RTT RETELLING METHOD: AN ALTERNATIVE APPROACH TO INTELLIGIBILITY TESTING

Angela Kluge

SIL International
2007

CONTENTS

0. ABSTRACT

This paper presents an alternative approach to intelligibility testing, the RTT (Recorded Text Testing) retelling method. This method constitutes a modified version of the standard RTT method as described by Casad (1974). The standard method uses a short text recorded from an L1 speaker of the speech variety being tested with questions about the text interspersed in appropriate places throughout the text. Respondents are required to answer these questions, which are in their first language (L1). This approach is based on the assumption that from the level of correctness that respondents have of the speech variety being tested, inferences can be made concerning the overall comprehension levels testees have of the speech variety being tested. Subsequently, conclusions are drawn concerning degrees of interdialectal intelligibility and which variety or varieties of a given language area, network or continuum might serve best as the written reference dialect.

In spite of its many strengths, experience over the years has indicated a number of difficulties with this method. Its question format is often found to be culturally inappropriate, requires indirect inference, and relies on difficult question selection.

To overcome these difficulties, the RTT retelling method has been developed. This modified version of the standard RTT method requires respondents to listen to a recorded narrative. The narrative is broken down into natural chunks of one or two sentences each. Rather than having to answer specific comprehension questions, respondents are required to retell the recorded text segment in their L1.

Following a brief review of the standard RTT method and problems regarding its question format, this paper presents a detailed description of the RTT retelling method, including test design, and testing and scoring procedures, as well as a brief discussion of some of the advantages and disadvantages of the RTT retelling method.

1. INTRODUCTION[1]

Intelligibility testing, or recorded text testing (RTT)**,** is widely used in language survey research to investigate levels of intelligibility among related linguistic varieties. Based on previous work by Voeglin and Harris (1951, in Casad 1974) and Wolff (1959, in Casad 1974), John Crawford of SIL International developed the first recorded text test, with preliminary studies carried out in 1964 (Casad 1974). Since then, SIL researchers have used the RTT method extensively to determine interdialectal intelligibility and "to evaluate the potential of particular linguistic varieties to communicate between speakers of different social or regional dialects" (O'Leary 1994:49).

In spite of its many strengths, there are also a number of difficulties that language surveyors have encountered over the years when applying the RTT methodology to assess intelligibility of related speech varieties. These difficulties mainly concern the following three aspects of the question format of the standard RTT method, outlined by Casad (1974): (1) the cultural appropriateness of asking questions; (2) issues of inferencing; and (3) the selection of adequate comprehension questions and their translation into the respondent's first language (L1).[2]

[1] I would like to thank Regina Hertz of SIL International for her helpful comments on this paper.

[2] See O'Leary (1994) for a more comprehensive discussion regarding difficulties that have been encountered with the RTT method, mainly regarding: "(1) The issue of defining intelligibility and deciding what we are aiming to measure through RTTs; (2) The problem of distinguishing inherent from acquired intelligibility in speech communities known to have contacts; (3) The difficulty of interpreting variation in RTT scores; and (4) The complexity involved in aiming to establish dialect and language boundaries through the use of RTTs" (1994:48).

Although the standard RTT methodology provides procedures which are expected to safeguard against the potential drawbacks of the RTT question format, these safeguards do not offer adequate solutions if test-taking difficulties due to cultural inappropriateness of asking questions and/or if inferencing are fairly wide-spread throughout the community under investigation.

To overcome the problems regarding the RTT question format, a modified version of the standard RTT method has been developed. This method requires respondents to listen to a recorded narrative which has been broken down into natural segments of one or two sentences each. Rather than having to answer specific comprehension questions, respondents are required to retell the recorded text segment by segment in their L1.

Before discussing the RTT retelling method in more detail, the standard RTT method is reviewed in Section 2 with special attention given to problems regarding the RTT question format. Section 3 presents a description of the RTT retelling method, with Section 3.1 focusing on the test design, Section 3.2 describing the testing procedures, and Section 3.3 discussing the scoring procedures. Finally, in Section 3.4, some of the advantages and disadvantages of the RTT retelling method are discussed.

