



12-2000

# Participant Reference in Narrative Discourse: A Comparison of Three Methodologies

Stephen A. Clark

Follow this and additional works at: <https://commons.und.edu/theses>

 Part of the [Psychology Commons](#)

---

## Recommended Citation

Clark, Stephen A., "Participant Reference in Narrative Discourse: A Comparison of Three Methodologies" (2000). *Theses and Dissertations*. 948.

<https://commons.und.edu/theses/948>

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact [zeinebyousif@library.und.edu](mailto:zeinebyousif@library.und.edu).

PARTICIPANT REFERENCE IN NARRATIVE DISCOURSE: A COMPARISON  
OF THREE METHODOLOGIES

by

Stephen A. Clark  
Bachelor of Arts, Wheaton College, IL, 1969  
Master of Science in Education, University of Kansas, 1974

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Arts

Grand Forks, North Dakota  
December  
2000

This thesis, submitted by Stephen A. Clark in partial fulfillment of the requirements for the Degree of Master of Arts from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

S. H. Lewis  
(Chairperson)

David J. Marshall

Stephen A. Markert

This thesis meets the standards of appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Harvey Knuff  
Dean of the Graduate School

9-15-00  
Date

PERMISSION

Title Participant Reference in Narrative Discourse: A Comparison of  
Three Methodologies

Department Linguistics

Degree Master of Arts

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Signature Stephen Clark

Date September 5, 2000



## TABLE OF CONTENTS

LIST OF TABLES .....	vi
ACKNOWLEDGEMENTS .....	viii
ABSTRACT .....	ix
CHAPTER	
I. THE ANALYSIS OF PARTICIPANT REFERENCE IN NARRATIVE DISCOURSE .....	1
Participant Reference: Some Theoretical Considerations .....	2
Three Methods for Analyzing Participant Reference .....	5
II. BACKGROUND INFORMATION ON THE SIO LANGUAGE .....	16
Geographic and Cultural Background .....	16
Pertinent Features of Sio Morphosyntax .....	17
III. ANALYSIS OF A SIO TEXT USING THE RECENCY/DISTANCE METHOD AND THE EPISODE/PARAGRAPH METHOD .....	26
Givon's Recency/Distance Method .....	29
Tomlin's Episode/Paragraph Method .....	38
Initial Evaluation of Both Methods .....	42
IV. ANALYSIS OF THE TEXT USING THE DEFAULT/MARKED METHOD .....	44
Subject Reference .....	45
Non-Subject Reference .....	61
V. DISCUSSION OF FINDINGS .....	69
Givon's Recency/Distance Method .....	69
Tomlin's Episode/Paragraph Method .....	71
Levinsohn's Default/Marked Method .....	72
Confirmation of the Usefulness of the Default/Marked Method .....	73
Concluding Remarks .....	75

APPENDICES.....	79
A: List of Abbreviations Used .....	80
B: Thematic Paragraphs (Episodes) in Text.....	81
C: Sample Page of Charting .....	82
D: Free English Translation of Text.....	83
E: Interlinearized Version of Sio Text .....	89
REFERENCES CITED .....	101

## LIST OF TABLES

Table	Page
1. Sio Pronouns .....	19
2. Subject Agreement Prefixes on Sio Verbs .....	20
3. Object Agreement Suffixes on Sio Verbs .....	21
4. Non-Subject Encoding Restrictions .....	22
5. Number of Participants Introduced in Each Grammatical Role.....	27
6. Frequency of Use of Nominal Encoding Strategies.....	28
7. Referential Distance for Subject Position .....	30
8. Referential Distance for Object Position .....	31
9. Referential Distance for Adjunct Position .....	32
10. Potential Interference for Subject Position .....	34
11. Potential Interference for Object Position.....	34
12. Potential Interference for Adjunct Position .....	35
13. Persistence for Subject Position.....	36
14. Persistence for Object Position .....	37
15. Persistence for Adjunct Position .....	37
16. Over-Coding and Under-Coding with Reference to Episode Boundaries .....	41
17. Coding Material Used for Subjects in Different Contexts: Initial Analysis .....	46
18. Default Encoding Levels for Subjects, by Context.....	49
19. Coding Material Used for Subjects in Different Contexts: Final Analysis.....	49
20. Coding Material Used for Non-Subjects in Different Contexts: Initial Analysis .....	63
21. Default Encoding Levels for Non-Subjects, by Context.....	64
22. Coding Material Used for Non-Subjects in Different Contexts: Final Analysis .....	64

23. Coding Material Used for Subjects in Different Contexts: The Kulambi Man .....	75
24. Default Encoding Levels for Subjects, by Context: The Kulambi Man .....	75

## ACKNOWLEDGEMENTS

The author wishes to thank Dr. Stephen H. Levinsohn, the Committee Chairperson, for his excellent advice and guidance in the preparation of this thesis. His quick and thorough review of each chapter, despite being busy with teaching responsibilities in Dallas at the time the thesis was being written, is deeply appreciated. It has been a pleasure and an honor to work closely with him.

Thanks also are given to Dr. Stephen Marlett, for his help in negotiating through the various requirements of the program, and to the outstanding instructors at the University, whose wisdom and enthusiasm were a great asset and encouragement. Included in this group are Dr. John Clifton, Dr. Tim Pulju, Dr. David Weber, and Dr. Lindsay Whaley.

Appreciation also is extended to Dr. David Marshall, who graciously added to his workload by agreeing to serve as a Committee member, and to Dr. Richard Cameron of the University of Illinois at Chicago, under whose instruction the author was privileged to sit for two courses.

Deep gratitude is expressed to the Applied Linguistics students at the Moody Bible Institute in Chicago, IL, whose eagerness to learn served as the author's motivation to pursue this degree, and to members of the Administration of the Institute who supported and encouraged the effort (especially Dr. Robert Woodburn, Vice-President).

Finally, the author thanks his wife, Dawn, without whose constant support and considerable personal sacrifice this program could not have even been started. As the only other speaker of the Sio language in the Western hemisphere, her assistance has been invaluable.



To Melikisede,

who was my language instructor, culture advisor, and personal friend during the eleven years I lived and worked with the Sio people of Papua New Guinea. His skill as a storyteller is unmatched, and it was through studying written transcripts of the many stories and legends he told that I was able to understand the nuances and beauty of the language. A wise and respected leader of the community and church for many years, it was Melikisede's desire that I speak "correct Sio" that kept me working through various plateaus and points of frustration in my language learning. To him I say,

*Ikura mbwera mbwera noko ne pavilɔŋa ara ŋinde isukana ŋana aloŋo Sio ŋgua. Wisitŋgu ara nɔ ku kawɔŋgu daŋge ŋalae pa nenda Maro Kindeni ŋanano.*

("Year after year, your wonderful assistance enabled me to learn the Sio language. With contentment I give great thanks to our God for you.")

## ABSTRACT

In recent years several linguists have developed quantitative methods for analyzing and describing how speakers of languages refer to activated participants in narrative discourse. One of the first of these was Talmy Givon, whose method (referred to by some as the Recency/Distance method) measured three factors in participant reference: referential distance, potential interference, and persistence. A few years later, Russell Tomlin proposed an alternative model, which he labeled the Episode/Paragraph method. He felt that Givon's method fails to account for the fact that the thematic paragraphs or "episodes" found in narrative texts largely determine the amount of coding material used. His method assesses the point within an episode at which a referring expression occurs.

Stephen H. Levinsohn developed a third strategy for analyzing referential systems in narrative texts. Informally called the Default/Marked method, he suggests categorizing subject and non-subject nominals into one of several contexts, then determining the default level of encoding for each context. Deviant tokens are then studied to determine the motivation for the deviation.

With such widely different approaches, the question arises as to whether one seems to more accurately and thoroughly describe participant reference than do the others.

A single, lengthy narrative text from the Sio language of Papua New Guinea was selected for the analysis. The first step was to prepare a chart showing the various constituents of each clause in the text. Tabulations were made regarding the participants in the text, including the clause position in which they were introduced and the amount of coding material used. Then, the Recency/Distance method was applied, and values were calculated for each of the three major factors. The next step was to identify the episode boundaries within the text, and apply the

Episode/Paragraph method to determine how many over- or under-coded tokens could be explained by their position within episodes.

Then, the Default/Marked method was applied to the text. Default encoding levels were obtained for each subject and non-subject context, using detailed analyses of each participant in every clause. Once this was complete, those tokens that had an amount of coding material other than the default level for that context were analyzed, to determine possible reasons for this marked encoding. Finally, the results obtained using each of the three methods were assessed.

It was determined that the Default/Marked method provides the most thorough analysis of participant reference. The Recency/Distance method demonstrated the relationship between referential distance and the amount of coding material, but produced less convincing results for the other two factors, and failed to account for certain important discourse features and structure. The Episode/Paragraph method proved to be even less effective, as the number of over- and under-coded tokens explained by the position in arbitrarily pre-determined episodes was quite small.

In contrast, the Default/Marked method accounted for all aspects of the referential system in Sio. Default levels for each of the contexts were easily obtained, and the motivation for non-default (marked) encoding was as predicted by the theory. Thus it serves as a comprehensive and flexible method, one which undoubtedly can be effectively applied to narrative discourse in any language.



## CHAPTER 1

### THE ANALYSIS OF PARTICIPANT REFERENCE IN NARRATIVE DISCOURSE

Texts of the narrative discourse genre can provide linguists with a wealth of data for research. These texts normally contain events that are controlled by "agents" and are contingent upon previous events, and they often involve some sort of tension (Longacre 1996:9; Dooley and Levinsohn 1999:3). Narrative texts typically contain multiple participants, a story line that contains a beginning, one or more climaxes, and a conclusion, as well as related features that have led to interesting linguistic analyses, especially during the last 25 years.

One of the features of narrative texts that has been studied in some detail is participant reference. Basically, researchers investigate how it is that participants are introduced or "activated" in a text, how they are referred to once they have been activated, and how they are brought back into the discourse following an absence. While much of the earlier work was mainly theoretical, linguists in the 1980s began developing and discussing methods to quantitatively assess the ways in which participants in a narrative are established and maintained.

Among the researchers in this area of study are Talmy Givon, Russell Tomlin, and Stephen Levinsohn. Each of these linguists has proposed strategies for assessing the linguistic forms of referring expressions, and while they all are looking at the same variables, the strategies they have developed are different. An interesting question is thus raised: what are the strong and weak points of each of these different linguists' methodologies, as assessed using a sample of a narrative text from a language spoken by a small group of people in the South Pacific nation of Papua New Guinea? Which method best explains the system of participant reference, especially in those instances in a text where an unexpected amount of linguistic encoding material is used?

It is the objective of this study to attempt to answer these questions, using a narrative text from the Sio language (see the following chapter for specific details regarding the classification and morphosyntactic features of this language). The outline of the study is as follows: Chapter 1 presents a brief overview of the theoretical issues of participant reference, as well as descriptions of the methodologies developed by the three researchers named above. Chapter 2 provides pertinent background information about the Sio language. Chapter 3 contains the analysis of a Sio text using Givon's and Tomlin's methodologies, and Chapter 4 presents the findings of the analysis of the same text using Levinsohn's method. Finally, Chapter 5 discusses the findings of the study, attempting to answer the questions raised in the preceding paragraph.

#### Participant Reference -- Some Theoretical Considerations

The discussion of why referring expressions take the linguistic forms that they do has led to the use of a number of terms by various researchers that often seem to overlap in definition. Givon (1983, 1990) speaks of "topicality"; Lambrecht (1994) refers to "accessibility"; and Prince (1981:230), while preferring the term "assumed familiarity", lists other frequently used terms, including Halliday's (1978) "predictability", Chafe's (1976) "saliency", and Clark and Haviland's (1977) "shared knowledge". However, despite this sometimes-confusing use of terms, there are a few key principles of participant reference with which most theorists seem to agree.

A narrative text is essentially a set of instructions from a speaker to a hearer on how to construct a particular discourse-model, and it typically contains discourse entities (or participants), attributes of these entities, and links between the entities (Prince 1981:235). In order for a narrative text to be coherent, the hearer must be able to construct a meaningful mental representation that accounts for the participants and events. Thus it is the speaker's task to make clear the identity of each participant and event as they are introduced and maintained in the discourse. How the speaker encodes participants (that is, the amount of linguistic material used



in making reference to them) to ensure that reference is unambiguous is one of three major tasks of a reference system. The other two are to signal the activation status and prominence of referents, and to handle disruptions in the flow of information (Dooley and Levinsohn 1999:55).

