In the two major sections of this paper the influence of John B. Carroll and Robert Lado on the teaching and testing of listening comprehension since 1961 is examined. The first section reviews the literature of that field and outlines four major periods roughly corresponding to the methodological orientation dominating professional thinking at the time. They include: the era of discretion (1961-1968), in which language learning was viewed as a set of empirical phenomena reducible to manageable and definable segments; the era of cognition (1969-73), which stressed the learner's understanding of what he was doing as he practiced elements of the target language; the threshold era (1974-81), which emphasized the functions of language in realistic situations; and the proficiency era (1982-1986), which focused on what speakers actually do with language in various specific situations, for various reasons, with various degrees of accuracy. The second section proposes and discusses the immediate recall protocol, a technique for listening comprehension test item development which is based on the testing insights gained during that 25-year period and which seems to focus the principles elaborated by Carroll and Lado. The protocol involves having learners listen to a foreign language text and write down, in their native language, everything that they can remember about the text. A list of references, sample texts and test items, and a selective bibliography are appended. (MSE)
FROM PRINCIPLES TO PROTOCOLS: THE CONTRIBUTIONS OF JOHN B. CARROLL AND ROBERT LADO TO THE TESTING OF LISTENING COMPREHENSION

Charles J. James

Paper presented at the Language and Testing Symposium, Quiryat Anavim, Israel May 11-13 1986
Charles J. James
University of Wisconsin
Madison, Wisconsin 53706 USA
"From Principles to Protocols: The Contributions of John B. Carroll and Robert Lado to the Testing of Listening Comprehension." 1

Introduction

In 1986 the language teaching profession honors two colleagues who have exerted considerable influence on the direction of language testing in the past 25 years, Robert Lado and John B. Carroll. The principles they set down have guided the language teaching profession in positive ways since the publication of the first edition of Robert Lado's book Language Testing: The Construction and Use of Foreign Language Tests and the first public presentation of John B. Carroll's paper "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students," both in 1961. If one were to review their works alone, a paper of considerable size would emerge. What this paper will attempt to do, however, is to look back over the past 25 years of testing in one specific area, listening comprehension, touching upon the works of Carroll and Lado, as well as the work of others who have been directly and indirectly influenced by them.

1 I am indebted to John H.A.L. de Jong for his critical but balanced review of early drafts of this paper.
The paper has two major sections: 1) a review of the literature on the teaching and testing of listening comprehension, and 2) a preview of a technique for assisting in the development of listening comprehension test items which combines many of the insights gathered from the research into listening comprehension during the past quarter century.

A number of articles outline various aspects of the nature of listening comprehension, and should be referred to, including Brown and Yule (1983), Byrnes (1984, 1985), Call (1986), Nagle and Sanders (1986), Richards (1983, 1985), and Rivers (1986). Five books are also worthy of note: 1) the anthology by Duker (1966) and 2) the monograph by Lundsteen (1971), both for the research and practical aspects of teaching listening comprehension in native speaker learning environments; 3) the handbook by Ur (1984), 4) the special edition of *Foreign Language Annals* on the receptive skills in language teaching (Phillips & Omaggio, 1984), and 5) the proceedings of a conference held in Bielefeld (Schumann, Vogel & Voss, 1984), all three for the teaching of listening comprehension in target language learning environments. In addition, since 1968 the American Council on the Teaching of Foreign Languages (ACTFL) has produced an annual volume on topics of interest to the language teaching profession.

---

2 Throughout this paper the term for a person's first or primary language (or "mother tongue") will be "native." For any other language, whether "second" or "foreign", the term will be "target."
Each of the sixteen volumes that have appeared to date contain chapters that treat one or more aspects of language testing. A number offer practical suggestions for the testing of listening comprehension. In the chronological bibliography at the end of this paper are listed those references in the ACTFL annual volumes pertaining specifically to listening comprehension testing.

PRINCIPLES

In the past 25 years listening comprehension has gone from a set of skills that was paid "ear service" but rarely emphasized in practical classroom language teaching to a major focus of attention in most target language curricula. The fortunes of listening comprehension have risen and fallen with the major emphases in language methodology, especially in the United States, which since the 1960s have seen at least four shifts in attitude towards the teaching and learning of languages. The history of listening comprehension testing since 1961 can thus be divided into four major "Eras," roughly corresponding to the methodological orientation that dominated professional thinking at the time: 1961 - 1968: The Era of Discretion, 1969 - 1973: The Era of Cognition, 1974 - 1981: The Threshold Era, 1982 - 1986: The Proficiency Era. The division is based upon the appearance of one or more documents which shifted the focus of the profession in a seemingly "new" direction, although the desire to trade in an old bandwagon was frequently the reason for the shift.
Dividing 25 years into eras requires one additional, but emphatic caveat: work in language teaching and testing has always been open to all interested colleagues and hotly debated at all levels of instruction. As a result the process has been more continuous and less disjointed than it might appear. For example, there were strong advocates of proficiency in the 1980s meaning of the word, even at the height of audiolingualism in the mid 1960s, including Carroll and Lado. Factors beyond the control of teachers and testers, such as the lack of adequate electronic equipment, often prevented us from doing what we have always known needed to be done. Each Era, then, has been an age of compromise and concession, not an age of narrow vision.

1961 - 1968: The Era of Discretion

Although listening was the first of the skill areas in the audiolingual model of "Listening - Speaking - Reading - Writing," which dominated the professional literature around 1961, practical considerations often restricted the implementation of listening comprehension exercises beyond those of imitation of the teacher's voice in pattern drills or the transcription of teacher-made texts in dictations. The reasons for the limitation of meaningful listening comprehension exercises include: 1) the availability of native speaker models was restricted, 2) non-native speakers were often reticent about offering models that might have been deficient phonologically and structurally, 3) audiolingual teaching emphasized the oral (re)production of phonologically accurate patterns, not the (re)creation of
realistic situations, and 4) the technology of the 1960s made it difficult to bring material into the average classroom, especially at the secondary school level.