2. STANDARD RTT METHOD

The standard RTT method, as described by Casad (1974), consists of a short text recorded from an L1 speaker of the speech variety being tested.

> "A subject listens to the text one time. The subject then hears the text a second time, with questions about the text interspersed in appropriate places throughout the text. The questions are dubbed into the text following the portion which contains the answer to the question. These questions are in the subject's mother tongue. A test has ten questions." (Blair 1990:73)

Figure 1 displays one example of the RTT question format, taken from a recorded narrative in standard Yoruba as spoken in Porto Novo, Benin. This narrative was used for intelligibility testing in a sociolinguistic survey of the Ede speech communities of Benin and Togo to explore whether and to what extent speakers of those communities understand standard Yoruba which is closely related to the Ede speech varieties. The example shows the first section of the Yoruba narrative with the matching comprehension question and expected answer.

Figure 1: Narrative in standard Yoruba from Porto-Novo (Benin) (Kluge 1999b, Appendix 8)

Section #1:

> "Eight years ago, I woke up one morning and wanted to go to another village. My wife told me I shouldn't go. I told her I was obliged to go. When my wife realized this, she went to call her friends to come and tell me clearly that I shouldn't go."

Question #1:

> "Why did his wife leave to call her friends?"

Answer #1:

> "To tell him not to go!"

The RTT methodology is based on the assumption that from the level of correctness that respondents display in answering a specific set of comprehension questions regarding a specific recorded text, inferences can be made concerning the overall comprehension levels respondents have of the speech variety being tested. Subsequently, conclusions are drawn concerning degrees of interdialectal intelligibility and which variety or varieties of a given language area, network, or continuum might serve best as the written reference dialect.

2.1. PROBLEMS WITH THE RTT QUESTION FORMAT

One of the main advantages of the RTT question format is that it requires respondents to provide precise answers. Since the expected answers tend to be rather short, it is fairly easy for the researcher to conduct a preliminary, on-the-spot evaluation as to whether the answer is acceptable or the researcher should probe for more information or replay a given section. Moreover, the recording and scoring of these answers requires little time.

There are, however, a number of problems with the RTT question format concerning (1) the cultural appropriateness of asking questions, (2) issues of inferencing, and (3) the selection of adequate comprehension questions and their translation into the respondents' L1.

2.1.1. CULTURAL APPROPRIATENESS

Showalter (1991:207) highlights some of the cultural problems that have been noted regarding the administration of questions in interviews:

> One of the main problems has been that the question-answer formula has often been interpreted as stimulus response behavior, without the interactional complication of conversation. Questions and answers are treated as though they are self-contained units, interpretable in themselves without a social context. This leads to misinterpretation of survey results (Mishler 1986:13-14; Briggs 1986:2-4). Another problem is that in many societies the primary function of questions may not be to elicit information, but to direct another's behavior. Questions are frequently used to scold children for improper behavior or to make requests or to give advice or to challenge another person. Thus, the use of a series of questions to elicit data is far from ideal. In fact, it may offend, embarrass, or even humiliate the respondents.

2.1.2. INFERENCING

Another problem arising from the RTT question format is inferencing at the level of text processing. This may result in unexpected answers to RTT questions. Presenting one problem example (see Figure 2), Graham (2000) explores the inferential nature of question and answer interchanges.

Figure 2: Example from Guinea-Bissau (Graham 2000:119)

Sentence:	"There were two birds in the tree."
Question:	"What was in the tree?"
Answer:	"Food"

In investigating respondent's responses, Graham (2000) noticed that the unexpected answer "food" was provided almost exclusively by elderly and middle-aged female respondents. Upon further investigation, Graham (2000:119) discovered "that these respondents had inferred that the birds were in the tree because the birds had found something to eat in the tree. There were many more examples of inferences from the elderly and middle-aged female respondents in their responses to such statement/question sets."

Exploring the inferential nature of question and answer exchanges, Graham (2000), in a review of the literature on question and answer exchanges, discusses Schank and Abelson's (1977) "scripts model of memory". Graham (2000:121) submits that question and answer exchanges are heavily influenced by "scripts" which the questioner and the respondents have developed throughout their life histories

> … to organize knowledge of the real world. These scripts organize information about situations and activities frequently encountered. They contain the important events that might occur in different situations and play an active functional role in the task of text processing.

