AD. Age ranged from 53 to 87 (mean age = 70.35; SD = 8.8). Performance on the Mini Mental Status Exam (MMSE) (Folstein et al., 1975) ranged from 2 to 28 (of a possible 30). Unfortunately, reliable information about age at onset of illness was not available in enough instances to present that information here.

Control subjects were a group of 43 normal older volunteers with no known history of cognitive impairment (age range: 50–82; mean age = 66.0; SD = 9.8; 13 men and 30 women). Education for both groups ranged from 8th grade to postgraduate degree; most subjects had a high school education, and the groups did not differ in average educational levels. Testing was always completed in 1 session.

RESULTS

The protocols were designed to result in essentially perfect performance from normal subjects. The 43 older normal controls in this study scored 100% on single words, 97.6% correct on familiar phrases, and 99.7% correct on novel phrases.

In general, the AD patients were able to perform the word comprehension task adequately; they obtained a mean 96% correct (SD = 7.0; range 75%–100%). The patients scored significantly better on the novel phrases task (mean 74% correct; SD = 19.0; range 40%–100%), than they did on the familiar phrase subtest (mean 49% correct; SD = 19.0; range 10%–100%; matched pairs t = 7.084; p < 0.01)\(^6\). Performance on single words was significantly better than performance on either

\(^6\) All t-test data are reported with the Bonferroni adjustment to correct for multiplicity of tests: for an overall 0.05 level of significance, each individual p must be less than 0.0125; for 0.01 level of significance, p must be less than 0.0025; for 0.001 level of significance, p must be less than 0.00025.
familiar ($t = 12.75; \ p < 0.001$) or novel phrases ($t = 7.1, \ p < 0.001$). To compare performance on the 3 tasks at different stages of severity, the clinical group was divided into 3 subgroups according to their MMSE scores: mild ($> 20; \ n = 9$), moderate ($15 - 20; \ n = 11$), and severe ($< 15; \ n = 9$) (see Fig. 3). A 2-factor, group x task ANOVA with repeated measures on task (words, familiar and novel phrases) revealed a significant main effect of group ($F(2,26) = 7.834; \ p < 0.01$), significant main effect of task ($F(2,26) = 104.699; \ p < 0.0001$), and a group x task interaction ($F(4,52) = 2.75; \ p < 0.05$). All 3 groups showed the same pattern, performing worst on familiar phrases and best on single words. This pattern held for even the mildly affected patients (MMSE > 20) who showed a substantial deficit in understanding familiar stimuli (mean $= 54\%$ correct; SD = 21.9), compared with novel stimuli (mean $84\%$ correct; SD = 13.3; matched pairs $t = 4.811; \ p < 0.01$). In fact, the discrepancy in performance on novel and familiar phrases generally held across the group: of the 29 AD subjects, 24 performed worse on the familiar phrases, 4 achieved identical scores on both tests, and only 1 performed better on the familiar phrases.

Correlational analyses comparing age, MMSE scores, and scores on the 3 language comprehension tasks yielded several significant correlations: MMSE scores
correlated significantly with performance on novel stimuli \( (r = 0.618; p < 0.001) \) and words \( (r = 0.481; p < 0.01) \); novel phrases and words were significantly correlated \( (r = 0.507; p < 0.01) \); as were novel and familiar phrases \( (r = 0.506; p < 0.01) \). Age, overall, was not correlated with MMSE scores, nor did age correlate with scores on any of the 3 comprehension tasks. Although the mild group was significantly younger than either the moderate or severe group, the pattern of performance and level of significance were constant across groups, suggesting that neither age nor mental status affect the pattern of performance on these 3 language measures. As mentioned above, nearly all subjects performed best on single words, and worst on familiar phrases. When the patients were divided into younger (65 and younger) and older (>65) groups, no significant differences between the 2 age groups were observed in MMSE scores or performance on the language tasks. Similarly, no gender differences were observed on any measure.