The era beginning with the early 1960s was dominated methodologically by audiolingualism, in turn dominated by structural linguistics and behavioral psychology. It was an era when language learning was viewed as a set of empirical phenomena reducible to definable and manageable segments. Holistic or global learning was never denied; yet was never emphasized. It was better to be discrete, and not talk about things that were larger than the sentence. Multiple choice testing represented the height of discretion. Although multiple-choice discrete-point testing had been use in the United States as far back as the early 1920s, when they were introduced as "new-type" tests, "scored by a special staff of clerks under expert supervision" (Cole, 1931: 391). Even then it was recognized that multiple-choice tests was not well suited to determining active or functional knowledge of a language (1931: 424-28). Nonetheless, the 1960s saw the development and implementation of test batteries based almost exclusively upon discrete-point testing procedures.

Listening comprehension was no exception. The Carroll-Sapon Modern Language Aptitude Test, Form A (1959) and the Pimsleur Language Aptitude Battery (1966), for example, included parts which attempted to tap a learner's ability to distinguish between elements in spoken minimal pairs, both in isolation and in longer
utterances. The Modern Language Association Tests (the MLA Cooperative Foreign Language Tests, Forms LA and LB for pre-secondary, Forms MA and MB for secondary, and the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students (later retitled MLA Cooperative Foreign Language Proficiency Tests, Forms HA, HB, and HC for university) picked this up to a lesser extent, but relied just the same upon a learner's ability to separate out the sounds of the target language as a prerequisite to comprehension of longer utterances in meaningful situations (MLA, 1962, 1964). In addition, the MLA Tests took everyday situations and broke them down into segments of usually no more than one or two short sentences, which required the examinee to recognize which of three or four responses was the most appropriate rejoinder to an utterance spoken on a tape (Clark, 1972).

The test batteries just described were, however, not readily available to the classroom teacher, and only reinforced the teacher's preconception that language teaching and testing involved breaking down language into small segments. It thus comes as no surprise that the primary testing technique for listening comprehension in most classrooms was, and in some cases remains, however, the dictation. At a time when electronics was beyond the grasp of all but a few enthusiastic hardware specialists and their allies in audio-visual departments, most teachers did not consistently use the listening materials provided for most textbooks. When they did, they discovered that
the material was heavily single-item oriented, be it through the imitation of dialogues or the insertion of words and phrases into blanks. The dictation, however, afforded the classroom teacher a technique for measuring a learner's perception of and reproduction of the words and sounds of the target language without the elaborate apparatus found in the language laboratories of the day. A number of articles have been written about dictation, some old and venerable (Cole, 1931: 339-42), some quite enthusiastic (Oller and Streiff, 1975), others more balanced in their analysis (Freudenstein, 1970). The simplest statement of principle on the value of the dictation comes from Lado: "Dictation is not a bad technique and should not be abandoned, but there are better ways to test" (1964: 161).

The Era of Discretion ended quietly in the United States with the publication of the results from the so-called "Pennsylvania Project," a large-scale attempt to contrast audiolingual methodology with methodologies that had preceded it, called "traditional" or "grammar-translation methodologies." Issues of Modern Language Journal (1969) and Foreign Language Annals (1969) devoted most of their pagination to various aspects of the project, and should be referred to. I say "methodologies" (plural) because we now realize that good teaching had always emphasized more than just the rote memorization or imitation of language models. For listening comprehension, all of this meant that the more mechanical aspects of mimicry-memorization, with the prelude to mimicry being the precise attention to the
phonological input of the message regardless of the informational content, could draw upon insights into the nature of human language use that scholars like Carroll and Lado had always emphasized were present and should be fostered.

Begin the Era of Cognition.

1969 - 1973: The Era of Cognition

Audiolingualism never pretended to do more than teach people how to produce language accurately, assuming that accuracy was the prerequisite for any other language use. Nonetheless a movement began in the late 1960s against the more mechanistic approaches to language learning. In 1961, at the beginning of the Era of Discretion, Carroll had already warned against over-enthusiasm for discrete-point teaching and testing:

If we limit ourselves to testing only one point at a time, more time is ordinarily allowed for reflection than would occur in a normal communication situation, no matter how rapidly the discrete items are presented. For this reason I recommend tests in which there is less attention paid to specific structure-points or lexicon than to the total communicative effect on an utterance. (1961: 37)

The movement was labeled with the term "cognitive code learning," a term credited to Carroll in 1964 (Chastain, 1971: 77). In essence the cognitive code principle (Carroll never referred to it as a "method!") required learners and teachers to know and understand what they were doing while they were practicing the elements of the target language. For some this
meant returning to grammar-translation techniques. For others it meant explaining more and more about the target language in the native language. For most it meant the beginning of a heated debate as to the best means to teach a target language, rather than a concerted effort to address the principal (principle!) issues raised by Carroll in 1961.

The best summary of the points of debate between audiolingualism and cognitive code is found in Chastain's methodology text (1971, 1976). In it, he outlines the psycholinguistic and methodological differences that would be exhibited in the classroom and in the learner when confronted with the one or the other type of learning environment. Neither Chastain nor anyone else questioned the validity of the model "Listening - Speaking - Reading - Writing," frequently abbreviated as the "Four Skills Approach" to language learning.