In question and answer exchanges, these scripts guide the questioner and the respondents in their expectations regarding the question answering process. Thus, Graham (2000:125) suggests the conclusion that

> the respondents that inferred the birds were in the tree because they had found food in the tree have developed a script that the other respondents have not developed. For example, these respondents may share food-gathering responsibilities in common, which they do not share with the respondents that gave the expected response to the sentence/question/answer set. Hence, these respondents may have developed a script that guides them to expect to find food which may be gathered, when they see more than one bird in a single location.

Another important aspect highlighted by Graham (2000:125) is the role of pragmatics in the question and answer exchange:

> The question answerer needs to be aware of what the asker knows already, and provide the missing information accordingly. The women represented in the inferencing in the problem example set may have assumed that the asker was aware of the two birds in the tree, in that the asker also heard the recorded sentence providing this information. These women may have thus assumed that the birds were known information, and therefore they needed to supply new inferred information.

Respondents, however, not only attempt to create meaning regarding the specific question and answer exchange but also regarding the testing situation at large: why do the researchers conduct their research in the first place, what are their motivational goals, and what are their expectations concerning the local population? In an attempt of creating meaning of the testing situation, respondents may draw the conclusion that they are required to supply information concurrent with

the researchers' implicit goals rather than based on the story's content. Figure 3 displays one such example from the Cabe survey in Benin.

Figure 3: Example from the Cabe survey, Benin (Kluge 1999a)

Text:	"After this his grandmother started crying."
Question:	"What did his grandmother start doing?"
Answer:	"She started praying."

During the Cabe survey, respondents repeatedly gave unexpected answers along religious/spiritual lines although the recorded text was secular in nature. Respondents were, however, aware of the religious motivations of many foreigners (foreign missionaries) who had visited their area in the past. It would seem reasonable to conclude that respondents assumed that the researchers, all of whom were foreigners, had a religious motivation for conducting the research and that the respondents, therefore, were attempting to answer the RTT comprehension questions in light of these putative motivational goals.

Thus, due to differing scripts and expectations regarding the overall research situation and the specific question as well as differing assumptions regarding the overall research situation, RTT respondents may provide unexpected answers to presumably straightforward RTT questions. These unexpected responses, in turn, may lower respondents' overall RTT scores, and thus result in significant skewing of the overall test results and thereby may affect the conclusions regarding language program decisions.

2.1.3. QUESTION SELECTION AND TRANSLATION

In her article "The role of recorded text tests in intelligibility assessment and language program decisions", O'Leary (1994:60) discusses a number of variables that may affect RTT results, three of which are directly related to the RTT question format:

- Text differences: relative difficulty or complexity of text and of comprehension questions asked.
- Slight inequivalence of translation of test questions into each hometown dialect.
- Selection of questions, that is, questions may or may not hit crucial items of difference. There may be wide variation caused by the influence of particular lexical items.

In discussing the issue of question selection in more detail, O'Leary (1994:62) suggests "that all questions are not equal in terms of their effect on comprehension scores." For a given RTT, there may be one or two questions which most respondents fail to answer, whereas the remaining questions may be answered correctly. For O'Leary (1994:62f), instances like these raise a number of questions:

> How do we interpret such cases? What if other questions had been asked instead of the ones that the subjects missed, that is, questions with similar response patterns to the ones that most people answered correctly? What is the significance of a set of subjects' misunderstanding of

one particular word? Often the subjects can pick the correct word or phrase to answer the question, but admit that they do not know exactly what it means.

O'Leary (1994) points out that the researchers would need to have significant knowledge of the dialects under investigation or be able to discuss these issues with the local population in order to explore adequately why respondents failed to answer the questions correctly. These kinds of issues can significantly influence scoring decisions and thereby have an affect on overall testing results, and ultimately even on survey conclusions and recommendations. Due to the limited scope of most language surveys, though, investigations of these complexities are usually unfeasible (O'Leary 1994).