Analysis of errors revealed a significant tendency to select concrete responses. Wrong answers on each of the 10 stimuli were pooled from all 29 subjects, and choices of the 3 kinds of foils (concrete, related, and irrelevant) were compared. Of a total of 145 errors, 81% were concrete choices, compared with 11% related choices and 8% irrelevant choices. This ratio was similar for all 3 subgroups, indicating a tendency towards concrete error responses for patients at all levels of severity.

**Discussion and Conclusions**

Using a picture pointing task in a comprehension protocol, we have shown differential dissolution of 3 features of auditory language comprehension in AD. Our results confirm suggestions that dementia patients have particular difficulty understanding abstract language, as observed by their particularly poor performance interpreting idioms and proverbs compared with their relatively better understanding of novel phrases and overall accurate interpretation of single words. This pattern held across mild to severe levels of impairment. When faced with alternative interpretations of familiar phrases, AD patients overwhelmingly chose concrete responses, suggesting that they were using lexical (single word), referential meaning to interpret the phrases. Although comprehension of novel sentences significantly correlated with MMSE scores, neither of the other 2 comprehension tasks (words and familiar phrases) appeared strongly related to overall cognitive functioning as measured by the MMSE. In the remainder of this report we address relevant psycholinguistic issues and clinical implications of these findings.

**Psycholinguistic Issues**

Why do AD patients have trouble understanding familiar phrases? At least 3 nonlinguistic explanations must be addressed: attentional and memory deficits, visual perceptual deficits, and overall level of difficulty associated with the familiar phrase subtest. In the tasks reported here, the novel phrases and the familiar phrases were matched in length (as well as word frequency and superficial grammatical structure). Therefore, significantly worse performance on the familiar phrase subtest cannot be attributed to a general memory or attention deficit, which would be expected to affect both tasks equally. Second, since poor word comprehension has been attributed to visual perceptual deficits (Kirshner et al., 1984; Rochford, 1971) it is conceivable that our results could be at least partially explained by difficulty in interpreting the pictures which accompany the familiar phrases. This is unlikely, as the line drawings for both novel and familiar phrases
were of similar quality and complexity (see Figs. 1 and 2). Third, it is tempting to invoke a difficulty metric, whereby familiar phrases could be seen as inherently more difficult to comprehend than novel phrases. If this were true, we would expect to see an effect of normal aging on the task, and a similar effect in all cases of brain injury, but we find neither. Normal controls performed nearly perfectly on both tasks, regardless of age, and age of the clinical subjects was not correlated with performance on the familiar phrase task ($r = -0.075$). Further, another brain injured population (left hemisphere damaged patients) demonstrated a reversed pattern, in that they understood familiar better than novel phrases (see below and Van Lancker and Kempler, 1987).

To understand what makes the familiar phrases psychologically different from the novel phrases, we must look at the processes involved in comprehension of the 2 sentence types. Psycholinguistic research with normal subjects indicates that, whereas the meaning of a novel sentence is derived from a sequential analysis of constituent parts, successful recognition of familiar phrases involves perception of an overall pattern (Gibbs, 1980; Horowitz and Manelis, 1973; Lieberman, 1963; Osgood and Hoosain, 1974; Simon, 1974). These studies have shown, for example, that normal subjects can identify and correctly interpret idiomatic phrases more quickly than novel phrases (Gibbs, 1980; Swinney and Cutler, 1979). Moreover, normal subjects appear not to remember the individual words that make up idiomatic phrases, whereas they do remember words constituting comparable novel phrases (Horowitz and Manelis, 1973; Osgood and Hoosain, 1974).

Further evidence comes from digit span studies using linguistic units, showing that subjects retain in memory almost as many familiar phrases as words (5–9). These results indicate that familiar phrases are psychological “chunks” similar to words, and that they are perceived and understood as wholes, without analysis of constituent parts (e.g. individual words) taking place at all. The data suggest that people do not process idioms by first processing individual words and attempting to analyze their literal meaning, and only then deriving the idiomatic meaning. Precisely how grammatical analysis is bypassed is not well understood, but it is possible that, at least in part, single unitary intonation envelopes which accompany familiar but not literal phrases (Van Lancker et al., 1981) serve as acoustic cues to initiate holistic rather than grammatical processing. Further, in order to understand a familiar phrase, one must be able to associate the overall entity with a unique meaning (Peters, 1983). It is likely that AD patients have difficulty either with recognizing the overall pattern of familiar phrases and/or with accessing the associated complex meaning. Previous research with AD patients lends support to both of these possibilities.