For listening the late 1960s and early 1970s the attempt was made to balance testing of the three components that Chastain outlined in the 1971 edition of his book: 1) Sound Discrimination, 2) Auditory Memory, and 3) Comprehension. Procedures for teaching and testing the first component did not need much change; they were basic to pure audiolingualism. The second component emphasized expansion of test material from single sounds, word, and phrases, to longer utterances, so as to give learners confidence in listening to longer, although still structured material: "As the students progress through the course, their ears become attuned to the second-language sound
system and auditory memory expands" (1971: 170). Comprehension, the third component, seems to be a factor primarily in what Chastain labels "advanced" listening, since the first two, sound discrimination and auditory memory, take up substantial amounts of time in light of the other skills (speaking, reading, writing) to be developed. Chastain expresses it this way:

At a more advanced stage and after much practice, the students will be able to concentrate on the content of familiar dialogues ... without being consciously aware of the manner of expression. ... In sound discrimination, words were used. In auditory memory, phrases and sentences were employed. Now, short conversations and oral reading as well become the bases for developing greater listening comprehension. (1971: 171)

Keep in mind that the availability of material on tape or on records for listening comprehension in the early 1970s was restricted to that produced for target language textbooks, and thus reflected the methodological prejudices of the authors and/or publishers. As a result, most of Chastain's suggestions for listening comprehension come from the classroom:

The students themselves can provide each other with a considerable amount of listening comprehension practice ... Listening to classmates stimulates motivation and provides an opportunity to relate to a live situation (1971: 172). Otherwise, realistic material for listening comprehension must be provided by teachers themselves, from radio broadcasts, movies,
records, and personal experiences, which teachers are urged to share with one another. Testing continued to include dictations, multiple-choice items, and similar discrete-point oriented techniques, but a genuine attempt was made to make listening tasks as realistic and communicative as possible, as long as the various stages in the development of language learning was maintained.

The reader familiar with Chastain's work will know that in the 1976 edition, he added three more components to the listening comprehension process: "perception of the message" (inserted between discrimination and auditory memory), "comprehension in the first [native] language," and "comprehension in the second [target] language" (1976: 289, 291-93). All five components together represent a refinement in a continuum of processes for which test items can be constructed. Refinement in the model of listening comprehension also meant a differentiation in the types of test tasks demanded of the learner.

1974 - 1981: The Threshold Era

If the Era of Discretion emphasized minimal pairs, recognition of individual words and phrases, and memorization of language patterns, and if the Era of Cognition emphasized the decoding of longer utterances, and if both relied upon multiple choice testing formats, then the Threshold Era emphasized the functions of language in realistic situations. Another name for the era might be the "Era of Communicative Competence." Many colleagues would almost automatically use the latter term in
reference to this supposed "successor" to audiolingualism and cognitive code learning. However, the appearance of a single document in 1974, and with it two interrelated terms, dominates the professional literature: The Threshold Level, published by the Council of Europe, first for English instruction (van Ek, 1975), then for French (Coste et al., 1976), German (Baldegger, Müller, Schneider, 1980), and Spanish (Slagter, 1979). With its concepts of "notions" (semantic categories such as "quantity," "time," "characteristics," and "space") and "functions" (speech act categories such as "exchange of information," "expression of emotional [and] intellectual attitudes," "operation of social conventions," and "performance of tasks"), language educators concerned themselves more and more with the communicative aspects of language, i.e., what people do with language, not just its segments. The communicative intent of listening comprehension exercises and test was thus emphasized, that is, how people actually deal with incoming oral (aural) language and how this is manifested in the structures and vocabulary they need when listening to target language material.

The Threshold Era did not foster a particular methodology or develop its own testing programs. Indeed, the term heard most often during the 1970s was "eclectic," which in essence meant that each teacher did whatever worked in his or her classroom on a day to day basis. Nonetheless during the Threshold Era a number of different methodologies arose. Some of these have listening comprehension as a focal point, including Total
Physical Response (Asher, 1969, 1981; Asher, Kusudo, de la Torre, 1974), which emphasizes listening and acting out basic tasks; Suggestopedia (Lozanov, 1978), which uses music and melody to help the learner relax in order to reinforce comprehension of the target language; and Community Language Learning (Curran, 1976), which puts the teacher in the role of "knower/counselor" who translates learners' native language utterances by whispering the target language equivalents softly into the learner's ear.

The work by Winitz et al. (1981), summarized in the term "Comprehension Approach," stresses the necessity of long periods of listening before the other skills, especially speaking, are introduced. The testing of comprehension takes the form of pointing to or drawing pictures, or reacting in the native language in various ways.

What seems to characterize the Threshold Era best is an emphasis on the actual uses to which language is put, differentiated across the four basic skill areas. Where the so-called "four-skills approach" definitely dominated the two preceding eras, from approximately 1973 onwards there arose the realization that language operates in several dimensions at once, which previous learning models may not have adequately addressed. In addition, the increase in language for special purposes has made it imperative to examine the basic rationale for putting two or more skills together in a learning program. While it may be appropriate to present German or French or Spanish or ESL as a uniform-looking subject-matter package in formal instructional
settings, in the real world the need for particular skills becomes highly differentiated. Thus, the four primary skill areas (listening, speaking, reading, writing) are now treated as separate from one another, depending upon the professional and personal needs of the "customers" of the language teaching "business," be they university students or business executives or travel agents.

In their book on teaching the spoken language, Brown and Yule begin the chapter on teaching listening comprehension with the following observation:

In the last ten years [since 1973] listening comprehension has begun to be taken seriously. Previously, where there was any interest at all, it seemed to be assumed that the student would just pick it up somehow in the general process of learning the foreign language. ... Sadly, this apparently natural process doesn't seem to produce the desired results (1983: 54).

In other words, testing listening comprehension is not easy. Native speakers do not normally speak slowly or clearly, yet students are most often taught to speak slowly and clearly by teachers who always address them slowly and clearly. What this tells us, of course, is that listening comprehension during the Era of Discretion and the Era of Cognition was dominated by overly correct, sanitized language which students simply "got used to" so much so that they were incapable of understanding native speakers on their home turf. The value of work such as
that done by the Council of Europe is the reintroduction of realistic situations which encouraged learners to deal with language outside of the artificial constraints of the studio and the language laboratory.

1982 - 1986: The Proficiency Era

Proficiency is not a new concept. The MLA Tests of the 1960s used the term as part of their titles. Both Carroll and Lado used the term in outlining testing procedures. However, since 1982 it has taken on a very specific meaning which focuses more clearly what previous discussions only suggested. Briefly stated, proficiency reflects what speakers (or listeners or readers or writers) actually do with language in a variety of situations, under a variety of circumstances, for a variety of reasons, with varying degrees of accuracy. The "variety" must, however, be specified and verified.