2.2. SOLUTIONS

Integral to the standard RTT methodology are two procedures, which are expected to safeguard against potential problems with the RTT question format.

Firstly, during the design phase, each RTT is control-tested with a panel of L1 speakers of the recorded speech variety in question. The main objectives of the hometown panel pre-testing are: (1) to ensure that the recorded text is representative of the speech variety in question, (2) to verify that the researchers who designed the test understand the text in the same way that the text is understood by the speakers of that variety, and (3) to ensure that the comprehension questions can be answered successfully by L1 speakers of the recorded text in question. Thus, potential problems related to question selection are expected to emerge during this pre-testing phase.

Secondly, during the actual RTT testing, each potential respondent is administered a hometown test that is a recorded text test in the main language. The main objectives of the hometown test are: (1) to screen out respondents with test-taking difficulties and (2) to double-check that the comprehension questions can be answered successfully by L1 speakers of the recorded text in question.

Although both the hometown panel pre-testing and the hometown testing provide valid means of identifying problematic questions and potential respondents who can not perform adequately, neither safeguard offers a solution if test-taking difficulties due to cultural inappropriateness of asking questions and/or to inferencing are fairly wide-spread throughout the community under investigation.

An alternative approach to the standard RTT question format would be intelligibility testing that does not require respondents to answer specific questions. In the next section, such an alternative approach is presented in more detail.

3. RTT RETELLING METHOD

The RTT retelling method requires subjects to listen to a narrative that has been broken down into natural segments of one or two sentences each and to retell the recorded text, segment by segment, in their L1 without having to answer specific comprehension questions.

An earlier version of this variation of the standard RTT method was developed by Ring (1981, 1995) and subsequently refined by Boafo et al. (1996), Kluge and Hatfield (2002), and Tompkins et al. (2002). To assess comprehension levels of speech varieties other than their own, respondents were required to listen to recorded segmented passages of speech and to paraphrase the passages they had just listened to in their L1. This approach to comprehension testing was employed in the context of the sociolinguistic surveys of Ghana's Central, Eastern, Northern, and Volta Regions to

assess the potential extensibility of existing materials to neighboring speech communities conducted by the Ghana Institute of Linguistics, Literacy and Bible Translation (GILLBT).

In 2005, this approach to comprehension testing was further refined and modified to be employed in the context of dialect surveys in the Papua province of Indonesia. In the following sections, the procedures for test design, the testing procedures, and the scoring procedures are discussed in more detail.

3.1. TEST DESIGN

The procedures for text selection and recording are the same as those required for the standard RTT-with-questions method as described in detail by Casad (1974), Blair (1990) and Grimes (1995). In short, an autobiographical text of about three minutes length is recorded from a native speaker of the tested language.

3.1.1. TEXT TRANSCRIPTION AND SEGMENTATION

In a first step of the text preparation, the recorded text is transcribed and translated both literally and idiomatically, using an interlinear script format such as suggested by Blair (1990:76).

In a second step, the transcribed and translated text is divided into 10–12 segments by breaking the connected discourse into natural segments of one or two sentences each.

Figure 4 presents the translation and segmentation of a narrative recorded in Western Sentani, a language spoken in the Papua province of Indonesia.

Figure 4: Western Sentani narrative (Papua province, Indonesia)

1. There is a story. We live on the western side of the lake and our lives (totally) depend on fishing. Through fishing, we eat and pay our children's school fees.
2. Every morning and afternoon I work as a fisherman. In the morning I pull the net in and in the afternoon I throw it out. Every morning and every evening goes like this.
3. One day, I threw out the net at Yokure Kou. I paddled until noon to get there to pull in the net.
4. As I pulled in the net, a crocodile was sleeping in the water under a Sago tree. I didn't see it. I pulled in the net with the crocodile behind me.
5. As I pulled, the back of the canoe nudged a Sago leaf which startled the crocodile. Immediately, it jumped out of the water to throw itself behind me onto the canoe.
6. As I turned around, I shouted out loud, "Aduh, aduh! Crocodile, crocodile!", so I exclaimed. There was no one to hear me. Not a single soul was there. I was all by myself.
7. I prayed, "Oh, Lord, thank you. Today bad luck was about to happen to me. Fortunately, the crocodile came down beside the canoe and missed me", so I prayed to the Lord.
8. From there, I pulled in the net very very fast and then paddled home. I paddled without stopping until I was back home.
9. At home, I told the story to my wife and children, "Listen! Today a crocodile almost got me and almost tore me to pieces. The Lord Almighty had mercy on me so that I lived and am safe."
10. Because of the crocodile, we look for fish with great caution, using spears and nets at night and at noon.