Although languages differ to some extent, there generally are four different linguistic devices used for referring expressions in narrative texts (Givon 1983:17). A full noun phrase represents the largest amount of "coding material", and is typically used to introduce new entities into the discourse, or to reintroduce them if they have been absent for some time. (In languages that differentiate formally between definite and indefinite noun phrases, the latter are more likely to be used for the first introduction of a participant.) Once a participant has been activated in a text, expressions with less coding material typically are used to refer to that participant, ranging from stressed or independent pronouns, to unstressed or bound pronouns ("agreement markers"), to zero anaphora (the least possible amount of coding material). These forms have been called "de-categorialized" nouns (Hopper and Thompson 1984:722-723).

Participants typically are introduced into a text in a non-active role in the clause. Once introduced, they are "discourse old" and are more likely to be subjects (Prince 1992:316), although factors such as the animacy of the participant may affect this (Comrie 1989:185ff; Cooreman 1983:451ff). The expectation is that the speaker will use the least amount of coding material that is required for the hearer to identify the referent (see Grice 1975:45-46 for discussion of this point, which he calls the "Maxim of Quantity"). Therefore as ambiguity of identification, disruption of the flow of information, or activation status decrease, the amount of coding material will increase (Dooley and Levinsohn 1999:56). This enables the analyst to predict how much coding material will be used for a particular referring expression.

This relationship between the speaker's assumptions about the hearer's mental representations, and formal sentence structure, has been proposed to be governed by the rules and conventions of grammar in a component called "information structure" (Lambrecht 1994:5).

Within this structure, the identifiability and activation status of referents correlate directly with the structural properties of sentences. Thus a speaker will, through meaningful choices, construct a referring expression that matches his/her hypothesis about the current mental representation of the discourse that exists in the hearer's mind. Identifiability deals with the speaker's assessment of whether a mental representation of the referent is already stored in the hearer's mind ("knowledge") (1994:87), while activation status involves an assessment of whether that mental representation is accessible to the hearer at the time of utterance ("consciousness") (1994:93ff).

However, despite the speaker's intention to clearly identify referents in order to make the text coherent, participant reference is not a simple or clear-cut process. There is no guarantee that the hearer will in fact correctly identify the intended referent. Rather, successful identification depends on the extent to which both the speaker and the hearer share the relevant knowledge and beliefs (Schriffin 1994:198). The speaker is constantly evaluating the state of the hearer's knowledge, and this evaluation largely determines the precise form of reference used (Prince 1992:310). Thus, although reference may be aided by certain grammatical properties of referring expressions, ultimately the hearer may need to guess at the identity of the speaker's intended referent (Green 1996:34-35). This has been elsewhere described as being due to an "inherent mismatch" between the speaker's and hearer's mental representations (Brown and Yule 1983:210).

Much more could be said about the theoretical aspects of participant reference. However, in focusing on theory, there is a potential to allow argumentation to carry the analyst "further and further from the data" (Napoli 1996:397). Therefore the remainder of this Chapter will consist of a description of the work of three analysts who have developed quantitative methods for tracking and assessing participant reference in narrative discourse.



## Three Methods for Analyzing Participant Reference

### A. Givon's "Recency/Distance Method"

Talmy Givon edited a volume in 1983 that presented the first comprehensive attempt to quantify participant reference. This study included articles dealing with a variety of languages, including Japanese, Amharic, Ute, Biblical Hebrew, colloquial Spanish, Hausa, Chamorro, and colloquial and written English. Focusing on what he called "topic/participant continuity", Givon identified (pp. 9ff) three types of topics: chain-initial (a newly introduced, changed, or reintroduced topic), chain-medial (a continuing and persistent topic), and chain-final (a non-persisting topic). The latter two he identified as "definite topics", while the first type consists of "indefinite topics". The ease with which a hearer can identify a topic (the topic's "availability") is affected by how long it has been absent from the discourse register; how many other topics are present in the preceding discourse; and the presence of additional semantic or thematic information that provides probabilities that assist in identification. Givon felt that the first two of these factors were the most concrete and measurable, and he set out to establish a method for measuring them.

Assuming that what continues is more predictable, and what is predictable is easier to process, Givon suggested (pp. 12ff) that the linguistic devices used to encode topics/participants, and the exact position of these referring expressions within the discourse, could be correlated with three primary "discourse measurements": referential distance, potential interference, and persistence.

1. Referential Distance -- This is a measure of the gap between the previous occurrence of the referent in the discourse, and its current occurrence in a clause. By counting the number of clauses "to the left" since the most recent mention of the referent (which need not be overt, as long as the referent was a semantic argument of the predicate of the preceding clause), Givon felt that this "look-back" value could help explain why a particular form of referring expression was

used. Each mention of a topic/participant was assigned a look-back value from one (mentioned in the immediately preceding clause, and thus maximally continuous) to 20. The use of 20 as an upper limit to referential distance was arbitrarily established.

2. Potential Interference -- This is a measure of the disruptive effect of other referents in the immediately preceding text on the availability and identification of the topic. The definition of "immediately preceding" varies by analyst, but generally includes the preceding three clauses. A topic is considered to interfere only if it is semantically compatible with the predicate of the clause under consideration (e.g. regarding animacy, agentivity, plausibility, etc.). If there is no potentially interfering topic, a value of one is assigned; if interference is suspected, a higher value is given (usually two, but in cases of high ambiguity, a value of three could be given).

3. Persistence -- While the previous two factors are anaphoric, this third factor is a measure of the referent's "decay" in the subsequent discourse. It serves as a measure of the speaker's topical intent; the higher the number of clauses in which the topic persists, the more "important" it is to the speaker. Thus a value is determined by counting the number of clauses "to the right" in which the topic continues an uninterrupted presence as a semantic argument of the clause. If it does not appear in the following clause, it is assumed to have decayed immediately, and is assigned a value of zero.

Givon predicted (p. 15) that paragraph-initial topics would have high referential distance and high persistence; paragraph-medial topics would have low referential distance and medium persistence; and paragraph-final topics would be low with respect to both referential distance and persistence.

Most of the studies included in Givon's volume adapted a significant strategy in counting clauses. Complement clauses, as well as clauses that represented direct speech quotations, were not considered to be "gaps" when the referent did not appear in the clause, but they were counted as occurrences when the referent did appear in the clause.



Givon said (p. 17) that topic accessibility is scalar in nature, and that languages have certain syntactic devices that serve as "coding points" to reflect how accessible a certain topic is. While there is considerable variation among languages, the general tendency is to encode the least continuous or accessible referents with a full noun phrase, and to code the most continuous/accessible ones with zero anaphora. In between these two points fall the other most common linguistic devices -- stressed/independent pronouns, and unstressed/bound pronouns. Thus a hierarchical scale of topic continuity, in order of increasing accessibility, is:

Full Noun Phrase > Stressed Pronoun > Unstressed Pronoun > Zero Anaphora

Givon summarized his theory in what he called the "Iconicity Principle", which states, "The more disruptive, surprising, discontinuous, or hard to process a topic is, the more coding material must be assigned to it" (1983:18). This iconic relationship between linguistic tokens and the speaker's assessment of the topic's availability to the hearer is foundational to Givon's approach.

Givon's 1983 volume included a number of studies in which this methodology, or a variant of it, was employed. [It should be noted here that Givon did not use the label "Recency/Distance Method" for his approach; this label seems to have been given to it by Tomlin (1987:455.)] Givon concluded (p. 182) that referential distance is the most reliable measurement of topic continuity and thus accessibility. He also noted that persistence is only an indirect reflection of topic importance (p. 185), and he pointed out that the direct object position often serves as the entry point for important topics/participants, but these topics convert into clausal subjects if they persist (p. 187).

Several of the studies included an analysis of the animacy of referents. Among the more thorough discussions was that of Cooreman in her analysis of Chamorro. She noted (1983:465) that animate referents are more frequent and are maintained as topics much longer than inanimate

referents (90% of the subjects in her text were animate). Similar findings regarding Biblical Hebrew were presented by Andrew Fox (1983:315). Thus animacy does seem to be an important variable to consider in studies of participant reference.

Another important observation was that one can expect more radical disruption in topic/participant continuity at discourse junctures where there are radical changes in time, place, action and participants. Neither Givon nor the other analysts elaborated substantially on this point, and there was no attempt to document this through specific examples.

In his later work, Givon expanded on his notions of topicality, suggesting that the two quantifiable aspects of topicality in discourse are referential accessibility (anaphoric) and thematic importance (cataphoric) (1990:902). He added a binary "switch reference" factor that measures whether or not the preceding clause has the same referent as an argument. Describing the measurement of thematic importance as "not easy, since it is in principle subjective" (p. 908), he expanded the factor of persistence or "decay" to include both local and global persistence, and hypothesized that high topicality correlates with low referential distance and higher topical persistence.

#### B. Tomlin's "Episode/Paragraph Method"

In 1987, Russell Tomlin edited a volume in which Givon's method for quantitatively assessing participant reference in narrative discourse (particularly, his concept of referential distance) was critiqued and suggested to be inadequate. In his own article in this volume, Tomlin stated that his goal was to explore the relationship between the thematic organization of narrative discourse, and the syntax of reference. He felt that while Givon did establish a relationship between referential distance and the amount of encoding material used, his method failed to address two specific classes of counterexamples: when full noun phrases refer to antecedents only one clause away, and when pronominal encoding of a referent is sustained for more than one clause (1987:456). Tomlin stated that the theory thus did not account for the full range of use of



referring expressions that is found in natural discourse, noting that "it admits a great many potential counterexamples but offers no systematic explanation for them" (1987:475).

In the same volume, Barbara Fox also addressed Givon's continuity hypothesis. She stated that Givon's method fails to account for thematic information in explaining the use of referring expressions, ignoring units of discourse organization above the clause level. Rejecting this "flat and undifferentiated" view of text structure and attention flow (1987:159), she suggested that rather than relying totally on rigid quantitative procedures in analyzing anaphora, some qualitative analyses of the hierarchical structure of texts must precede the counting of clauses.

Clancy (1980:143-144) suggested that the main discourse structures that influence speaker choice of specific referring expressions are "episode boundaries" (including "changes of action") and "world shifts" (as when the text digresses from the main story line). He noted that full noun phrases tend to be associated with these points in a text. Tomlin, building on this theory, developed what he calls the "Episode/Paragraph" method of analyzing participant reference.

The basic assumption of Tomlin's model is that the syntax of reference is directly a function of the episodic or thematic boundaries of a narrative text, at a relatively local level. At such boundaries, the speaker uses additional coding material to reorient the hearer's attention. Thus the observed alternation between full nouns and lesser encoding, while certainly reflecting the maintenance of reference across adjoining clauses, is mediated by the higher-level organization of texts into thematic units (episodes). In short, speakers will use a full noun to reinstate reference across an episode boundary, and use lesser encoding to maintain reference within a particular episode (1987:457). As long as attention is maintained on a referent, less encoding material will be used; but whenever attention focus is disrupted, reference is reinstated with a full noun, regardless of the number of intervening clauses.

Tomlin recognizes the fact that full nouns also may be used in unexpected places by speakers to resolve potential ambiguity of reference (a factor which Givon accounts for in his "potential interference" measurements), but he does not attempt to address this in his method.

Thus Tomlin's method relies heavily on the analyst being able to identify where episode boundaries occur in a text. Tomlin admits that this is a weakness in his method (1987:457), as the identification of such boundaries depends on vague concepts such as "relevance" and "salience"; but at least it maintains a focus on the speaker's strategies, rather than the listener's. He insists that episode boundaries should be identified explicitly and independently of linguistic information, and that they often correspond with major changes in time, place, and participants, although ultimately they are defined by the sustaining of attention on a particular paragraph-level theme (1987:458).

The method developed by Tomlin assesses four factors: "absolute production", or the number of propositions and episodes in a text; "discourse density", or the proportion of propositions to episodes; "referential distance" (as measured by Givon); and "episode boundary results". It is this fourth factor which constitutes the heart of the method, so this will be explained in more detail below.