The history of proficiency testing in the United States begins with the Foreign Service Institute's oral proficiency interview in the 1950s. Since 1980 the FSI and other governmental agencies have operated under the rubric of the "Interagency Language Roundtable" (ILR). Under the leadership of the American Council on the Teaching of Foreign Languages (ACTFL) and the Educational Testing Service (ETS), the ILR rating scales have been adapted for non-governmental, "civilian" use. In 1982 ACTFL published the ACTFL Provisional Proficiency Guidelines, which it revised in 1985, dropping the term "provisional" (ACTFL, 1982, 1985).
Although the techniques for rating listening comprehension proficiency are not yet at the same stage of development as those for speaking, there are a number of techniques under discussion for classroom teaching and testing of listening comprehension with a proficiency orientation. The best sources to date are two books by Omaggio (1983, 1986), in which she outlines a number of listening comprehension proficiency test formats, all of them situational in nature, which include listening for particular verb forms present in a conversation overheard in a cafe, or taking notes on information needed in the search for an apartment, or true/false or multiple-choice questions about a university lecture, or a native-language summary of a narrative (related to the immediate recall protocol to be described later), or global understanding of a commercial advertisement, etc.

Summary

Summarizing the development of listening comprehension testing over the past 25 years, we see the evolution of techniques ranging from those which segment the audio stream into minimal pairs or into individual sounds to be reproduced in a written transcription (dictation) or individual situations to which the learner is expected to react through pictures or actions, to those techniques which attempt to bring whole texts into play with all the segmental and suprasegmental information available for listener processing. The primary contribution of Carroll to the discussion is the emphasis on the fact that listening consists of a number of interrelated processes and
skills which are difficult to test separately (at least with the instruments available during the 1960s and 1970s). The teaching of listening is, as Carroll phrases it, "a matter of training of processes that lead the individual to pay closer attention to what he hears and to organize meanings for better retention, comparison, and inference" (1971: 131). The primary contribution of Lado to the discussion is the development of basic test techniques for measuring listening comprehension, but with many words of caution about overuse of any one of them in the absence of an understanding of the complex nature of listening. His caveat about dictation has already been repeated. In the discussion to follow, his seven "cautions" regarding the development of listening comprehension test items will be discussed in more detail.

PROTOCOLS

With this historical preface, I would like to discuss a technique which seems to focus the principles that Carroll and Lado have elaborated with regard to listening comprehension: the immediate recall protocol. I believe that the immediate recall protocol may prove to be one of the best means for determining the extent of listening comprehension of a target language text. It involves asking a learner to listen to a target language text, and then writing down, in the native language, everything that the learner can recall about the text.

In its purest form, what the learner writes will be in complete sentences, however a given language defines complete
sentences. The text is read aloud or played electronically only once. Learners are not allowed to take notes as they listen. There are also no external stimuli such as pictures or charts. The protocol is then compared with a transcript of the original text and scored on the basis of recall of details, called "idea units," in the text. The segmentation into "idea units" can be downward, that is, at the word, phrase, or clause level, depending on the content of the text. The text can, however, also be segmented upwards, that is, certain dominant, inferential ideas may be separated out as additional units. In practice, however, this latter process can usually be subsumed under the first.

The reasons for my enthusiasm about the immediate recall protocol are as follows: 1) the offset hypothesis, as described by Lowe, indicates that the receptive language skills develop more rapidly in a target language than the productive skills (1984); 2) it is usually easier to express oneself in the native language than in a target language (Ostyn and Godin, 1985); 3) protocols can help reveal the processes underlying the organization of texts, not just those underlying the retention of specific details (Meyer, 1975); 4) protocols can assist in the development of other types of test items (Lange, 1986). The immediate recall protocol seems to reflect what most people would perceive as a realistic method for expressing comprehension of a target language text (James, 1986).

There are a number of disadvantages inherent in the
immediate recall protocol: 1) if the teacher does not know the learner's native language, it will obviously be difficult to evaluate his/her protocol; 2) even granting knowledge of the learner's native language, it may be difficult to determine the quality of his/her native writing language skills; 3) confusion can arise in the student between a protocol, which asks for everything that the learner recalls, versus a summary, which asks for only the salient details, thus potentially weakening the reliability of the instrument; 4) especially in the listening modality the length of time available for efficient processing of different kinds of spoken material varies considerably, although the optimal time is probably relatively short, making longer texts, such as plays and lectures, impractical for this technique. One further difficulty, albeit a minor one, arises from the potential for transcription, that is, learners writing down verbatim everything they hear. Experience tells me, however, that this is unlikely simply because few people possess a "phonogrammic" memory, permitting them to retain in and replay from their short-term memories what they hear from a recording or a live situation verbatim. Adding the dimension of translation from the target language to the native language would only increase the burden on the short-term memory.

How does the immediate recall protocol fit into the history of language testing, especially into the history of listening comprehension testing? The term is found as early as 1953, in, for example, the Brown-Carlsen test, designed to test the global
understanding of continuous discourse in children in grades 9 to 13. One part is called "Immediate Recall," but it tended to test memory for numbers and serial orderings, that is, highly specialized kinds of recall procedures (Carroll, 1971: 128). The research on the immediate recall protocol in its present sense was originally carried out by Meyer (1975) to determine native language processing of reading texts. Her work in the area has been replicated for teaching target languages, specifically German, by Bernhardt (1983, 1986).

Lado (1979) contributed indirectly to the discussion of protocols in a study which compared the translation of written texts (from Spanish to English) when the original (Spanish) text could be read and referred to directly ("immediate translation"), and when the text had to be recalled and translated from memory. Among other things he determined that those learners who translated with the text in front of them made more errors than those who read and studied the text, then wrote a summary. This led him to conclude that "translation with text in view induces significantly more errors [in] unskilled translators, i.e. students." Of greater significance for the use of the immediate recall protocol is his conclusion that

in recall tasks across languages Ss process language into ideas which are stored in deep memory and recalled to construct utterances in a process that approximates full linguistic performance for communication (1979: 571). In other words, learners, especially those in the undergraduate...
population at an American university, can process texts in their native language in ways that reveal just how much and how well they comprehend. The more precise decoding processes required for translation are usually underdeveloped in such a population, and may distort the extent of comprehension present.