In a third step, pauses of one to two seconds in length are inserted between the identified segments. These pauses allow the researchers, who in most cases are not speakers of the tested language, to identify segment breaks more easily and thus stop the recording at the appropriate moment during the actual testing.

3.1.2. HOMETOWN PANEL PRE-TESTING

Before a newly designed RTT test can be presented to speakers of other speech varieties, the test needs to be control-tested with a panel of L1 speakers to:[3]

1. Ensure that the recorded text is representative of the speech variety in question.
2. Verify that the researcher who designed the test understands the text in the same way that the text is understood by mother-tongue speakers of that language.

[3] The "hometown panel pre-testing" of an RTT should not be confused with the "hometown test" that is used to screen potential testees during the actual RTT testing.

3. Identify the core elements that L1 speakers regard as pertinent for each segment to establish a base-line scoring system based on panel subjects' responses (see Section 3.2).

This pre-testing should be conducted with a hometown panel of eight to ten L1 speakers who come from the same community and speak the variety of the recorded text.

The L1 panel subjects are tested individually and listen to the story over headphones. As in an actual RTT testing situation, subjects listen to the text twice. During the first audition respondents listen to the text in its entirety. During the second audition they listen to the text segment by segment. After a segment has been played, the tape is paused and respondents are required to retell the respective section in their L1.[4]

Respondents' responses are translated by an interpreter into the survey team's working language, e.g., Indonesian or English, and written down by the researcher. If respondents retell only part of a recorded segment or retell a segment in unexpected ways, the researcher should probe for the missing parts. If some of the L1 panel respondents consistently retell the same sections in unexpected ways, this could indicate that (1) the L1 speakers of the recorded variety regard certain segment items as superfluous, (2) the recorded text does not represent a good sample of speech, or (3) there are problems with the transcription and translation of the text which may have escaped the researchers' scrutiny during the actual transcription/translation process. All probing question and answer exchanges and additional observations regarding respondents' reactions should be documented so that the researchers can refer back to them when establishing the base-line scoring system (see Section 3.1.3).

To facilitate the comparison of respondents' answers across segments, form sheets are recommended for the hometown panel pre-testing such as indicated below in Figure 5.[5]

Figure 5: Form sheet for hometown panel pre-testing

Date: _____ Location: _____ Researcher: _____

Segment	Testee #1	Testee #2	Testee #3	Testee #4	Testee #5
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					

[4] If the survey team decides to conduct RTT testing without letting RTT subjects listen to the story in its entirety first, then these same procedures should be followed during the Hometown panel pre-testing (see Section 3.1.2).

[5] For each testee note gender and approximate age.

3.1.3. IDENTIFICATION OF CORE ELEMENTS AND ESTABLISHING A BASE-LINE SCORING SYSTEM

One main objective of the hometown panel pre-testing is to identify the core elements that L1 speakers regard as pertinent for each segment.

To determine the level of exactness which can be expected from an L1 speaker of the language variety in question, each panel respondent's response is compared to the original narrative and to the responses given by the other panel respondents. Those elements of the original text that were not mentioned by all panel subjects are discarded.[6] Likewise, added information mentioned by some or all of the panel subjects but not found in the original text is discarded.

Since it is not always obvious which elements of a segment should be regarded as core elements and which ones should be discarded, this phase of the RTT design is preferably conducted as a group effort by several members of the survey team, since the team effort enables more standardized identification.