Regarding the first (overall perception of a unitary phrase), there is reason from other observations to believe that AD patients are impaired in apprehehding overall patterns. They are known to have difficulty perceiving and reproducing both nonsymbolic visual patterns such as geometric forms (Cummings and Benson, 1983) and symbolic, communicative gestures (Kempler, 1984, in press). In both these areas, it is notable that AD patients can often reproduce some of the details but not the overall Gestalt or pattern. These findings suggest that AD patients may demonstrate a generalized impairment in what has been variously called holistic, integrative, global, or synthetic processing, and has been associated with right hemisphere specialization (Bever, 1973; Bogen, 1969a, b; Bradshaw and Nettleton, 1983; Bryden, 1982; Lezak, 1983). We would expect such a general deficit to affect perception and interpretation of familiar phrases, which also, by definition, have holistic form and integrated meanings.
AD patients are also known to have difficulty accessing meanings. The finding that AD patients selected significantly more concrete than related or irrelevant wrong answers lends empirical support to the popular belief that dementia diminishes the ability to process abstract meanings. While this is true, we prefer a more operationally definable formulation of this notion. Recall that a primary difference between familiar and novel phrases is that familiars involve associating a fixed form with a complex and integrated meaning, while the novels require sequential and grammatical analysis to derive a meaning. Since AD patients are known to have particular trouble accessing meaning (e.g. Bayles, 1982; Bayles and Kaszniak, 1987), we suggest that the deficit observed on the familiar phrase subtest is related to the integrated, complex meanings associated with these phrases. Since comprehension of familiar phrases requires the retrieval of complex meanings, and does not rely on syntactic abilities, it is understandable that idiom and proverb interpretation would be disproportionately compromised in AD. The relative preservation of grammatical abilities may actually defeat the AD subjects in this task, because correct interpretation of familiar phrases requires non-utilization of these very grammatical mechanisms which are best preserved (Gibbs, 1980; Kempler et al., 1987). In summary, we propose a tentative and necessarily speculative explanation for the proverb and idiom deficit, whereby AD patients have difficulty first perceiving the overall pattern of the familiar phrase and second, matching that phrase to its meaning.

Clinical Issues

Language profiles of AD patients have been compared with language profiles associated with focal aphasias (Appel et al., 1982; Cummings et al., 1985). The data presented here, taken together with comparisons of left and right brain damaged patients on this same task (Van Lancker and Kempler, 1987), indicate specific differences between the language abilities of patients with focal aphasias and those with AD. Van Lancker and Kempler studied 39 consecutively available, right handed patients with documented unilateral focal brain damage of vascular origin, with the same protocol and procedure described here. All were right handed, native English speakers (5 women, 34 men; mean age 62.8; range 45–80) with between 10 and 19 years of education. Twenty-nine of the patients were left brain damaged and aphasic (mean age 62.3; average 63.5 months post onset). Aphasia type was diagnosed by a speech-language clinician using standardized instruments: 8 patients had fluent aphasia, 10 were non-fluent, 3 were globally aphasic, 5 were amnestic, and 2 were mixed. Eleven subjects were right brain damaged (mean age 63.4; average 19.5 months post onset). The aphasic patients were able to understand familiar phrases significantly better than they understood novel phrases, a pattern opposite from that observed with AD patients. In contrast, the AD patients performed like the right hemisphere damaged patients insofar as they were able to process and recognize novel phrases significantly better than familiar phrases (see also Benton, 1968; Myers and Linebaugh, 1981; Winner and Gardner, 1977). Figure 4 shows the performance of 28 left hemisphere aphasic patients, 11 right hemisphere damaged patients and 29 AD patients on the familiar and novel phrase subtests. The similarity between AD patients and right hemisphere patients may reflect the necessity of a functionally intact right hemisphere for successful completion of the familiar phrase task. The qualitatively different performances of left hemisphere damaged aphasic and AD patients suggests that the “aphasia” associated with AD is distinct from that associated with focal left hemisphere.
lesions, despite earlier reports of the similarities (Appel et al., 1982; Cummings et al., 1985).