The immediate recall protocol technique described here has a number of immediate precursors. Chastain, in the second edition of his methodology text, supported comprehension processes in the native language as a necessary preparation for comprehension in the target language (1976: 291). Omaggio, in the first of two books on the subject of proficiency teaching and testing, described the "native language summary" (1983: 22/23) as a realistic and reliable technique. This was based in turn upon a testing and scoring procedure described by Schulz and Bartz (1975: 85/86), with primary difference being that the text is played twice, only factual information is demanded, and complete sentences are not required.

The renewed interest in recent years in listening comprehension (Brown and Yule, 1983, Rivers 1986, Ur 1984, Winitz 1981) has already been mentioned. It is now possible to resurrect Newmark's and Diller's plea for having students spend more of their time listening to natural speech and authentic models of the foreign language and the need for ... the systematic development of listening comprehension not only as a foundation for speaking, but also as a skill in its own right. (1964)
For those of us who teach German there is a substantial amount of well edited, pedagogically sound material available for training listening comprehension, from agencies as Inter Nationes in Bonn and from publishers such as Hueber in Munich and Heinle & Heinle in Boston. Put another way, we no longer have any technical excuses for not increasing the amount and the variety of listening material in our classrooms. There are too many walkmen around to ignore.

In his article "Fundamental considerations ..." Carroll makes the following statement:

The four skills of listening, speaking, reading, and writing must ... be regarded as integrated performances which call upon the candidate's mastery of the language as a whole, i.e., its phonology, structures, and lexicon (1965: 317/18). He continues by acknowledging that it is possible to test very specific items of language knowledge and skill. He then says:

I do not think, however, that language testing is complete without the use of the approach ... requiring an integrated, facile performance on the part of the examine (1965: 318). From his own research of the time he relates that he had excellent success in ascertaining levels of audio-lingual training by a listening comprehension test in which aurally-presented sentences of increasing length and rapidity are to be matched with an appropriate picture out of four presented. The examine is not concerned with specific structure-points or lexicon, but with the total meaning of the
sentence, however he is able to grasp it. (1965: 318) The last statement underscores the fact that each individual perceives incoming language material individually. This individualism may be simply an idiosyncratic alignment of otherwise formalistic subroutines, or it may involve operating at several levels of consciousness and unconsciousness simultaneously that we do not as yet understand. In any event, comprehension involves total meaning and individual reaction to that total meaning. It is holistic and idiosyncratic.

Richards expands on this thought in a recent publication. The primary contribution to our discussion here is his observation that

Listeners make use of two kinds of knowledge to identify propositions: knowledge of the syntax of the target language and real world knowledge. Syntax knowledge enables the listener to chink incoming discourse into segments or constituents. ... But knowledge of the world is also used to help identify propositions, enabling listeners to sometimes bypass the constituent identification press (1985: 190).

He goes on to discuss nine so-called "medium factors" which underlie the relationship between knowledge of the world and knowledge of the language: 1) the clausal basis of speech, 2) reduced forms, 3) ungrammatical forms, 4) pausing and speech errors, 5) rate of delivery, 6) rhythm and stress, 7) cohesive devices, 8) information content, 9) the interactive nature of conversation (1985: 193-97). In the discussion of the immediate
recall protocol to follow, the factors of rate of delivery, cohesive devices, information content, and the interactive nature of conversation will crop up again in other guises.

In describing the techniques for testing "auditory comprehension," Lado discusses the following seven "cautions" in the preparation of test items (1961: 218-20). Although he was writing primarily about the construction of discrete-point multiple-choice test items, his remarks are still relevant to the discussion of the immediate recall protocol.

1. Context. Lado says: "Give as much context as needed and no more." If the purpose to focus on one particular aspect of the spoken text, then a minimum of context is necessary. If, however, the interrelation between parts of connected text also play a role in the retention and recall of text, then the context should be as large as possible, even if superficially extraneous material remains. Giving students a title or a brief scenario helps alleviate "context grabbing" while a listening text comes deluging at them through the tape.

2. Length of lead material. "There is no optimal length for auditory comprehension leads." This is true. Experience shows, however, that there is an outside extreme for short-term memory processing within the format of the immediate recall protocol. Miller (1956) speaks of "seven plus or minus two" as the formulaic limit of retention of (seemingly) disconnected material. In working with material for the immediate recall protocol, the outer limit seems to be one minute running time,
with approximately 200 words per minute the optimal speed (Rivers, 1981: 173).

3. **Content.** Lado phrases it this way: "The content of the lead material should as much as possible be equally familiar or equally unfamiliar to all the students." Information in the text must be accessible from the learner's experience of the world. Again, in practice this means supplying unusual terms, especially proper nouns (place names, personal titles, etc.) to the learner.

4. **Problem dominance.** "One can easily ... inadvertently include another problem that may be even more difficult than the one we wish to test." This caution can be dealt with especially well with the immediate recall protocol where the learner determines more or less exactly what the problems are in processing the text. Once the learner has gotten used to writing down everything he/she recalls from the text, the protocols should improve in terms of quantity of detail and quality of expression.

5. **Limitation of pictures.** "[T]here is a point of diminishing returns in the use of pictures, a point which is reached when the pictures become so complicated or so remote from the idea they are supposed to suggest that they cannot be understood at a glance but have to be deciphered." There is a school of thought which encourages the use of visual material for listening comprehension training, such as Nord's "Picture Bingo" or "Picture Grammar Form" (1980, 1981). At the same time, the listening modality and the visual modality involve activation of
different processing locales in the brain (Walsh & Diller, 1981), as well as different processing strategies in the short-term memory (Kroll, 1975: 154-56). In an experiment on the efficacy of visual cues in listening comprehension tests, Mueller, replicating a study by Omaggio (1979), discovered that as listening proficiency increases, learners are less dependent upon visuals than learners at the beginning stages of language learning (1980). Examinees may thus be distracted by pictures and other visual aids when they (the examinees) are forming mental visualizations of the propositions being presented aurally. This is not to say that listening comprehension should not be trained with the aid of visual materials, whether pictures or charts or printed text. The testing of listening comprehension should, however, be trained and tested as much as possible without dependence on visual stimuli, so that the learner becomes independent of external visual reinforcement.