Once all responses have been discussed and the core elements have been identified for each segment, the base-line scoring system can be established. For each segment, the total number of core elements is determined and each core element is given one point. Based on the number of core elements, each segment is given a segment score, e.g., "3 points" or "5 points". Since the total number of core elements per segment may differ, total points per segment may also differ. Figure 6 displays some of the hometown panel answers and the identified core elements for segments 4 and 5 for the above-mentioned Western Sentani narrative.

Figure 6: Western Sentani narrative (Papua province, Indonesia) – Core elements for segment #4 and segment #5

Segment #4

As I pulled in the net, a crocodile was sleeping in the water under a Sago tree. I didn't see it. I pulled in the net with the crocodile behind me.

Hometown panel answers:

1. He was **pulling** like this and **didn't see** the big **crocodile between the Sago roots**.
2. While the man was **investigating the net**, he **didn't know** that there was a **crocodile there under a Sago tree**.
3. While he was **investigating**, he **didn't see** that a **crocodile was under the Sago (trees)**.

Core elements:

[6] Alternatively, the survey team may decide to set the threshold lower, for example at six out of eight or seven out of ten subjects.

– (he/the man) **pulled/investigate net**
– **doesn't know/see (crocodile)**
– **crocodile in the water/under Sago tree/between Sago roots**

3 core elements = 3 points

Segment #5

As I pulled, the back of the canoe nudged a Sago leaf which startled the crocodile. Immediately, it jumped out of the water to throw itself behind me onto the canoe.

Hometown panel answers:

1. As **he pulled**, the back of the **canoe hit the trunk**. The **crocodile was startled**. It **jumped out of the water beside the canoe**.
2. **I pulled it in** with the crocodile behind me. The tip of the **canoe touched and hit the Sago**. The **crocodile was startled**. It **jumped onto the tip of the canoe**.
3. While **he pulled in the net**, the **canoe hit a Sago leaf**. The **crocodile was startled** and **jumped onto the canoe**.

Core elements:

– **(he/I) pulled (in the net)**
– **canoe touches/nudges Sago leaf/tree/trunk**
– **crocodile startled**
– **crocodile jumps/throws itself onto/beside canoe**

4 core elements = 4 points

The results of the Western Sentani hometown panel pre-testing indicated a total score of 35, with segment scores of between two and four per segment, as displayed below in Figure 7.

Figure 7: Western Sentani narrative (Papua province, Indonesia) – Segment scores

Segment		Segment score
Segment #	1	4
Segment #	2	3
Segment #	3	4
Segment #	4	3
Segment #	5	4
Segment #	6	4
Segment #	7	2
Segment #	8	3
Segment #	9	4
Segment #	10	4
Total RTT score		35

3.2. TESTING PROCEDURES

The RTT testing procedures are the same as outlined for the hometown panel pre-testing: subjects are tested individually and listen to the recorded text over headphones; they are required to retell the recorded text segment by segment in their L1.

Regarding the number of auditions, the team may decide to choose one of two alternatives, both of which have advantages and disadvantages:

1. Two auditions: During the first audition, respondents listen to the text in its entirety. During the second audition, respondents listen to the recorded text segment by segment and are required to retell the respective segments in their L1.

   Advantage: Subjects are given access to the overall content before being required to retell the recorded text segment by segment. Overall, this procedure of listening to a cohesive discourse appears to be more natural and closer to daily communicative interactions.

   Disadvantage: Subjects may focus on the text in its entirety and therefore be tempted to retell more than one segment at a time. Subsequently, they may leave out some details of a particular segment.

2. One audition: Respondents listen to the recorded text segment by segment, without the opportunity to listen to the text in its entirety first. After the first audition of a segment, respondents are required to paraphrase the respective segment in their L1.

   Advantage: Subjects focus on one segment at a time.

   Disadvantage: Overall, this procedure appears to be less natural since the cohesive discourse is broken down and presented in segments, without giving subjects access to the overall content.

In light of the different advantages and disadvantages of the two approaches, it is recommended that the survey team choose one of the two approaches before the commencement of the hometown panel pre-testing and subsequently maintain consistency in its testing procedures throughout the entire research project. Such consistency is important in order to ensure that respondents' responses are comparable to those of the hometown panel and also comparable to each other.