Tomlin classifies each referring expression in a text as either a "hit" or a "miss", depending upon whether or not it is consistent with his hypothesis. If the syntactic form of the expression is a full noun that also is the first mention of the participant since an episode boundary, or if it contains lesser encoding material and is not the first mention in that episode, it is a "hit". However, if it is a full noun but is not the first mention after an episode boundary (excluding cases of ambiguity resolution), or if it uses lesser encoding but is the first mention, it is a "miss".

The study presented by Tomlin did not consist of assessing hits and misses in a spontaneous narrative text (as is the case in this present study), but rather was a carefully



structured experiment where the episode boundaries were pre-determined either by the use of slides or by changes of scenery in a video. Subjects were asked to make up a story based on what they saw as the slides or video screening progressed, and their use of referring expressions in light of these pre-determined episode boundaries was assessed. As anticipated, whenever a slide or scenery change occurred, the subjects were likely to reorient attention in the story they were telling by using full nouns in referring expressions; yet while the slide or scene was not changing, reference was maintained with pronouns. In analyzing the "misses", it was found that most were in the area of "over-coding", and tended to be due to ambiguity resolution, or because the speaker had temporarily digressed to insert some non-narrative evaluative comments (an instance of Clancy's "world shift"), and was returning to the main discourse (1987:469-471).

Barbara Fox's study consisted of an analysis of participant reference in four popular contemporary English written novels (primarily of the "science fiction" genre). She found counterexamples to Givon's predictions, which led her to state that referential distance is not, in and of itself, a factor in the determination of anaphora in narrative texts. Rather, the hierarchical structure of narratives is more influential (1987:162). Regarding the "under-coding" of referring expressions, she found that minimal coding material is maintained until another participant's goals and actions are introduced, unless those goals and actions are interactive with the first participant's. Thus the only situation that requires the reintroduction of full encoding for a participant is when another character is introduced that is not interacting directly with the first (1987:163-164). In general, over-coding appears to mark the beginning of a new narrative unit.

Thus Tomlin's method represents a reaction to, and an expansion of, Givon's work. It attempts to explain some of the referring expressions in a text that are over- and under-coded, using the episodic or thematic structure of narrative texts rather than relying on surface features that are linear and isolated from this underlying hierarchical structure. Fox stated that one of the crucial tasks facing analysts is "the development of models of text structure which can be

fruitfully used in the study of linguistic coding" (1987:172) -- and Tomlin's model is a move in that direction (Carlson 1987:4).

### C. Levinsohn's "Default/Marked Method"

In an effort to more satisfactorily describe participant reference, and particularly to account for the apparent under- and over-coding of significant numbers of referring expressions left unexplained by Givon, Stephen H. Levinsohn has developed an eight-step methodology (1999:63-67) which may be called the "Default/Marked Method". Like Tomlin's method, Levinsohn's recognizes that there are factors beyond those identified by Givon that impact reference in narrative texts. However, Levinsohn rejects the pre-identification of episodes and episode boundaries, and looks instead at a combination of both linear and hierarchical structural features of texts. The methodology is adaptable to any specific language, and provides a precise step-by-step procedure.

The first step in Levinsohn's method is to draw up an inventory of the possible ways in which a language can refer to participants. These then are assigned to one of four general categories that parallel those identified by Givon: nouns (with or without qualifiers), stressed pronouns, unstressed pronouns, and zero anaphora. Levinsohn does not present these categories as being universal, though they are typically found in languages (Levinsohn, personal communication).

The second step is to prepare a chart of the participant encoding in a narrative text from the language being studied. This involves listing each clause separately, identifying both subject and non-subject participants (as well as other information) in each clause, and then classifying each referring expression into the four categories identified in the first step. The third step is to track the participants through the text, using numbers for ease of identification (thus the first participant introduced in a text is assigned the number one, and this number is written next to



every subsequent appearance of that participant in subject or non-subject positions, regardless of the amount of encoding material used).

In step four, attention is focused on the subject and non-subject positions. First, the subject in each clause is classified into one of four "contexts":

S1 -- The subject is the same as in the previous clause or sentence\*;

S2 -- The subject was the addressee of a speech reported in the previous sentence;

S3 -- The subject was involved in the previous sentence in a non-subject role other than  
S2;

S4 -- Other changes of subject than those covered by S2 and S3.

\* = The S1 context also extends to situations where the subject and non-subject of the previous sentence combine to form a single, plural subject.

A fifth context, INTRO, marks the first time the participant is mentioned in the text (thus Levinsohn's model differentiates between initial reference and continuing reference).

Then, each activated non-subject is assigned to one of four contexts:

N1 -- The referent occupies the same non-subject role as in the previous sentence;

N2 -- The addressee of a reported speech was the subject (speaker) of a speech reported  
in the previous sentence;

N3 -- The referent was involved in the previous sentence in a different role (other than  
that covered by N2);

N4 -- Other non-subject references than those covered by N1-N3.

The fifth step is to propose a default encoding for each context, either by actual statistical counting or by inspecting the data. This would be the encoding that occurs most frequently in that context, indicating a: absence of discontinuities or surprises (as these terms are used by Givon), or other complexities. Then as the sixth step, any referring expression which has received an encoding level other than that proposed as the default level for that context is

identified. Levinsohn proposes that when the amount of coding material is less than the default level, it is assumed that the referent is being treated as prominent (see additional discussion below), there is only one major participant on stage, or a cycle of events is being repeated. When the coding material is more than predicted, this usually is due to its occurrence immediately following a point of discontinuity, or for highlighting the referent.

Step seven involves modifying the proposals made in Step 5 regarding default encoding, based on the review made in Step 6. This essentially amounts to a "fine-tuning" of parameters that are generally applicable to most languages, in order to make them as descriptive as possible of the language being studied. Finally, in the eighth step, each remaining "deviant" token is analyzed, to determine the motivation for the special encoding, and to draw generalizations.

As was mentioned above, Levinsohn focuses on the relative importance of different participants in a text. A major participant is one that is active for a large part of the narrative, has a leading role, and typically is introduced formally (i.e., with linguistic material that instructs the hearer not only to activate the participant, but also to be prepared to fit the participant into a major role). A minor participant, in contrast, is activated only briefly and without a formal introduction. The use of referring expressions partially depends on whether or not the participant is a major one.

Levinsohn's approach identifies participants that are particularly prominent in the text, and uses this prominence to explain unexpected levels of encoding in the text. He refers to this as a "VIP (Very Important Participant)" strategy, and explains that participants can be thus highlighted both in the text as a whole, or within one section of the text. This concept goes beyond some subjective sense of prominence, but looks for linguistic cues to support the fact that this part of the text is "about" this participant, in a key way. This VIP strategy, which Fox claims "does not work" in written English narratives (1987:160), was first introduced by Grimes (1975), in his discussion of "thematic participants".

A "Global VIP" is the most prominent participant in the entire text, and may receive special (usually minimal) encoding as a result. It also is possible to identify a participant as a "Local VIP", one that is especially prominent in a certain section of the discourse, and is so marked.

Thus the major differences between Levinsohn's method and the others presented in this Chapter include 1) the establishment of default values before the encoding of individual tokens is analyzed (cf. Tomlin); 2) the focus on explaining individual referring expressions rather than computing statistical norms (cf. Givon); and 3) the use of highlighting and the VIP strategy as key explanations for over- and under-coding of specific tokens (neither Givon nor Tomlin explore these factors).

In conclusion, each of these three methodologies has a different focus and different underlying hypotheses about how participant reference should be assessed. However, all three are capable of being tested, and Chapters 3 and 4 of this paper present the findings of a study that applies the three methodologies to a narrative text from the Sio language. First, however, some basic features of Sio that have an impact on participant reference need to be addressed.



## CHAPTER 2

### BACKGROUND INFORMATION ON THE SIO LANGUAGE

The Sio language is spoken by approximately 3,000 people who live in five small villages along the northeastern coast of the Huon Peninsula, Morobe Province, Papua New Guinea. According to Grimes (1996), the formal classification of the language is Austronesian, Malayo-Polynesian, Central-Eastern, Eastern Malayo-Polynesian, Oceanic, Western Oceanic, North New Guinea, Ngero-Vitiaz, Vitiaz, Sio. The author and his family lived in the Sio village of Lambutina from 1985 to 1996, conducting a literacy and translation program under the auspices of the Summer Institute of Linguistics.

#### Geographic and Cultural Background

The Sio language area is bounded by major rivers to the west and east, by the ocean to the north, and by the foothills of the Finistere mountains to the south. The people live in five villages located directly on the beach. Access is either by sea craft or by coastal foot trails from the government station of Wasu, some 10 miles to the west. Once known for their production of clay pots as part of an extensive trade system around the Vitiaz Strait (Harding 1967), the Sio people are subsistence farmers, planting large gardens of yams, taro and other root vegetables, and supplementing their diet with fish and fruit. Cash is obtained primarily through the sale of copra.

The first contact between the Sio people and the outside world occurred in the late 1800's, when a German plantation company came by ship to recruit workers. In 1907, a group of Lutheran missionaries from Germany sought to establish a presence in the area, and in 1911



Michael and Maria Stolz built a home between the current Sio villages of Laelo and Balambu. The Stolzes lived there until 1930, bringing Christianity as well as a number of technological and medical changes. With World War Two came a time of upheaval, as fighting between the Japanese and the Allies forced the people from Sio Island (where they had lived for many generations) to the mainland, and then into hiding in caves in the mountains for over a year. Following the War, another expatriate German Lutheran missionary, Hans Wagner, lived in Sio for about 10 years (all historical references are taken from personal communication with Sio residents, or from Evangelical Lutheran Church of Papua New Guinea 1985).

Most of the people of Sio are multilingual, although the use of the vernacular is highly preferred. Melanesian Pidgin is widely used in contacts with outsiders. English, which is the official language of Papua New Guinea for government and education, is understood by most younger people, but is rarely used outside of these formal settings. Some of the older people speak either Yabim or Kote, which were the church languages in much of Morobe Province until the mid-1950s. A few Sio people are fluent in Komba or Gitua (neighboring languages to the south and east).

### Pertinent Features of Sio Morphosyntax

A complete description of the morphosyntax and phonology of the Sio language is available in other places (Clark and Clark 1987; Clark 1994), so this chapter will focus on features that are pertinent to the analysis of Sio narrative discourse: definite versus indefinite noun phrases, pronouns, subject agreement prefixes on verbs, object agreement suffixes on verbs, zero anaphora for non-subject reference, repetition of clauses, and the use of the continuative marker. Each of these is described below.

A. Definite versus Indefinite Noun Phrases

Sio has the possibility of constructing noun phrases that are either definite (usually marked by one of the demonstrative pronouns/deictics, /*ɲine*/ "this" and /*ɲinde*/ "that") or indefinite (usually marked by the indefinite pronoun /*tonge*/ "a, an"). Examples of both can be found in the text being analyzed in this study, even within a single clause (see 3 below):

- 520 **naŋa tini-ɲgu    pɔka wɔŋgu ɲine**                    "I am disgusted about this drum"  
*Is    skin-1sPOSS wrinkled drum    this*
- 141 **ɲineŋga zo    tonge koto tonge i-mɔ**                    "Then one day a rat came"  
*then    time a    rat    a    3s-come*
- 3 **nɔndi    ɲinde nde mira tonge**                    "That quartz is a rock..."  
*rock.name that    =    rock a*

However, this strategy is not always utilized, and a number of noun phrases are unmarked for definiteness. Example 196 below contains what appears to be a definite noun phrase, and example 136 contains a noun phrase that seems to be indefinite -- but neither is overtly marked:

- 196 **aku koto i-lo**                    "and (the) rat went"  
*and rat    3s-go*
- 136 **rua si-ɲgunu tɛŋgo**                    "the two of them built (a) fortification"  
*two 3p-erect wall*

It appears that the use of the demonstratives in narrative discourse is linked to the prominence of the noun phrase referent, in that only thematic participants (local VIP's, to use Levinsohn's terminology) or thematic props are thus marked. Once a key participant or prop has been activated in the text, it is marked in subsequent appearances with a demonstrative when it becomes the topic about which a comment is made. It then remains the local VIP/thematic prop until another participant is thus highlighted. Further, the indefinite pronoun /*tonge*/ "a, an" is





non-subject positions, none of them refers to an inanimate participant. This leads to the following restriction:

**RESTRICTION 1** -- Inanimate referents may not be encoded with pronouns.

### C. Subject Agreement Verbal Prefix

The Sio verb contains an obligatory subject agreement prefix, which indicates the person and number of the subject (the only other possible verbal prefix is a reflexive marker /**pa-**/, which makes the subject of the verb into the simultaneous recipient/undergoer of the action being predicated). There is high regularity in the subject agreement paradigm, with only one small set of phonologically determined irregular verbs (see below), and one otherwise irregularly inflected verb, /**yoka**/ "walk". The subject agreement prefixes are presented in the following table:

TABLE 2 -- Subject Agreement Prefixes on Sio Verbs

PERSON	NUMBER	
	Singular	Plural
First	<b>a-</b>	<b>ta-</b> = Inclusive <b>ka-</b> = Exclusive
Second	<b>ku-*</b>	<b>ka-</b>
Third	<b>i-</b>	<b>si-</b>

\* For verb stems starting with the voiceless velar plosive /**k**/ and consisting of two or more syllables, the prefix for second person singular changes to /**pw-**/, and the stem-initial consonant is deleted. Thus what would otherwise have been \*/**ku-kɔki**/ "2s-go.up", becomes /**pw-ɔki**/.