6. Choices in goal language. "The use of the goal [target] language in the alternatives [of a multiple choice test format] to check auditory comprehension is particularly susceptible to the introduction of irrelevant problems." The protocol avoids this by keeping the amount of guidance in the target language to a bare minimum. Most learners seem to have little difficulty comprehending directions in the target language as long as these directions relate directly to predictable tasks.

7. Choices in native language. "The caution here is against the exclusive use of the native language, because of the
deadening effect it will have on classroom teaching." One advantage of the immediate recall protocol is that it involves the comprehension of the target language through the medium of the native language, but not necessarily composition in it or using it for oral discussion. Indeed, many students find the technique highly satisfying, since even advanced students process the target language through their native language, much as they would not admit it openly. The protocol technique allows students to separate the native language they will never forget from the target language they are attempting to master. Indeed, I have seen no research on how long it takes for learners to begin thinking consistently and sustainedly in a target language. I suspect that a language learned in a classroom setting will never be internalized efficiently so as to be readily accessible in the myriad of situations in which most language is used outside of the classroom.

Sample Text and Sample Protocols

The sample listening comprehension text selected for discussion here was designed for teaching German (Jacobs, Karlsson, Vesalainen 1982). It was used during a composition-conversation course conducted during the Fall Semester of 1985 with a group of students who had already had the equivalent of four semesters of German at the University of Wisconsin. The text contains 140 words, across 54 "idea units." The running time is 43 seconds, or 195 words per minute, well within the range of 160 - 220 words per minute suggest as "normal" text
length (Rivers, 1981). The topic is familiar, namely, planning a vacation, although the locale (Finland) is familiar to most American students only as a place very far away. Specific place names like "Kangasniemi" and "Helsinki" thus have to be given to students before listening to the text, so as not to distract them with lexical specifics while attempting to process clausal features.

More interesting than any statistics that might be generated are the sample protocols given in Appendix B, as well as a sample of test items produced on the basis of student protocols, in Appendix C. A number of things must be stated about this sample. First of all, the title of text contains the label "Text B." Text A is a longer, more detailed version of the content, used in class as a preparation for Text B, which in turn was part of an examination. Thus, the context was already familiar to students, although they did not know ahead of time that this particular text would be part of an examination. As a result, the protocols may seem remarkably detailed and difficult to distinguish qualitatively from one another, except in the extremes. A number of things were learned from working with this particular text. First of all, because the students had heard a much longer version (Text A) all had some idea as to what the dialogue was about. At the same time, a number of items were left out of Text B which many students put into their protocols, leading me to believe that the did not really listen to the dialogue, even though they were admonished to do so. Secondly, a number of
instances occurred where students inserted whole German words heard on the tape (Blockhaus, Sommerhaus, PKW, Juni) or wrote English with pseudo-German spellings (kost, sommer, Finland). Third, there was considerable variety in the style and structure of each protocol. While most students wrote good English, a number were sloppy, with spelling errors, incomplete sentences, false starts, illegible handwriting. Many of these phenomena, of course, had already revealed themselves in the students' previous writing, both in German and in English, but their writing seemed to have been aggravated when operating in two languages simultaneously. At the other extreme were students who, not surprisingly, wrote as smoothly and accurately in English as in German. Many students wrote very clear, detailed, stylistically polished protocols. One could make a legitimate case for testing people in their native language skills as they progress in learning a target language. A visit to colleagues in the Department of English is definitely called for!

Finally, it became clear after this particular experiment that students were good at embellishing a minimum of detail comprehended with a maximum of stylistic devices mastered over years of verbalizing in their native language. In other words, the protocols with the least amount of detail were not necessarily the ones with the fewest words. This might lead one to conclude that the protocol might not be the best means of determining comprehension, since a handful of item comprehended could easily be "padded." As a testing instrument, then, the
immediate recall protocol may require considerable training on the part of the researcher before the various levels of processing can be sorted out.

The purpose in giving this particular text and its attendant protocols is to demonstrate how it is possible to generate listening comprehension items based upon actual student comprehension, not merely upon teacher projection of supposed comprehension. The protocols demonstrate just how much students are capable to recalling if given the opportunity to do so. Some students can, in fact, be quite creative, such as KS, who outlined the protocol as a dialogue, or MD, who recalled considerable detail in the proper order.

If administrative constraints dictate the use of mechanically scorable test items, such as multiple choice, running a text through a group of students first gives the test maker more confidence in the items ultimately generated. For example, items generated upon comprehension of quantities, dates, ages, etc. will probably be handled very easily by most listeners at this particular level, because most protocols contain some numerical references. The breakdown in comprehension seems to take place once different kinds of details start coming together rapidly, such as size (50 square meters) followed by quantity (6 beds) followed by household rooms (kitchen, sauna) followed by location (on a lake). The last items, in fact, involves a cognate between English and German which in German involves two homophones, all of which may throw off the more incautious
students into relocating the entire scene from a lake (der See) to the ocean (die See), producing responses such as "directly on the sea." The question would naturally arise to what extent these kinds of errors should be cycled into comprehension questions about the text.

The multiple-choice content items in Appendix C attempt to balance the need for specific listening tasks (items 1, 2, 6, 7, 8) with global comprehension tasks (items 3, 4, 5, 9). Item 10 asks students to select from a number of protocols, based upon actual protocols. All of the test protocols contain the same basic information, and none is either correct or incorrect upon the basis of one specific piece of factual information.