Respondents are required to retell or paraphrase a given segment in their L1 with an interpreter translating the responses into the survey team's working language, e.g., Indonesian or English. The researcher should write down the complete answer rather than just noting "right" or "wrong" since "obviously right" or "obviously wrong" responses sometimes turn out to be half-correct.

When respondents' responses are incomplete or incorrect, the researchers may probe for missing parts and replay the particular segment. Again, it is important to maintain consistency across researchers and throughout the entirety of the research project in terms of the extent of probing and the number of replays. All probing question and answer exchanges as well as replays should be well documented so that the researchers can review and discuss them if necessary.

3.3. SCORING PROCEDURES

Once the RTT testing phase has concluded, each response is compared to the respective base-line response that had been established during the hometown panel pre-testing. To obtain full credit, RTT respondents are expected to mention all elements included in base-line responses. Thus, each response that provides the required core elements is immediately assigned the full segment score.

Variations from the base-line responses are listed on a separate sheet of paper or in a separate Word document, which includes the respondent's reference number, reference to the respective RTT text, and the segment number. Once all responses have been reviewed, the researchers convene to discuss and score the deviating responses.

Evaluating one RTT text at a time, the research team discusses deviating responses segment by segment. Comparing these responses to the established core elements and the responses given by other subjects, the deviating responses may be given a score of 0.5 or 0 points. Elements not included in base-line responses are not expected to be mentioned by RTT respondents. Likewise, respondents are not given extra credit if they do provide these elements.

Figure 8 displays one example of a deviating response on the Western Sentani RTT, segment #5. While the respondent correctly mentioned three out of the four core elements, she maintained that the man, rather than the crocodile, was startled. Therefore, she was given 0.5 point for the core element "crocodile startled", resulting in an overall segment score of 3.5 rather than 4 points.

Figure 8: Western Sentani narrative (Papua province, Indonesia) – Scoring segment #5

As I pulled, the back of the canoe nudged a Sago leaf which startled the crocodile. Immediately, it jumped out of the water to throw itself behind me onto the canoe.

4 core elements = 4 points

- **(he/I) pulled (in the net)**; **canoe touches/nudges Sago leaf/tree/trunk**; **crocodile startled**; **crocodile jumps/throws itself onto/beside canoe**

RTT testing response:

> While **pulling in the net**, the **canoe hit a Sago leaf**. The man was **startled**. The **crocodile jumped (to get into) onto the boat**.

Segment score: 3.5

Once all deviating responses have been discussed and evaluated, the scoring of the RTT responses can be completed, and for each respondent the segment scores can be added up to obtain the overall score for a given RTT text. Once the overall scores have been calculated for each RTT and for each subject, each script should be re-checked by a second surveyor to ensure scoring reliability.

To facilitate future scoring and ensure scoring consistency throughout an entire research project detailed electronic documentation is recommended for:

1. hometown panel pre-testing responses,
2. identified core elements, and
3. deviating RTT testing responses and their assigned scores.

Such documentation ensures that whenever the same RTT text is used again, scoring will be based on the already established base-line scoring system. Thus, future deviating responses that are identical to those given during previous surveys will be assigned the same scores as those previous responses. New deviating responses and their assigned scores need to be added to the already established electronic document so that they are available as scoring guidelines for further future research.

3.4. ADVANTAGES AND DISADVANTAGES

The RTT retelling (RTT/RT) method has a number of advantages when compared to the standard RTT-with-questions method (RTT/Q). The main advantage of the RTT/RT retelling method, when compared to the standard RTT/Q method, is the fact that comprehension of an entire text is tested, rather than that of selected sections only. A second major advantage is that in many more traditional societies, retelling a story is more appropriate and less threatening than answering questions. An additional advantage is that this method does not require the design of comprehension questions and the translation of these questions into the speech varieties of the communities under investigation.