### D. Object Agreement Suffixes

There also is a set of object agreement suffixes, although there are significant restrictions as to when they may be used. These suffixes are used to indicate the person and number of the

direct or indirect object of the verb, and also to indicate the person and number of the object of a prepositional phrase involving one of the following prepositions: /pa/ "to", /ɲana/ "about", and /kuku/ "with". When an object agreement suffix is used, it usually replaces the overt object; in fact, the only instance in which an object agreement suffix can grammatically co-occur with an overt object is when the object is third person plural (although some speakers question the grammaticality of such redundancy). The normal object agreement suffixes are presented in the following table. As this table shows, there is no third person singular object agreement suffix in Sio. This has an immediate impact on the encoding of non-subject participants in a text, as a third person singular non-subject can only be encoded with a noun phrase, a pronoun, or zero anaphora.

TABLE 3 -- Object Agreement Suffixes on Sio Verbs

PERSON	NUMBER	
	Singular	Plural
First	<b>-na</b>	<b>-ninda</b> = Inclusive <b>-ma</b> = Exclusive
Second	<b>-no</b>	<b>-mi</b>
Third	--	<b>-nzi</b>

Thus the following restriction is proposed:

**RESTRICTION 2** -- No reference to a third person singular participant in a non-subject position in the clause may utilize an agreement suffix.

Also, as is the case with pronouns (as was explained above), there are similar restrictions regarding the use of object agreement affixes that pertain to animacy. There are 44 tokens of object agreement suffixes in the Nondi text, and none of them refers to an inanimate participant.

Taken together with the findings regarding pronouns, the following significant restriction can now be stated:

**RESTRICTION 3** -- The only permissible encodings for an inanimate participant in a non-subject position are noun phrases or zero anaphora.

A final restriction involves the use of the zero anaphora encoding strategy (see additional discussion below). There are 83 such tokens in the text under study, but when the referent is plural, none of these referents is animate. Thus the lowest level of encoding for a plural, animate non-subject is an agreement suffix. This leads to the following restriction:

**RESTRICTION 4** -- For plural non-subject participants, only inanimate referents can be zero encoded.

The following table summarizes the encoding restrictions that are placed on non-subjects regarding both number and animacy. Note that the phrase "No Restriction" is to be understood as meaning, "No restriction providing the referent is third person".

TABLE 4 -- Non-Subject Encoding Restrictions

NUMBER	ENCODING STRATEGY			
	Noun Phrase	Pronoun	Affix	Zero
Singular	<i>No Restriction</i>	<b>Animate Referent, only</b>	<b>No 3s; Animate Referent, only</b>	<i>No Restriction</i>
Plural	<i>No Restriction</i>	<b>Animate Referent, only</b>	<b>Animate Referent, only</b>	<b>Inanimate Referent, only</b>

#### E. Zero Anaphora for Non-Subject Position

Since there is an obligatory subject agreement prefix on verb stems, the use of an overt subject is redundant in cases where the referent is clearly understood. However, for the non-



subject position in clauses, there are additional factors involved in zero anaphora. The activation status of the referent certainly is one factor, as illustrated in the following example,

58 **aku si-kasoŋa 0**                    "and they asked 0 (her)..."  
*and 3p-ask 0*

where the identity of the subject is indicated by the third person plural verbal prefix, while the non-subject's identity must be inferred by the hearer from the preceding text (which is easily done in this and similar instances of zero anaphora).

Yet there are other reasons for zero anaphora for non-subject references. One is the use of what might be called serial verb constructions, where two or three verbs are so closely connected that the insertion of any additional material between them is awkward. These constructions usually involve verbs of motion. For example, consider Clauses 172-174 in the text:

172 **ku i-kai kanzi loo**    173 **i-mo**    174 **i-o 0 ndue**  
*and 3s-get 3pPOSS water    3s-come    3s-put 0 down*

"She got their water, and came, and set (it) down"

The zero anaphora for the direct object (water) in the above clause series may be partly explained by the clarity of the referent, but also is dictated by the verb series itself. In Sio, whenever there is a series of same-subject clauses that share a single intonation contour, any non-subject entity that is common to more than one of the clauses is only referred to once.

Finally, selectional restrictions of verbs may make the overt presence of the object unnecessary. For example, the verb /**zua**ra/ "shove" requires a direct object that is inanimate, large and heavy, yet still capable of being moved with effort. Thus in the sequence toward the end of the text, when the twins shove their brother's canoe out into the ocean, the only participant on stage that would meet these requirements is the canoe (other recently mentioned participants are their brother, the sail, the paddle, and a brace). Zero anaphora is the result:

- 641 **si-zuara 0** "They shoved 0 (the canoe), and it went down..."  
*3p-shove 0*
- 642 **ku i-ndue**  
*and 3s-go.down*

#### F. Repetition of Clauses

Another feature of Sio morphosyntax that affects discourse participant reference analysis is the technique of repeating clauses or portions of clauses. For example, consider the sequence of Clauses 560-563, as compared with 564-566:

- |                                                          |                                                 |                                          |                                    |
|----------------------------------------------------------|-------------------------------------------------|------------------------------------------|------------------------------------|
| 560 <b>ku si-kawea 0</b><br><i>and 3p-snatch 0</i>       | 561 <b>si-kai 0</b><br><i>3p-get 0</i>          | 562 <b>si-toa</b><br><i>3p-go.ashore</i> | 563 <b>si-mɔ</b><br><i>3p-come</i> |
| 564 <b>si-kai 0 lee lee</b><br><i>3p-get 0 CONT CONT</i> | 565 <b>ku si-toa</b><br><i>and 3p-go.ashore</i> | 566 <b>si-mɔ</b><br><i>3p-come</i>       |                                    |

"...they snatched it, taking it, going toward shore, coming. They took it on and on, and went ashore, and came..."

In this sequence, much of the material is identical to the preceding clauses (compare 564 with 561, 565 with 562, and 566 with 563). This seems to mark a clear episode boundary, and it does affect the calculation of variables measured by the methodologies being examined in this study.

However, this is not the only example of repetition of material in the text. Sometimes this feature occurs in what seems to be the middle of an episode. In these cases, the duplication may be for the purpose of highlighting. For example, the sequence of Clauses 452-462 contain several clauses that represent full repetition of preceding clauses. (In fact, these clauses contain what is referred to as "tail-head linkage", where the end of one clause is repeated as the beginning of the following clause.) The following is an English translation of a portion of these clauses, with reduplicated material indicated in italics:



452 "So they called out to the birds in their nesting places 453 to come. 454 *They called out to the birds in their nesting places,* 455 and they all *came,* 456 and they sent off all the birds. 457 *They sent off all the birds...*"

In this sequence, the repetition does not appear to mark episode boundaries or other discontinuities, but rather seems to be to highlight the action.

#### G. Continuative Marker

Finally, the use of the Sio continuative morpheme /*lee*/ is significant as a discourse device. This is an aspect marker that is used to indicate that the action or state being predicated in the clause goes on for an extended period of time (repetition of the morpheme signals either unusually long duration, or the continuation of the state/action). As such, it is an ideal marker of discontinuity of time, especially when used with the existence verb /*mo*/ "be, exist (animate referent, only)". The text being analyzed in this study is divided into episodes or thematic paragraphs (see Appendix B for a listing of the episode boundaries), and it is noteworthy that the morpheme /*lee*/ occurs in one of the opening clauses of 12 of the 40 thematic paragraphs. In fact, the most natural translation of the common construction [ /*mo*/ "be, exist" + /*lee*/ "CONT" ] (for example, see Clauses 24, 56, 91, 99, 107), would be "time went by".

With these features of Sio morphosyntax in mind, attention is now turned to the analysis of the Sio text using the participant reference analytical methodologies developed by Givón and Tomlin (Chapter 3), and by Levinsohn (Chapter 4).

## CHAPTER 3

### ANALYSIS OF A SIO TEXT USING THE RECENCY/DISTANCE METHOD AND THE EPISODE PARAGRAPH METHOD

In this chapter, an analysis of the Sio narrative text will be presented using the two different strategies already discussed: Givon's "Recency/Distance" method, and Tomlin's later approach, the "Episode/Paragraph" method. The analysis of the text using Levinsohn's methodology is presented in Chapter 4.

As background information, the text selected for this analysis is entitled "Nɔndi", which is the Sio word for a very hard, quartz-like rock that is found in rivers, and is also the name given to a mythical creature who had a removable, quartz-like skin. A free English translation of the text is included as Appendix D; but briefly it is the story of how two young brothers defeated this creature in battle, and then made a highly resonant drumhead from a piece of the quartz-like skin. Their older brother stole this drum from them, but ultimately they retrieved it and killed him, too.

The story was told by an elderly man named Melikisede, a resident of the Sio village of Niususu, and the unanimously acclaimed prime story-teller of the entire language group. For many years, Melikisede held key positions of authority both in the local village government and in the Lutheran Church structure. This is one of approximately 25 full-length narrative texts told by Melikisede, recorded on audio cassette tape, and later transcribed and printed for distribution in a village literacy program. Many of the stories he told were of the legend/myth genre, and his clan has proprietary rights to them. This particular text was recorded in Lambutina village on August 19, 1988.

The text consists of a total of 649 clauses, although, following the Givon model, clauses representing direct speech acts have been removed from the analysis. Since conversation is a key

feature of many narrative texts, this resulted in almost 25% of the clauses being removed.

However, the remaining 490 clauses provide ample data for the purposes of this study.

There are a total of 36 referents/participants in the text that serve as the subject of at least one clause. Of these, 15 are animate (humans, animals and other living beings), and 21 are inanimate (including canoes, spears and arrows, log fortifications, water, the ocean, etc.). Within the analyzed clauses of the text, an animate being is the subject in the vast majority of the clauses (440 out of 490, or approximately 90%). There are an additional 20 participants (all but two are inanimate) that serve as direct objects of clauses but not as subjects, and another 17 inanimate participants that occur only in adjuncts. Thus there are a total of 73 potential referents in the text. Using the entire text as a data base, the two younger brothers appear in a total of 239 clauses, making them the most frequent referents in the text. Other participants appearing frequently include Nondi (109 clauses), the older brother (109 clauses), and the boys' mother (90 clauses).

All but two of the 73 participants are encoded with a full noun phrase the first time they are mentioned. The other two are encoded with an agreement affix in their first mention.

The following table shows which position or grammatical role in the clause the participants occupy the first time they are mentioned. It is noted that, following Givon, the "Object" category includes only nominals that serve as the direct object complement of the verb, while the "Adjunct" category includes the objects of prepositional phrases and nominal complements of nonactive verbs. This table reflects the fact that almost 84% of participants are introduced into the text in a non-subject position.