SUMMARY

The use of the immediate recall protocol for listening comprehension testing is in its beginning stages. In the United States, the years since 1982 have been dominated by discussion of language proficiency, which has in turn been dominated by the Oral Proficiency Interview, as developed by the ILR, ETS, and ACTFL. Although rating scales are available for reading, writing, and listening, none enjoys such robust psychometric health as the speaking scale. What is needed, especially for listening comprehension, are instrumentalities which yield ratable samples as reliable as those that can be generated for speaking. When we look back over the past 25 years, we realize that colleagues like Carroll and Lado have been advocating the implementation of materials and techniques that reflect realistic
language use, whether in the target language or the native language. The immediate recall protocol attempts to mediate between purely discrete-point testing procedures with teacher-generated items and global procedures with student-generated samples. The results to date seem somewhat messy, considering what students will produce when given the opportunity. At the same time, what students produce is always more fascinating than what we teachers come up with. The prospects for the next 25 years of listening comprehension testing are encouraging.
REFERENCES


Frau: Bitte sehr?
Mann: Wir wollen im Sommer nach Finnland fahren. Haben sie da Angebote?
F.: Ja. Sie können zum Beispiel eine Kreuzfahrt machen.
M.: Mhm.
M.: Aha.
F.: Sie können aber auch in einem Feriendorf wohnen oder ein Sommerhäuschen mieten.
M.: Ich glaube, das letzte ist etwas für uns.
F.: Schauen Sie sich das Angebot nur in Ruhe an.
F.: Und Sie wollen die Schiffspassage Travemünde-Helsinki.
M.: Ja. Und zwar für den sechzehnten Juni, drei Wochen. Für zwei Erwachsene und drei Kind unter zwölf und einen PKW.
F.: Dann wollen wir mal dieses Formular ausfüllen ... Wie ist Ihr Name, bitte?
M.: Aber warten Sie mal! Was kostet denn das alles.

Bitte sehr? /
Wir wollen / im Sommer / nach Finnland / fahren. /
Haben Sie da Angebote? /
Ja. Sie können ... machen /
zum Beispiel eine Kreuzfahrt /
Oder es gibt dieses Pauschalangebot hier. /
Schiffsreise / Travemünde / Helsinki / Travemünde [zurück] /
plus Rundfahrt / und kurzer Aufenthalt /
in den schönsten / Gegenden Finnlands /
Sie können aber auch / in einem Feriendorf / wohnen /
oder ein Sommerhäuschen / mieten. /
Ich glaube, das letzte ist etwas für uns. /
Schauen Sie sich das Angebot nur in Ruhe an. /
Das hier, das klingt ganz gut:
Blockhaus / 50 / Quadratmeter / sechs / Schlafplätze /
kleine / Küche / Sauna / direkt / am See. /
In der Nähe von / Kangasniemi. /
Sie wollen die Schiffspassage Travemünde-Helsinki. /
Für zwei / Erwachsene / und drei / Kinder / unter zwölf /
und einen PKW. /
Dann wollen wir mal dieses Formular ausfüllen ... /
Wie ist Ihr Name, bitte? /
Aber warten Sie mal. / Was kostet denn das alles?

[54 "idea units"]
APPENDIX B - Sample Protocols (plus number of "idea units")

[KS - 34] L - May I help you? M - Yes, we would like to Finnlann. Do you have offers? L - Yes we have tours. There is also this package deal. Or you could rent a summer house. L - You can look at these offers yourself (in peace?) M - Hmm. This one sounds good condo 50 cubic meters. 6 beds, sauna, small kitchen, on the lake near Kangasniemi. L - Then you'd like the ship package from Travemünde to Helsinki. M - Yes. 16 June, 3 weeks, 2 adults. 3 children under 12 and a motor vehicle. L - OK. I'll book it. What's your name? M - Wait! How much does it all cost?!

[SW - 33] A man enters a travel agency and inquires about trips to Finnland for a summer holiday. The travel agent has many suggestions including a trip from Travemünde to Helsinki and back with lots of stops in different areas of Finland. She also suggests a vacation village or cottage. He opts for the cottage and she tells him to look at them on his own. Then he finds what he is looking for. A cottage that has 50 m² of space, 6 sleeping places, a small kitchen, a sauna, and is right on a lake. The travel agent then books them a place on the ship for June 16th to return 3 weeks later for 2 adults, 3 kids under 12 years of age and a car. She asks him to help fill out the forms and when she asks his name he balks because she never told him the price.

[MD - 32] A man asks a travel agent what vacation opportunities does she have for Finland. She mentions that he can have _____ or a travel package which includes a roundtrip ship voyage from Travemünde to Helsinki & back or also offered are vacation towns (Feriendorf) or summer houses/cabins. He says the summer house is best for them. He looks through a brochure & finds a cottage/cabin that he likes. The cottage is 50 m², has 6-sleeping places, a small kitchen, a sauna & is located directly on the sea. He notices that it's near Kangasniemi & asks the lady where that is. She tells him it's in the middle of Mikkeli & Jyväskylä. She asks him if he wants that offer and a ship ride from Trav. to Helsinki: He says yes on the 16 June for 3 week for 2 adults & 3 children under 12 and a VKW. He asks how much it costs. tape ends.

[AM - 28] A gentleman goes into a travel agent to look for a summer vacation in Finland. He has three children under 12, and a wife. He would like to leave June 16 for 3 weeks. He looks through a catalogue ad decides that he would like to rent a summer house right on a lake. The house is 50 sq. meters, it has 6 rooms, a small kitchen, a sauna ad is near the city of Kangasiemi. He would also like to rent a PKV (a type of car). He will go (with family) from Travemünde to Helsinki ad back Helsinki to Travemunde. He wants to know what all this is going to cost!
A man asks a woman for help in planning a trip this summer in Finland. The woman suggests a package deal including Helsinki and Travemünde, living in a village with other vacationers, or living in a cottage. The man likes the idea of a summer cottage best. He selects a cottage of 50 cubic meters, near Kangasniemi. It has 6 beds, a small kitchen, and a sauna: it's also right on a lake. He reserves it for 2 adults, 3 children (and he mentions that he has a car), beginning on June 16 and lasting 3 weeks. The woman is ready to fill out a necessary form, but the man wants to know the cost of the deal first.