However, there are also a number of disadvantages. Identifying the core elements and establishing the base-line scoring system is quite time consuming. Additionally, writing down respondents' responses takes more time than recording subjects' answers to RTT questions as required for the

standard RTT/Q method. Likewise, the scoring of the retold segments is more time consuming than the scoring of subjects' answers to RTT questions. Finally, determining whether a potential subject has passed the hometown test is more involved than with the standard RTT/Q method: it is easier to determine whether a potential subject has answered 8/10 or 9/10 questions correctly than to calculate whether a potential subject has achieved e.g., 28-30 points out of 35 total points.

4. CONCLUSIONS

Intelligibility testing forms an indispensable component of SIL's sociolinguistic investigations that have as their major objective assessing the potential extensibility of materials among related speech varieties. To investigate levels of intelligibility, recorded text testing is most commonly conducted by employing the standard RTT/Q method as outlined by Casad (1974). Experience over the years with this method, however, has indicated a number of difficulties due to its question format. To address these difficulties, this paper has proposed an alternative method which does not require respondents to answer specific questions. Instead, they are required to retell a recorded and segmented text in their L1.

It is suggested here, that this alternative approach to assessing intelligibility has a great deal of potential. The proposed RTT/RT method is thus submitted for further evaluation, discussion, use, and/or amendment.

5. REFERENCES

Blair, Frank. 1990. *Survey on a shoestring: A manual for small-scale language surveys*. Dallas: Summer Institute of Linguistics and University of Texas.

Boafo, Ebenezer, Christine Kehl, and Deborah H. Hatfield. 1996. Sociolinguistic survey of the Awutu-Efutu language area. Accra: GILLBT. ms.

Briggs, Charles L. 1986. *Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research*. Cambridge: Cambridge University Press.

Casad, Eugene H. 1987. *Dialect intelligibility testing*. Dallas: Summer Institute of Linguistics.

Graham, Steve. 2000. Questions, inferences, and answers. *Notes on Sociolinguistics* 5(3-4):117-127.

Grimes, Joseph E. 1995. *Language survey reference guide*. Dallas: Summer Institute of Linguistics.

Kluge, Angela. 1999. Cabe RTT: Questions for hometown and practice text. Cotonou: SIL International. ms.

Kluge, Angela. 1999. Sociolinguistic survey of Ede language communities of Benin and Togo. Volume 1: General overview. Cotonou: SIL International. ms.

Kluge, Angela and Deborah H. Hatfield. 2002. Sociolinguistic survey of the Safaliba language area. *SIL Electronic Survey Reports* 2002-041. Dallas: SIL International. Online URL: http://www.sil.org/silesr/abstract.asp?ref=2002-041.

Mishler, Elliot G. 1986. *Research interviewing: Context and narrative*. Cambridge: Harvard University Press.

O'Leary, Clare F. 1994. The role of recorded text tests in intelligibility assessment and language program decisions. *Notes on Sociolinguistics* SI2:48–72.

Ring, J. Andrew. 1981. *Ewe as a second language: A sociolinguistic survey of Ghana's Central Volta Region*. Research Review 12(2&3). Legon: University of Ghana, Institute of African Studies.

Ring, J. Andrew. 1995. Revisiting the Central Volta Region: Avatime / Santrokofi / Bowiri. Tamale: GILLBT. ms.

Schank, Roger C. and Robert P. Abelson. 1977. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale: Lawrence Erlbaum Associates Publishers.

Showalter, Catherine J. 1991. Getting what you asked for: A study of sociolinguistic survey questionnaires. In: Kindell, Gloria E. (ed.). 1991. *Proceedings of the Summer Institute of Linguistics International Language Assessment Conference, Horsleys Green, 23-31 May 1989*. Dallas: Summer Institute of Linguistics, 203-222.

Tompkins, Barbara, Deborah H. Hatfield, and Angela Kluge. 2002. Sociolinguistic survey of the Gua language community. *SIL Electronic Survey Reports* 2002-037. Dallas: SIL International. Online URL: http://www.sil.org/silesr/abstract.asp?ref=2002-037.

Voegelin, Charles F. and Zellig S. Harris. 1951. Methods for determining intelligibility among dialects of natural languages. *Proceedings of the American Philosophical Society* 95:322–329.

Wolff, Hans. 1959. Intelligibility and inter-ethnic attitudes. *Anthropological Linguistics* 1(3):34–41.