TABLE 5: Number of Participants Introduced in Each Grammatical Role

Subject	12 (16.4%)
Object	34 (46.6%)
Adjunct	27 (37.0%)
TOTAL	73 participants



As has been mentioned, there are four different levels of referential encoding in Sio: a full noun phrase, a free pronoun, an agreement affix, or zero encoding. The following table summarizes the frequency of use of the different encoding strategies, both by animacy and by grammatical role. As noted in Chapter 2, the reason for the gaps in the row showing the totals for inanimate noun phrases is that in Sio only a noun phrase or zero anaphora is allowed to encode inanimate non-subjects. Thus, for example, neither the third person plural pronoun /*kinzi*/ "they" nor the corresponding affix /-*nzi*/ can refer to canoes or any other inanimate non-subject referent.

This table reveals that while the majority of subjects (70.2%) are encoded only by an agreement affix, clausal objects and adjuncts are much more likely to be referred to using full noun phrases (58.4% and 75.7%, respectively). A significant number of objects (33.5%) contain zero encoding, so that the listener must infer the referent's identity from the context.

TABLE 6: Frequency of Use of Nominal Encoding Strategies

ANIMACY	GRAMMATICAL ROLE											
	NP	Subject			NP	Object			NP	Adjunct		
		Pron	Affix	Zero		Pron	Affix	Zero		Pron	Affix	Zero
Animate	61	51	317	11	27	3	11	16	16	3	8	13
Inanimate	22	0	27	1	74	0	0	42	62	0	0	1
TOTALS	83	51	344	12	101	3	11	58	78	3	8	14
	16.9%	10.4%	70.2%	2.5%	58.4%	1.7%	6.4%	33.5%	75.7%	2.9%	7.8%	13.6%

In order to prepare the text for analysis, each clause was charted to identify the various component parts. For Sio, this resulted in a chart with seven columns, entitled Comment, Conjunction, Time Phrase, Subject, Verb, Object, and Adjunct. A sample page of this 52-page chart is included in Appendix C.

With this background information in mind, attention is now turned to the analysis of the text using the methodology developed by Givon, as well as adjustments to that methodology suggested by Tomlin.

### Givon's Recency/Distance Method

As was explained in the first chapter, Givon's method includes measuring three different variables to assess the ways in which a language enables speakers to refer to participants in a text. These variables include referential distance, potential interference, and persistence. The results of the analysis of the Sio text on each of these variables are presented below.

#### A. Referential Distance

This feature, which is also called "look-back", measures the distance in number of clauses back to the most recent previous mention of the participant. Givon arbitrarily limits the upper value of this measure to 20; thus each reference to a participant receives a look-back value of from 1-20, with the smaller numbers indicating a more recent mention. The first time the participant is mentioned in the text, a value of 20 is assigned. The model predicts that the higher the referential distance, the more likely it is that the speaker will use more material to encode the referent. In Sio, a participant who has been off-stage for a number of clauses would be expected to be reintroduced with a full noun phrase, while one who is on-stage would be expected to be encoded simply with an agreement affix, or even with zero anaphora.

There are at least two factors in Sio which may indirectly affect the measurement of referential distance. First, one of the prominent participants (or "VIP's" to use Levinsohn's term) in the story, the creature Nondi, is almost always referred to by his proper name. The classification of proper names can be problematical in text analysis, but for this study a proper name is considered to be a full noun phrase. It may be that, in references to this particular participant, the proper name is sometimes used when an affix might normally have been used.

The second factor is that the Sio discourse feature of repetition of clauses (which marks discontinuities and highlighting -- see previous chapter) occasionally results in the use of a noun phrase for a participant who is already well-established and on-stage. In cases where the clause being repeated contains a referring expression encoded with a full noun phrase, the referential distance values are skewed by this repetition.

Nevertheless, the text can easily be analyzed for referential distance using Givon's method, and the following three tables contain the results of this analysis of the Sio text. There are two numbers in each box; the one in italics (labeled 'N') refers to the number of tokens found in the text, and the other number (in bold face, labeled 'R.D.')

The results presented in Table 7 clearly support the hypothesis that increasing referential distance correlates with additional coding material. For all subjects combined, as well as for animate and inanimate subjects measured separately, noun phrases > pronouns > agreement affixes > zero encoding in terms of the "look-back" value (where '>' means "is greater than"). (Note that none of the results presented in this paper have been tested for statistical significance.)

TABLE 7: Referential Distance for Subject Position

ANIMACY	AMOUNT OF CODING MATERIAL				TOTAL
	Noun Phrase	Pronoun	Affix	Zero	
Animate Subject	<i>N=61</i> R.D. = <b>5.34</b>	<i>N=51</i> R.D. = <b>2.00</b>	<i>N=317</i> R.D. = <b>1.37</b>	<i>N=11</i> R.D. = <b>1.09</b>	440
Inanimate Subject	<i>N=22</i> R.D. = <b>10.68</b>	<i>N=0</i>	<i>N=27</i> R.D. = <b>1.78</b>	<i>N=1</i> R.D. = <b>1.00</b>	50
All Subjects Combined	<i>N=83</i> R.D. = <b>6.76</b>	<i>N=51</i> R.D. = <b>2.00</b>	<i>N=344</i> R.D. = <b>1.40</b>	<i>N=12</i> R.D. = <b>1.08</b>	490



For clausal objects, as shown in Table 8 below, the referential distance for referents encoded by noun phrases is substantially higher than that for lower levels of encoding. (As was noted earlier, inanimate non-subject referents cannot be encoded with an agreement affix, and plural animate non-subject referents cannot be encoded with zero anaphora. Therefore for this and all other tables referring to non-subject roles, the "Affix" and "Zero" columns have been combined.)

The results for adjuncts, reflected in Table 9 on the following page, parallel those found for clausal subjects, with each level of increased encoding being matched by an increased average referential distance.

TABLE 8: Referential Distance for Object Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Object	<i>N</i> =27 R.D. = <b>9.93</b>	<i>N</i> =3 R.D. = <b>2.33</b>	<i>N</i> =27 R.D. = <b>1.44</b>	57
Inanimate Object	<i>N</i> =74 R.D. = <b>13.45</b>	<i>N</i> =0	<i>N</i> =42 R.D. = <b>1.71</b>	116
All Objects Combined	<i>N</i> =101 R.D. = <b>12.51</b>	<i>N</i> =3 R.D. = <b>2.33</b>	<i>N</i> =69 R.D. = <b>1.61</b>	183

#### B. Potential Interference

This second factor in Givon's method measures the disruptive effect of other referents in the immediately preceding text upon participant identification. Thus, if a sequence of clauses refers to two or more participants of similar animacy, number, etc., the speaker might need to use a higher level of encoding to ensure that the hearer can identify the referent. By contrast, the

TABLE 9: Referential Distance for Adjunct Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Adjunct	<i>N</i> =16 R.D. = <b>8.69</b>	<i>N</i> =3 R.D. = <b>3.67</b>	<i>N</i> =21 R.D. = <b>1.95</b>	40
Inanimate Adjunct	<i>N</i> =62 R.D. = <b>16.66</b>	<i>N</i> =0	<i>N</i> =1 R.D. = <b>1.00</b>	63
All Adjuncts Combined	<i>N</i> =78 R.D. = <b>15.03</b>	<i>N</i> =3 R.D. = <b>3.67</b>	<i>N</i> =22 R.D. = <b>1.91</b>	103

absence of one or more potentially confusing participants in immediately preceding clauses enables the speaker to use less coding material.

Givon's method involves assigning a value of one if there are no potentially disruptive referents in the immediately preceding three to five clauses, and a higher value if there are such referents. In the literature, most researchers assessing this variable simply assign a value of two if there are one or more potentially disruptive referents, and that is the strategy followed in this study.

In identifying the presence of "interference", especially for non-subject positions, it is important to consider the selectional restrictions of verbs (see Chapter 2). In Sio, for example, the verb /*ra*/ "strike" requires that the object of the action be a drum. So even if there are other inanimate participants in the immediately preceding clauses that are capable of being struck, there would be no potential trouble in identifying a drum as the correct referent. This allows the speaker to delete the object altogether, as is done in several clauses of the text. Other verbs with

clear selectional restrictions also exist in the text, such as /**tapa**/ "pull a bulky item", which requires a direct object capable of being pulled but only with some exertion, as a sail on a canoe.

In this particular text, it is noted that the two brothers are almost always mentioned together, most often with the pronoun /**kinzi**/ "3p" or /**kinzi rua**/ "3p two/both of them", though occasionally with the abbreviated construction /**rua**/ "two". In contrast, the participants with whom they interact most frequently, and always one at a time (Nondi, their mother, and their older brother) are singular in number. Because of this, there may be less potential interference in this text than there might be in a text with multiple participants of similar number.

The following three tables show the results of the assessment of potential interference among the various participants in the Nondi text. As with the tables showing referential distance, the italicized numbers are the number of tokens found in the text ('N'), and the bold-face numbers are the values for interference ('Interf').

Table 10 (see following page) shows the results for subjects. These results suggest that for animate clause subjects a noun phrase is more likely to be used than the other possible levels of coding when there are potentially confusing referents in the immediately preceding text. However, the differences among the values assigned to the other three coding levels are very small (1.29, 1.26, and 1.27). This leads to the conclusion that in cases where a full noun phrase is not used to disambiguate an animate referent in a potentially confusing portion of the text, none of the alternative levels of coding is preferred over the others. With regard to inanimate subjects, no definitive statements can be made.

The results of the analysis for the object position (Table 11 below) are less clear. It does appear that a noun phrase is more likely to be used for direct objects in situations of potential confusion than is the lowest level of encoding, but it is impossible to draw further conclusions.



TABLE 10: Potential Interference for Subject Position

ANIMACY	AMOUNT OF CODING MATERIAL				TOTAL
	Noun Phrase	Pronoun	Affix	Zero	
Animate Subject	<i>N=61</i> Interf = <b>1.43</b>	<i>N=51</i> Interf = <b>1.29</b>	<i>N=317</i> Interf = <b>1.26</b>	<i>N=11</i> Interf = <b>1.27</b>	440
Inanimate Subject	<i>N=22</i> Interf = <b>1.50</b>	<i>N=0</i>	<i>N=27</i> Interf = <b>1.56</b>	<i>N=1</i> Interf = <b>1.00</b>	50
All Subjects Combined	<i>N=83</i> Interf = <b>1.45</b>	<i>N=51</i> Interf = <b>1.29</b>	<i>N=344</i> Interf = <b>1.28</b>	<i>N=12</i> Interf = <b>1.25</b>	490

TABLE 11: Potential Interference for Object Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Object	<i>N=27</i> Interf = <b>1.30</b>	<i>N=3</i> Interf = <b>1.33</b>	<i>N=27</i> Interf = <b>1.04</b>	57
Inanimate Object	<i>N=74</i> Interf = <b>1.53</b>	<i>N=0</i>	<i>N=42</i> Interf = <b>1.12</b>	116
All Objects Combined	<i>N=101</i> Interf = <b>1.47</b>	<i>N=3</i> Interf = <b>1.33</b>	<i>N=69</i> Interf = <b>1.09</b>	173

For adjuncts, the results (see Table 12 on the following page) are in the expected direction: for animate and inanimate adjuncts, the higher the average potential interference, the more coding material is used.

TABLE 12: Potential Interference for Adjunct Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Adjunct	<i>N</i> =16 Interf = <b>1.38</b>	<i>N</i> =3 Interf = <b>1.33</b>	<i>N</i> =21 Interf = <b>1.05</b>	40
Inanimate Adjunct	<i>N</i> =62 Interf = <b>1.45</b>	<i>N</i> =0	<i>N</i> =1 Interf = <b>1.00</b>	63
All Adjuncts Combined	<i>N</i> =78 Interf = <b>1.44</b>	<i>N</i> = <sup>3</sup> Interf = <b>1.33</b>	<i>N</i> =22 Interf = <b>1.05</b>	103

### C. Persistence

The third factor in Givon's method is what he calls persistence, or "decay". It simply is a measure of the number of subsequent clauses, without interruption, in which a participant plays a role. When a participant is on-stage continuously, the persistence value will be high, whereas a participant who has no role in the immediately following clause does not persist (a case of "immediate decay") and receives a value of zero.

A potential problem with this measurement, as exemplified in several places in the Sio text, is that some participants appear frequently, but often with two or three clauses inserted between their appearances. Although the assigned values would suggest that the participant does not persist, the situation is in fact quite the opposite. For example, in Clauses 352-363, the two brothers are testing out various parts of the quartz skin to see which part produces the best noise for a drum. The following is a literal translation of a few clauses in this section:

355 They struck the abdomen,  
356 but it didn't resonate.  
357 They struck the head,  
358 but it didn't resonate.

359 They struck the back,  
360 but it didn't resonate.

In each of the three clauses that have the brothers as the subject, the persistence value must be zero, since they have no role in the immediately following clause. Yet they clearly persist through this section of the text, being absent only for a single clause each time.