Early identification of AD is crucial for developing and implementing treatments which will prevent and relieve the effects of the disease. Since much research on AD is plagued by problems of accurate diagnosis, it is notable that the accuracy of diagnosis in our subjects can be expected to be high. Recent experience with biopsy (Harbaugh, personal communication) suggests that restrictive clinical criteria such as those used in the present study yield an accuracy of diagnosis approaching 95%. Our own data confirming that the 5 patients who underwent cortical biopsy (including 2 mildly affected patients with MMSE scores of 21 and 28) were accurately diagnosed as Alzheimer patients further validates our diagnostic procedure. Although several techniques have been proposed for early differential diagnosis (e.g., Jarvik et al., 1982), few behavioral tests have been successful in this regard. Language deficits have held particular promise, insofar as they have been present in mildly demented cases (Huff and Growdon, 1986) and they have been linked to clinical variables such as age at onset, family history, and rate of progression of AD (Filley et al., 1986; Folstein and Breitner, 1982; Knesevich et al., 1985). Although some patients present clear deficits in word-list generation early on in the course of the disease (Cummings et al., 1985), this symptom does not differentiate them from patients with aphasia due to focal left hemisphere damage. The Familiar and Novel Language Comprehension test described here seems to highlight an early-appearing, subtle intellectual defect which
may appear before other, more overt aphasic symptoms such as anomia, paraphasias and poor word-list generation. If the test continues to distinguish AD patients from those with aphasias due to focal lesions, and continues to be sensitive to early cognitive deficits, it may be beneficially added to diagnostic and neuropsychological batteries.

ACKNOWLEDGMENTS

Illustrations were drawn by Susan Black. We thank Catherine O’Connor and Cathy Cornelius for assistance with testing and Wendy Mack for statistical consulting. We appreciate the cooperation of the Mahood Senior Center of Santa Monica for allowing us to recruit normal volunteers from among their patrons. This work was supported in part by The UCSD Alzheimer Disease Research Center (NIA #PS0-AGO-5131, Robert Katzman, P.I.). We are grateful to Drs. J. Cummings and S. Y. Tsai of the West L.A. VA Medical Center, Drs. W. Ashford and J. Schaeffer of the UCLA Geriatric Outpatient Clinic, and The UCSD Alzheimer Disease Research Center for patient referrals.

REFERENCES

Kempler D: Lexical and pantomime abilities in Alzheimer’s Disease. Aphasiology, in press.


**APPENDIX**

**Single Words**

1. cigarette
2. iron
3. spoon
4. umbrella
5. baby
6. car
7. hat
8. pen
9. sink
10. drum

**Familiar Phrases**

1. While the cat's away, the mice will play.
2. I'd like to give you a piece of my mind.
3. He's saving up for a rainy day.
4. The truth, the whole truth and nothing but the truth.
5. She took a sudden turn for the worse.
6. He's turning over a new leaf.
7. That's enough to drive a man to drink.
8. She's got him eating out of her hand.
9. It's like talking to a brick wall.
10. He's living high on the hog.
**Novel Phrases**

1. When the happy girl pushes, the angry boy swings.
2. The clown likes to pull the girl on the wagon.
3. He’s taking her for a little ride.
4. The clown, the small clown, is not in front of the girl.
5. She’ll push her brother to the ground.
6. He sits there looking away.
7. The happy girl has a wagon to pull.
8. She’s glad to be giving him a wagon ride.
9. She likes being in the wooden swing.
10. He’s sitting deep in the bubbles.