The man wants to take a vacation in Finland. He is talking with the travel agent about possible summerhouses to go to various places he could stay. Travel from Travemünde to Helsinki by ship is agreed to be the best. He decides to stay at a small cabin (w/ a small kitchen) near Kangasniemi. He will be traveling w/ his wife & three kids so he doesn't want a cruise, but the cabin is just right for him & his family. The cabin is directly on the Ocean. They will be leaving in June. Kangasniemi is between Mikkeli & Jyväskylä.

The man wants to take his wife and three children on a three week vacation in Finland. The travel agent offered two packages. The man wanted to stay in a cottage in Kangasniemi and she offered to arrange a ferry trip from Helsinki but he preferred to drive a van. He wanted to know how much it would cost. They arranged what he wanted. His starting point was Travemünde and he arranged to take a ferry from there to Helsinki.

We want to go to Finland. What kind of package would you like we have many different types. A Resort with a kitchen and bay view. When would like to go? The 16th of July for three weeks, with my wife and three children which are under 12. How much does this cost. Would like a hotel or a summer resort?

A man enters a travel agency to ask about the possible trips that are offered to Finland. The travel agent starts listing them and the man says the last one, a small cottage on the see be the best. She gives details about this trip: the cottage is a "Blockhaus," that is, quite small and located near Kangasniemi (between Mikkeli and Jyväskylä). He takes this one and gives the details: how big his family is, the date of trip and the travel arrangements. She starts to fill out the receipt but he interrupts and asks how much it costs.

This man wishes to go to Finland with his family. He speaks with a travel agent who is trying to offer him suggestions as to where he might go. He decides to rent a summer house located on a lake. 50 square meters in area with a sauna. It is located near Helsinki.
APPENDIX C - Sample Test Items

1. The man's family consists of himself,
   a. his wife and two children.
   b. his wife and three children.
   c. two adults and two children.
   d. and two children.

2. They plan to travel to Finland from Germany
   a. by ship.
   b. by car.
   c. by plane.
   d. by ferry.

3. When the man came into the office, he knew
   a. he wanted to visit a country by the sea.
   b. where in Finland he wanted to go on vacation.
   c. that he would be driving part of the way.
   d. that he wanted a cabin by a lake.

4. Which of the following questions should have been asked by the travel agent, but was not?
   a. What is your name?
   b. Would you like to look at catalogues?
   c. Are you interested in package tours?
   d. Would you like an estimate of the costs?

5. Which adjective best characterizes the travel agent's attitude toward the customer?
   a. cooperative.
   b. indifferent.
   c. impatient.
   d. calm.

6. The sounds of Finnish are almost the same as those of German and are written with the same letters. Which one of the following Finnish words is mentioned on the tape?
   a. Kansilasti
   b. Kangasniemi
   c. Kantotuoli
   d. Kansankieli
7. Which one of the following words did you hear on the tape?
   a. Häuschen
   b. Mäuschen
   c. Päustchen
   d. Bäumchen

8. After the man reads the basic information, what does he say?
   a. Das letzte ist etwas für uns.
   b. Das ist das letzte.
   c. Endlich etwas für uns.
   d. Das letzte verletzte uns.

9. If you were the woman, what would you say at the end of the tape?
   a. Das kostet nicht viel.
   b. Wollen Sie gleich bezahlen?
   c. Sie reisen am 15. ab.
   d. Wie ist Ihre Adresse?

10. Which of the following sets of statements best summarizes the conversation between the travel agent and the customer?
   a. A gentleman goes to a travel agent to book a summer vacation in Finland. He has three children under 12, and a wife. He would like to leave June 16 for 3 weeks. He looks through a catalogue and decides that he would like to rent a summer house right on a lake. The house has 6 rooms, a small kitchen, a sauna and is near the city of Kangasiemi. He will go from Travemünde to Helsinki and back. He wants to know what it is going to cost.
   b. He wants to travel to Finland. He said that the last offer was the best for his family. The trip will start the 16th of June. He has three children under twelve years old. The cabin which his family will stay in is by Kangasniemi. At the end he asks the travel agent how much the trip will cost. The area they will be staying at is one of the most beautiful parts of Finland.
   c. The man wants to take his wife and three children on a three week vacation in Finland. The man wanted to stay in a cottage in Kangasniemi and she offered to arrange a ferry trip from Helsinki but he preferred to drive a van. He wanted to know how much it would cost. His starting point was Travemünde and he arranged to take a ferry from there to Helsinki.
   d. This man wishes to go to Finland with his family. He speaks with a travel agent who is trying to offer him suggestions as to where he might go. He decides to rent a summer house located on a lake, 50 square meters in area, with a sauna. It is located near Helsinki.
Listening Comprehension since 1961:
A Selective Bibliography

(Page numbers preceded by "pp." after the main citation refer to
the teaching and testing of listening comprehension, to the
extent that the entire citation does not deal exclusively with
listening comprehension.)

1961

Lado, Robert. Language Testing: The Construction and Use of

Carroll, John B. "Fundamental Considerations in Testing for
English Language Proficiency of Foreign Students." Testing the
English Language Proficiency of Foreign Students. Washington, DC:
1961. [see 1965 reprint]

1962

Spearitt, D. Listening Comprehension - A Factorial Analysis.
Melbourne, Victoria: Australian Council for Education Research,
1962.

1964

Lado, Robert. Language Teaching: A Scientific Approach. New York,

Newmark, Gerald & Edward Diller. "Emphasizing the Audio in the
20. [see 1965 reprint]

1965

Carroll, John B. "Fundamental Considerations in Testing for
English Language Proficiency of Foreign Students," 313-21 in
Harold B. Allen & Russell N. Campbell, eds. Teaching English as a
article]

Newmark, Gerald & Edward Diller. "Emphasizing the Audio in the
Audio-Lingual Approach," 98-100 in Harold B. Allen & Russell N.
Campbell, eds. Teaching English as a Second Language. A Book of
1965, 1972. [see 1964 article]

1966

Pimsleur, Paul. "Testing Foreign Language Learning," 175-214 in
Albert Waldman, ed. Trends in Language Teaching. New York, NY:


---

1968


---

1969


1970


1971


1972


1973

1974


1975


1976


1977


1978


1979


1980


1981


1982


1983


1985


1986