The following three tables present the results of the analysis of persistence ('Perst') among the various grammatical roles of the Sio text. The results presented in Table 13 below serve to dramatically establish the persistence of animate over inanimate subjects; inanimate subjects tend to decay almost immediately, while animate subjects last for an average of five to nine clauses. In fact, the combining of animate and inanimate subjects in the third row totally masks this dramatic distinction. Yet within the various levels of coding for animate subjects, there seems to be no consistent pattern, as referents with zero anaphora actually persist longer than those encoded with a noun phrase. Thus for animate subjects it seems not to matter how much coding material one uses.

It also is noteworthy that subjects encoded with pronouns persist far longer than do similarly-encoded non-subjects (see Tables 13-15), which may indicate that pronoun encoding of a subject reflects that participant's prominence in the text.

TABLE 13: Persistence for Subject Position

ANIMACY	AMOUNT OF CODING MATERIAL				TOTAL
	Noun Phrase	Pronoun	Affix	Zero	
Animate Subject	<i>N=61</i> Perst = <b>6.56</b>	<i>N=51</i> Perst = <b>8.73</b>	<i>N=317</i> Perst = <b>5.35</b>	<i>N=11</i> Perst = <b>7.36</b>	440
Inanimate Subject	<i>N=22</i> Perst = <b>0.41</b>	<i>N=0</i>	<i>N=27</i> Perst = <b>0.22</b>	<i>N=1</i> Perst = <b>0.00</b>	50
All Subjects Combined	<i>N=83</i> Perst = <b>4.93</b>	<i>N=51</i> Perst = <b>8.73</b>	<i>N=344</i> Perst = <b>4.95</b>	<i>N=12</i> Perst = <b>6.75</b>	490



Very similar results are obtained for objects as were obtained for subjects, as may be seen in Table 14; animate objects persist dramatically longer than inanimate objects, the amount of coding material for objects seems to be irrelevant in measuring persistence, and combining animate and inanimate objects together obliterates the importance of animacy.

TABLE 14: Persistence for Object Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Object	<i>N</i> =27 Perst = <b>8.30</b>	<i>N</i> =3 Perst = <b>2.67</b>	<i>N</i> =27 Perst = <b>6.41</b>	57
Inanimate Object	<i>N</i> =74 Perst = <b>0.68</b>	<i>N</i> =0	<i>N</i> =42 Perst = <b>0.29</b>	116
All Objects Combined	<i>N</i> =101 Perst = <b>2.71</b>	<i>N</i> =3 Perst = <b>2.67</b>	<i>N</i> =69 Perst = <b>2.69</b>	173

The results for the adjuncts, shown on Table 15 below, again parallel those obtained for subjects and objects. Animate adjuncts persist, while inanimate adjuncts quickly decay.

TABLE 15: Persistence for Adjunct Position

ANIMACY	AMOUNT OF CODING MATERIAL			TOTAL
	Noun Phrase	Pronoun	Affix/Zero	
Animate Adjunct	<i>N</i> =16 Perst = <b>4.81</b>	<i>N</i> =3 Perst = <b>2.67</b>	<i>N</i> =21 Perst = <b>2.14</b>	40
Inanimate Adjunct	<i>N</i> =62 Perst = <b>0.32</b>	<i>N</i> =0	<i>N</i> =1 Perst = <b>0.00</b>	63
All Adjuncts Combined	<i>N</i> =78 Perst = <b>1.24</b>	<i>N</i> =3 Perst = <b>2.67</b>	<i>N</i> =22 Perst = <b>2.05</b>	103

In summary, while Givon's method does produce interesting statistical norms, there is no procedure for evaluating and interpreting those instances when an "unexpected or surprising" amount of coding material is used. There is no mention of alternative explanations for over-coding (such as points of discontinuity in the text, or speaker highlighting) or under-coding (such as participant prominence). Further, the factors of potential interference and persistence do not seem to have been very useful, and in fact Givon did state in a later work that referential distance is the factor that seems to contribute the most to explaining the scalar nature of topicality (1990:912). Thus the method leaves much to be desired.

#### Tomlin's Episode/Paragraph Method

A summary of Tomlin's method was presented in Chapter 1. One of his main objections to the model promoted by Givon is that the referential distance factor fails to explain counterexamples (i.e., references where the amount of linguistic coding material used is more or less than what would be predicted by the number of clauses since the previous mention of the referent). Tomlin feels that these counterexamples may be clearly accounted for if the position of the referring expression within thematic episodes or paragraphs is considered.

Tomlin's study compared the use of full nouns (as an example of the highest amount of coding material) with the use of pronouns (representing low amounts of coding material). While this was necessitated by his use of English as the language of analysis, for this analysis of participant reference in Sio, it seems more productive to compare the use of full noun phrases with the use of agreement affixes and zero anaphora.

A key starting point for Tomlin's method is to divide the text into "episodes" or "conceptual paragraphs", which are best identified by evaluating sustained attention on a particular paragraph-level theme. An episode boundary occurs when attention shifts in the flow of information, and these boundaries often correspond with major changes in time, place, and participants.

Once the boundaries between episodes are determined, the amount of linguistic material used to identify the participant(s) in a new episode is studied, and each token is labeled as a "hit" or a "miss". Tomlin assumes that speakers will use a full noun for the first mention of a participant after an episode boundary, and will use less coding material to sustain reference to a participant within an episode. Therefore a "hit" is a referring expression that meets this expectation, while a "miss" is one that does not.

Since Tomlin's method requires that episode boundaries should be determined by attention to themes, and not by analyzing linguistic cues, the process is quite subjective (Tomlin concedes that episode boundaries are not always easily identified). In identifying the episodes for this analysis, for those cases where an episode break seemed reasonable but the precise location of the boundary was not clear, intonation contours were used to make the determination. Sio narrative texts contain periodic long pauses, immediately preceded by distinctly falling intonation. Both the pauses and the falling intonation are more pronounced than are similar features found at sentence boundaries, so these locations were used as episode boundaries when thematic information was insufficient.

Before proceeding with the reporting of results in this study, it should be reiterated that Tomlin's study was carefully structured, in that subjects were given a series of slides for which they were to make descriptive statements. Thus the researcher actually defined the episode boundaries for the subjects. In this present study, the storyteller has no such cues, and thus episodic boundaries are natural and much less easy to identify.

The episode boundaries in the Sio text that have been pre-determined are listed in Appendix B, with a brief description of the action that takes place within the episode. These boundaries serve as the key reference points for determining whether Tomlin's method accounts for those referring expressions that have been over- or under-coded.



In addition to determining episode boundaries, a concurrent step was to identify all of the tokens in the text which had either more linguistic coding material than would be expected based on the number of clauses since the referent's last mention, or less coding material.

#### A. Over-coded Nominals

If the referential distance is three clauses or less, and a full noun is used, the token is considered to be over-coded. With such a relatively short distance since the last mention of the referent, a full noun is surprising (normal coding would be an affix or zero) and requires some explanation.

Following Tomlin, any token for which the referent was determined to be ambiguous was removed when the totals were being counted. The underlying reasoning for this decision is closely parallel to Givon's concept of "potential interference". Ambiguity was determined to exist when there was more than one plausible referent mentioned in the text since the most recent mention of the participant (plausibility takes animacy and number into account, as well as selectional restrictions of certain verbs). The over-coding was assumed in these situations to be the speaker's attempt to disambiguate the referent's identification by the hearer.

When the text was analyzed, it was determined that 28 subjects were over-coded (this represents 33.7% of the 83 subject noun phrase tokens). Of this total, only six were classified as "hits", being the first mention of the participant in a new episode (even with slight adjustments being made in the episode boundaries). The remaining 22 tokens are unexplained by the episode boundary hypothesis. For the direct object tokens, a total of 14 (14.9% of the 101 object noun phrase tokens) were determined to be over-coded, but only two of these (Clauses 197 and 370) were "hits", so the other 12 remain unexplained. A total of six of 78 adjunct noun phrase tokens (7.7%) were over-coded, and four of these (Clauses 8, 281, 568, and 577) were classified as "hits" (allowing for occasional, slight ambiguity in the pre-determined boundaries). Thus for the 48

noun phrase tokens in the text that had more coding material than expected, Tomlin's episode boundary explanation accounted for only 12 of these, or 25.0%.

### B. Under-coded Nominals

An under-coded nominal is one which is represented by an agreement affix or zero anaphora, implying a referent that has recently been mentioned in the text, but which has an actual referential distance of three or more clauses. Tomlin's explanation would be that these referents have already been mentioned somewhere in that particular episode, and this factor accounts for the low amount of coding material despite the relatively high referential distance.

In the Sio text, only 26 tokens were under-coded. Fourteen of these were subjects, and in eight cases a "hit" was counted, since it was not the first mention of the participant in that particular episode. The percentage of hits was higher for non-subject tokens; nine of 10 under-coded objects were hits, as was one of the two under-coded adjuncts.

The following table presents a summary of the results of this analysis for both over-coded and under-coded tokens.

TABLE 16: Over-Coding and Under-Coding with Reference to Episode Boundaries

CLAUSE POSITION	CODING					
	Over-Coding			Under-Coding		
	Hit	Miss	Total	Hit	Miss	Total
Subject	6	22	28	8	6	14
Object	2	12	14	9	1	10
Adjunct	4	2	6	1	1	2
Totals	12	36	48	18	8	26

Perhaps in the case of under-coding, the "misses" represent a more intriguing area of inquiry than the "hits". Since this is the first mention of that participant in a certain episode, why is the amount of coding material not higher (at least a pronoun, if not a full noun phrase)?



Of course, it is possible that the results of this utilization of Tomlin's method have been affected by improperly determined episode boundaries. In the majority of instances, however, there are other factors at work, such as the status of participants, some higher-level grouping of episodes, highlighting, repetition, etc. This will be discussed further in subsequent sections.

In general, then, Tomlin's method was not effective in explaining deviant levels of encoding. The pre-determining of episode boundaries, which is a crucial part of the method, is at best a subjective guess on the part of the researcher, and when this was done for the Sio text, few of the deviant tokens were explained by their presence at specific points within episodes. The value in the approach is that it at least goes beyond the straight statistical calculations of Givon in attempting to explain the form of individual tokens; yet its shortcoming is that it, too, does not fully describe the motivation for speaker choices in encoding referring expressions.

#### Initial Evaluation of Both Methods

It has been shown that both the "Recency/Distance" method developed by Givon and the "Episode/Paragraph" method of Tomlin have some utility in analyzing participant reference within narrative texts. Givon's methodology shows that referential distance, potential interference, and persistence in the text all have an influence on how much linguistic material is used to encode a nominal, although only referential distance was found to be useful in this particular study. Tomlin provides a starting point for explaining why certain nominals don't seem to fit the expected pattern regarding coding material and referential distance, pointing out that episode boundaries within the text can intervene and account for at least some of the counterexamples.

However, there are additional factors that still are not addressed. For example, does the status or prominence of the referent somehow affect the coding decisions a speaker makes? What about situations where a cycle of events, or at least a clause of the text, is being repeated, or



where the speaker is highlighting the participant for some reason? Is there some other methodology that will enable an even more precise accounting of counterexamples with regard to participant encoding? In the following chapter, the "Default/Marked" method developed by Levinsohn is applied to the Sio text, and it will be shown that his method does indeed provide more satisfactory and comprehensive explanations of the tokens that still are not explained.

## CHAPTER 4

### ANALYSIS OF THE TEXT USING THE DEFAULT/MARKED METHOD

In the preceding chapter, methodologies developed by Givon and by Tomlin were applied to the Sio text, to determine their utility in accounting for participant reference. They both provided explanations for a portion of speaker decisions regarding the amount of coding material to be used, although neither accounted for a number of counterexamples in the data. In the present chapter, the methodology developed by Levinsohn is applied to the same text.

As was explained in Chapter 1, Levinsohn's "Default/Marked" method identifies several different possible contexts for each subject or non-subject token in a text. Subject tokens are classified as **INTRO** (for the first mention of the participant), **S1** (the same referent was the subject of the immediately preceding clause), **S2** (the subject was the addressee of speech reported in the previous clause), **S3** (the subject was in the previous clause in a non-subject role), or **S4** (other contexts). Each non-subject token in the text is assigned to one of the following contexts: **N1** (the referent had the same non-subject role in the previous clause), **N2** (the referent was the speaker of speech reported in the previous clause), **N3** (the referent was in the previous clause in a different role), or **N4** (other contexts). Once tokens are classified, charts are developed to show the amount of coding material used for tokens within each context.

The next step in Levinsohn's method is to propose a default encoding for each context, based on an inspection of the data. Once this is determined, tokens that have received encoding that is different from the default encoding for that context are examined, to determine why more or less coding material than would be expected was used for that reference. Modifications are then made to the context definitions to reflect the data. Finally, each remaining token that

deviates from the default encoding is analyzed to determine the motivation for that particular "special" encoding.

Levinsohn's methodology assumes that over-coding of tokens is almost always done either to reflect textual discontinuities (including shifts in time, scene, participants, and action) or to highlight the participant and/or action to which reference is being made. Under-coded tokens are explained by the participant's local or global prominence, the absence of other participants on the stage, or by the fact that material is being repeated (Dooley and Levinsohn 1999:66-67). As the results presented in this chapter will show, these assumptions are very strongly supported in the analysis of the Sio text, as practically all deviant tokens are thus explained. The VIP strategy, along with the default levels, provide an excellent framework for analyzing the text.

As an aid in tracking the use of demonstratives in relation to the prominence of the participants and props (see pp. 18-19 of this paper), the following framework is presented. The creature Nɔndi is the first Global VIP, starting with the first clause of the actual narrative (Clause 10), and continuing until his death in Clause 324. The older brother then becomes the Global VIP upon his reactivation in Clause 370, as will be demonstrated later. The twin sons are marked as Local VIPs by the demonstrative /ɲinde/ in Clauses 57, 346, and 425, and they retain this status until attention either reverts to the Global VIP (as in Clauses 197 and 370), or turns to another Local VIP (see Clauses 141 and 504).

In the following discussion, the results of the analysis of the Sio text will be presented in two separate sections. Results obtained for subject tokens will be presented first, followed by the results obtained for non-subject tokens.

### Subject Reference

The tabulation of data for both subject and non-subject reference was conducted in two stages. First, the tokens were classified strictly according to the contexts as defined by



Levinsohn. Then, as suggested in his methodology as "Step 7", the contexts were slightly redefined where appropriate, to more precisely reflect the features of Sio participant reference.

The results of the initial counts for subject tokens are presented in Table 17 below:

TABLE 17: Coding Material used for Subjects in Different Contexts: Initial Analysis

AMOUNT OF CODING MATERIAL					
CONTEXT	Noun Phrase	Pronoun	Affix	Zero	TOTAL
INTRO	5	--	1	--	6
S1	7	13	256	19	295
S2	5	1	9	1	16
S3	13	5	31	4	53
S4	36	33	34	1	104

A preliminary analysis of the numbers in Table 17 shows that for two of the five contexts, a default encoding is readily apparent. For the **INTRO** context, five of the six tokens (83.3%) are encoded with full noun phrases, and for the **S1** context, 256 of the 295 tokens (86.8%) are encoded with an affix. In an additional two contexts, a less-convincing case can be made: for the **S2** context, nine of the 16 tokens (56.3%) are agreement affixes, and for the **S3** context, 31 of the 53 tokens (58.5%) also are encoded with an affix. However, for the **S4** context, the results are not clear; while 36 of the 100 tokens (36%) are encoded with noun phrases, an almost equal number of tokens are encoded with agreement affixes (34) or pronouns (33). Thus the first step will be to resolve the apparently conflicting data in the **S4** context.

The **S4** context is a rather broad one, encompassing any context other than those covered by the other four. It generally contains references in which the participant already is on-stage in the text, but has not been mentioned for two or more clauses (suggesting that the referent no

longer occupies center-stage). For this reason, it would be expected that the amount of coding material would be toward the higher end of the scale (a pronoun or noun phrase, rather than an agreement affix), to enable quicker identification by the hearer. On the basis of this expectation, a logical strategy would be to assume the default encoding to be a noun phrase, and to determine why the large number of references with affix encoding are in this context.

A review of the 35 highly under-coded tokens (affix or zero) provides a quick solution to this situation. In a total of 23 of the cases, the subject was also the subject two clauses earlier, but a single (usually brief) clause intervened between the two mentions, thus forcing an S4 context assignment. In each of these cases, there is no possibility of referential ambiguity. For example, consider again Clauses 357 to 361:

357 They (the two brothers) struck the head  
 358 but it didn't resonate.  
 359 They struck the back  
 360 but it didn't resonate.  
 361 Then they wanted...

By strictly adhering to the standard definition of the S4 context, the subject in Clauses 359 and 361 must be classified as S4, even though the intervening clause introduces no possible ambiguity regarding the identity of the referent appearing as the subject two clauses earlier.

In addition to these 23 tokens, there are another four tokens in the S4 context where the subject was also the subject three clauses earlier, and for which there is no other plausible referent that could make identification ambiguous. For example, consider Clauses 424-427:

424 he (the older brother) struck (the drum) on the path  
 425 and those two young men heard (it) from the grasslands.  
 426 The two started running.  
 427 He went down

The subject of Clause 427 (the older brother) must be assigned to the S4 context as originally defined. However, since he also was the subject in Clause 424, and there is no other plausible referent in the intervening clauses, the smaller amount of encoding material is not



problematic. (It may also be that this is an example of VIP encoding for the older brother -- see below.)

For this reason, in doing an analysis of Sio texts, a new context, S1A, has been created, to include subjects that appeared in the subject role within the preceding two to three clauses, provided that there is no other plausible referent in the intervening clauses since the preceding mention. This results in the 27 subject tokens mentioned above being moved from S4 to S1A (see Table 19, next page, for a revised summary), so that for S4, a total of 36 of the 77 tokens (46.8%) are encoded with noun phrases. (It is noted that in all but six of these 27 tokens, the number of the under-coded subject -- singular or plural -- is different from the number of the intervening subject[s], which makes accurate identification of the referent even more likely.)

The fact that an additional 33 of the 77 S4 tokens (42.9%) are encoded with pronouns also needs to be addressed. Specifically, it must be determined what causes the speaker to use a pronoun so frequently in the S4 context, since in no other context are pronouns so frequently used (Table 17 reveals that fewer than 10% of the tokens in any of the other contexts are pronouns). This issue will be covered in the discussion of the "deviant" tokens found in the various contexts.

Therefore the encoding levels presented in Table 18 on the following page are proposed as the default levels for the six contexts to which subject tokens may be assigned. Levinsohn (personal communication) predicts that the default encoding will never decrease as one moves from Context S1 to S4, and Sio is consistent with this prediction. Table 19, which immediately follows Table 18, presents the revised numbers of tokens in each context. An asterisk indicates non-default encoding that must be accounted for, and default numbers are in bold face.

An analysis of each of the deviant tokens within the various contexts is presented following Tables 18 and 19 (note that the S1A context is not discussed, since all tokens received the default encoding level). The question to be answered is this: why did the speaker not use the default encoding level? As will be shown, Levinsohn's theory does provide an answer.



It should be noted that a few of the deviant tokens may simply be due to speaker error or fatigue; the story was told orally in a single session, with almost no back-ups or self-corrections heard on the tape. The fact that so few errors were made in such a long text is rather remarkable.

TABLE 18: Default Encoding Levels for Subjects, by Context

CONTEXT	DEFAULT ENCODING
INTRO	Noun Phrase
S1	Agreement Affix
S1A	Agreement Affix
S2	Agreement Affix
S3	Agreement Affix
S4	Noun Phrase

TABLE 19: Coding Material used for Subjects in Different Contexts: Final Analysis

AMOUNT OF CODING MATERIAL.					
CONTEXT	Noun Phrase	Pronoun	Affix	Zero	TOTAL
INTRO	5	--	1*	--	6
S1	7*	13*	256	19*	295
S1A	--	--	27	--	27
S2	5*	1*	9	1*	16
S3	13*	5*	31	4*	53
S4	36	33*	7*	1*	77

Following Levinsohn's theory, the discussion of deviant tokens will be organized around the most common reasons given: highlighting or discontinuities for over-coded tokens, and

prominence or repetition for under-coded tokens. Other reasons will be introduced when necessary.

#### A. INTRO Context

The only deviant token is in Clause 1, where the speaker says "I am telling a story...", but uses the first person singular subject agreement affix rather than a full noun phrase, as is used when every other participant is introduced. The Default/Marked method applies only to third-person referents, and since the speaker's identity was obvious to the hearer at the time the story was being told, this token can be removed from the analysis.

#### B. S1 Context

The seven deviant S1 noun phrase tokens are analyzed below; six are explained by highlighting, and the other one by its presence at a point of discontinuity.

Highlighting -- In Clause 7, a full noun phrase ("that man's skin") is used for the subject which was the same subject in the preceding clause, also encoded as a noun phrase ("that skin"). This probably is due to highlighting being placed on the quartz-like quality of the skin. In Clause 157, the rat is the uninterrupted subject of Clauses 156-166, so it is unexpected that it would be encoded with a full noun phrase ("the rat") in each of the first two clauses in this section. This over-coding may be the speaker's attempt to slow down the story line as it approaches the first climax (the theft of the piece of skin). In Clause 248, the subject ("those spears and arrows") also appeared as the subject two clauses earlier ("the spears and arrows"). This may be to highlight the contra-expectation of sharpened weapons being unable to penetrate the creature's skin.

Clause 320 occurs at the conclusion of the scene where the creature is killed. Like Clause 157, this over-coding appears to be a device to slow down the story line as the climax of the creature's death is reached. In Clause 463, the subject noun phrase ("all the various birds") is over-coded, since the birds were the subject of the preceding clause, where only an affix was

used. Its use here seems to slow down the story line as the climax of their failure to conquer the elements is reached. Finally, in Clause 476, the noun phrase "the birds called nzoŋi birds", refers to the same subject as in the preceding clause ("the nzoŋi birds"). This probably is being done to highlight the fact that it is *these* birds, rather than the others, who will move the story forward.

Discontinuity -- Clause 266 seems to represent a parenthetical statement that provides background information to explain what is happening. After two events in which the creature magically turns food fragments into food, the speaker seems to want to clarify that the creature had no food of its own. Since this clause is thus off the main event line, this is a point of discontinuity or "world shift" (Clancy), so the full noun phrase is appropriate.

As for the 13 pronoun tokens in the S1 Context, attention is first given to the nine tokens where the pronoun used is /*rua*/ "two", which is an abbreviated form (0 + ADJ) of the usual third person plural (dual) pronoun /*kinzi rua*/ "they two" (PN + ADJ). This form occurs in spoken Sio far more frequently than any other third person pronominal form (when the referent is dual). As partial evidence, the S1 context in the Sio text "The Kulambi Man" (to which more reference will be made in Chapter 5), does not have a single S1 token that is a pronoun (out of 144 tokens). It is noteworthy that there is no pair of participants acting together anywhere in that text. Thus it may be that the high number of deviant S1 pronoun tokens in this text is an artifact of the reality that among the major participants is a set of twins who almost always act together.

Nevertheless, the tokens need to be considered. Four of these tokens occur in one small segment of text, in Clauses 128, 130, 133, 136. There are no other animate participants in 128-136, so the overuse of the form /*rua*/ in this short segment may simply be stylistic. However, in one other token (Clause 118) there is a special type of subject combination occurring, where the plural subject of the clause includes a combination of the singular subject referents of the



preceding clauses. This attempt to disambiguate the referents appears to explain the apparent over-coding of this token.

The remaining four tokens of /**rua**/ (Clauses 110, 351, 426, 578) will be considered along with the other four pronoun tokens below. Most of these examples of over-coding are explained by their presence at points of discontinuity.

Highlighting -- The one example of highlighting is in Clause 426. Having been encoded in the preceding clause (425) with a full noun phrase ("those two young men"), it is surprising to see /**rua**/ "two" as the subject in 426. This redundant encoding may be to slow down the story and increase the tension.

Discontinuity -- In Clause 8, the use of a pronoun instead of just an affix probably reflects the fact that when a stative clause has no verb, and lacks a conjunction at the beginning, an overt subject is obligatory. Since the speaker here chose a pronoun instead of a noun phrase, he has selected the lowest possible amount of encoding material for this token.

Clause 36 should be interpreted as part of a contrastive construction that goes from 36 to 42. The pattern is, "They did X; as for her, she did Y." The use of pronouns in Clause 36 and in Clause 39 (to be discussed later under S4) is probably what Levinsohn refers to as "anticipatory points of departure" (1999:33-35), in which the overt reference to "they" in 36 anticipates the shift of reference to "she" in 39.

In Clause 111, the presence of a time phrase signals a temporal discontinuity, which accounts for the full repetition of the extended noun phrase found in Clause 108 (and referred to above).

Clause 351 begins a sequence where the twins methodically test out various portions of the creature's skin to determine its usefulness as a drumhead. The preceding clause is a generic statement, while this clause cites a specific example -- an example of action discontinuity

(Levinsohn 2000:5). In Clause 515, the use of /**kinzi**/ "they" seems to be an attempt to disambiguate the identity of the referent. The subject of 515 is not the addressee of the quote in 510-514, as would be expected, so it may be that the identity of the subject is being made explicit. Alternatively, it could be a background comment (action discontinuity), as the speaker reiterates the purpose of the speech. Finally, Clause 578 also appears to be a point of action discontinuity in the text, as the speaker reiterates and amplifies the statement made in 576.

The final category of apparently deviant tokens involves 19 which contain zero anaphora. Nine of these tokens, each of which represents under-coding in light of the default level for this context, can be explained by what might be called a "part/whole" relationship. In these cases, one of the referent's body parts is presented as the subject (using a noun phrase), but it merely represents the participant, so should not be considered to be different from the participant (see Hopper and Thompson, 1984, for further discussion of the treatment of body parts as being undifferentiated from their possessors). An example would be in Clauses 220-223:

220 He (Nɔndi) searched and searched,  
 221 but he couldn't find it.  
 222 Therefore he reached out his hand (literally, "his hand reached out")  
 223 and took a leaf from a breadfruit tree

It is apparent that the hand in Clause 222 merely represents an extension of the creature, so an S1 designation is appropriate even though the body part is encoded (as it must be) by a noun phrase. So, rather than considering these to be examples of over-coding, it is suggested that they actually are special cases of zero-encoding, where the S1 referent is manifested by zero, and the body part is introduced as an extension of its possessor.

The other 10 zero-encoded tokens are found in the following clauses: 40, 140, 218, 237, 290, 348, 466, 517, 616, and 640. Nine of these are idiomatic constructions that contain a conjunction and an adjunct (usually the negation word /**tia**/, or the word /**marumbu**/ "finished"). The hearer is expected to fill in all of the content, and it usually involves something like "the

expected event didn't happen", or, "that was the end of it". Essentially, these are sentence fragments, lacking a main verb.

The only other token in this category is Clause 348, which has a similar conjunction-plus-adjunct construction, but the adjunct is an adjectival complement of the understood subject, the creature's body: /aku sɔmbu/ "and rotten", which is understood to say, "and the creature's body became rotten".

To summarize the findings regarding S1 subjects, it is confirmed that the default encoding level is an agreement affix. Noun phrases may be used for highlighting or at points of discontinuity; pronouns (which may be text-specific) are usually used at points of discontinuity; and zero anaphora is limited to sentence fragments in which most of the content must be supplied by the hearer.

### C. S2 Context

Since nine (56.3%) of the 16 S2 subject tokens are encoded with an agreement affix, that has been proposed to be the default level. There are seven deviant tokens, and these are discussed below.

Highlighting -- Clause 475 contains a noun phrase that appears to be highlighting that it was those particular birds that responded, rather than the others.

Discontinuity -- Clause 255 appears to mark a clear discontinuity, as the scene changes from the battle to the description of the rest break activities. This would appear to account for the over-coding of this S2 token.

Disambiguation -- For the nine non-deviant S2 tokens, the average number of clauses preceding the token which contains the reported speech is relatively small -- 2.3 clauses. Thus both the speaker and the addressee remain well activated in the hearer's mind, with little chance for



referential ambiguity. However, it seems that a longer stretch of speech encourages the producer of the text to place more coding material into the clause in which the addressee becomes the subject. This essentially reminds the hearer who it was that was being addressed by the longer stretch of speech. Thus Clauses 156 (following a reported speech segment that is 13 clauses long), 185 (nine clauses long), 196 (six clauses long), and 449 (nine clauses long) contain an over-coding for the purpose of clarification (Clause 156 may also be thus encoded as a pre-climax slowing device). With regard to Clause 185, it is noted that this is an example of what Levinsohn calls "countering speech", where the second speaker changes the direction of the conversation. Such speech clauses often take marked encoding (Levinsohn, 2000:231-234).

With regard to the one apparently under-coded S2 token (zero encoding in a context with a default level of affix), reference is made to the previous discussion about body parts representing their possessors. In Clause 556, even though the birds are the addressees, it is their "hands", encoded with a noun phrase, that respond as the subject of the clause that follows the reported speech. As suggested earlier, the S2 referent is manifested by zero, and the body part is introduced as an extension of its possessor.

Therefore the default coding of an agreement affix for the S2 context is supported. It is only in situations of possible referential confusion, or in other exceptional situations as described above, that a different level of encoding is used. (It does not seem profitable to attempt to explain the use of noun phrases vs. pronouns, as there is only one S2 pronoun token in the data).

#### D. S3 Context

The situation for deviant tokens in the S3 context is much like that already described for the S1 context, which also has an agreement affix as the default encoding level. Since an S3 subject also appeared in the preceding clause (though in a non-subject role), the question is why a higher level of encoding is used in some cases.

There are 13 S3 tokens that are noun phrases, and these can be explained by arguments already presented in other contexts:

Highlighting -- The over-coding in Clauses 15 and 28 ('all of the Sio women and men') highlights the inclusiveness of the referent. (An episode break has also been proposed before Clause 15.) In Clause 18, the full repetition of coding material ("that village") between the non-subject position in Clause 17, and the subject position in this clause, seems to highlight the climax -- the peoples' flight to avoid destruction. In Clause 198, the creature's proper name, Nōndi, occurs in the non-subject role in the preceding clause, and in the subject role here. It may be that this is being done to highlight his reestablishment in a dominant role. The same principle may be operating in Clause 273, to highlight the magical power of this creature, and in Clause 363, where the magical power of the skin of the neck is first displayed. The full repetition of "their mother" in Clause 439 from the preceding clause may be to highlight the fact that this is an indirect example of countering speech.

In Clause 455 the focus seems to be that it was all of the birds who responded, not just some of them, and this is highlighted by the use of a noun phrase.<sup>1</sup> Also, the subject of Clause 500 is encoded by the full noun phrase, "those little birds, those very ones", which is unexpected since they appeared as the object of the preceding clause. This appears to be a device to highlight them as the heroes of this section of the story.

A final example of over-coding to highlight is found in Clause 630. In this clause, which follows a brief section where the twins are presented as acting separately and thus are encoded by third person singular agreement affixes on the verbs, there is possible confusion as to the identity of the referent of this subject, who also is third person singular. Thus even though it seems clear

---

<sup>1</sup> While reading the final draft of the thesis, following the defense, Levinsohn noticed that this reference is a pronominal quantifier, not a noun phrase. Since the general argument of the thesis is unaffected by this correction, the text has been left unchanged.



that the older brother is the subject, the speaker over-codes him. This clause also serves as the climax of the episode, which confirms highlighting as the motivating factor for a higher level of encoding.

Discontinuity -- A noun phrase is used to encode the subject of an S3 context in two of the opening clauses of the text: Clauses 2 and 3. Both of these clauses represent speaker comment, which is a form of action discontinuity. Also, in the middle of the first confrontation between the mother and the older brother, the mother betrays the trust of her twin sons and reveals the location of the drum to their older sibling. Clause 416 appears to be off the event line, providing background information as to why the mother did this. Thus the apparent over-coding is explained.

The four tokens representing apparent S3 under-coding (zero instead of an affix) are found in Clauses 244, 279, 317, and 419. As already has been discussed, this utilization of a participant's body part (or, as in Clause 279, water droplets) as the subject/actor in lieu of the whole participant is simply a unique form of default encoding for this context.

As for the five tokens of S3 over-coding that involve pronouns, each case appears to involve some sort of discontinuity. In Clause 94, the use of the abbreviated pronoun form /**rua**/ "two" follows a very complex prepositional phrase in the non-subject position of the previous clause. The pronoun form here may simply re-establish the subjects' identity after this unusually long interruption (or, it may highlight their success). The use of the pronoun /**kinzi rua**/ in Clause 242, and /**rua**/ in Clause 295 is probably due to discontinuities (the switching of attention from Nondi to the brothers).

In Clause 450, the pronoun form /**rua**/ again seems to signal a discontinuity. An episode boundary had been proposed in the previous chapter to exist at Clause 449, but it is equally



plausible that the break begins at 450. A similar explanation may account for the over-coding of the subject in Clause 568, as a break in action appears to be occurring there, despite the earlier assessment that the break occurred at Clause 564 (both may actually be points of discontinuity).

As with the discussion of pronouns in the S1 context, the small number of tokens in the S3 category, combined with the fact that all five tokens involve the use of some form of the unique "they two" pronoun, would support the observation that these tokens occur mainly because of the fact that the subject of many clauses is a set of twins acting as a unit. In fact, the other text, "The Kulambi Man", has only one S3 pronoun token out of 33 tokens in that context.

Thus the agreement affix is clearly established as the default encoding for the S3 context, and exceptions can be explained by the factors proposed by Levinsohn.

#### E. S4 Context

Finally, the apparently under-coded S4 subject tokens need to be examined, to determine why a referent who has been absent from the story line for two or more clauses reappears with merely an agreement affix for encoding (or, in one case, with zero anaphora). As was discussed earlier, a total of 27 of the originally identified 35 tokens with agreement affix encoding have been moved to S1A, as there is no other plausible referent to whom it might be referring (these were mentioned just two or three clauses earlier in the text). The seven remaining tokens in this category, plus the one with zero anaphora, are described below.

Prominence -- For four of the tokens, an interesting explanation is provided by Levinsohn's "VIP" hypothesis; the under-coding is due to the fact that the subject is, either at a local or global level, the most significant participant on stage, so minimal coding is used. The most dramatic example of this is in Clause 549, where the subject is the older brother. Despite the fact that this is the first mention of this participant in 101 clauses, the encoding level is an agreement affix. The assumption appears to be that the identity of this key participant will be clear to the hearer despite

his long absence. Other examples of VIP reference are found in Clause 25 (Nondi); Clause 473 (the twins); and Clause 591 (the older brother).

Other Reasons -- There is one token with a unique explanation. In Clause 449, an idiomatic phrase is used in which the subject is absent and the verb is encoded with a third person singular prefix. The hearer is expected to search back through the text and fill in the missing subject (which appears to be the content of the preceding speech segment, that is, what the mother has just said).

The reason for the last two under-coded tokens seems to be indefinite reference. In Clauses 504-563 (the episode involving the dance on the island where the drum is retrieved) there are multiple participants, both singular and plural. In Clause 504 and 524, the subject is the people of the island (plural), who do not appear to be local VIPs, as the birds are also activated. Thus it is unusual that the coding for these two subjects would be an agreement affix, as it results in a situation of referential ambiguity; who gave the bad drum to the bird (504)? Who struck the good drum (524)? It may simply be that these are examples of indefinite reference, which is frequently observed in normal conversational Sio.

This accounts for the under-coding involving agreement affixes. However, there are another 33 tokens that have not received the default level for the S4 context (a noun phrase), but rather are encoded with a pronoun, which also constitutes a lower level of encoding.

The selection of a pronoun instead of a noun phrase has already been discussed, in both the S1 and S3 contexts. It is noted that in 24 of the 33 S4 pronoun tokens (72.7%), the pronoun used is one of the two forms of the idiomatic third person plural/dual pronoun, /**kinzi rua**/ or /**rua**/. Since only one of 49 S4 tokens in The Kulambi Man text is a pronoun, and this text contains no pair of participants acting together (as the twins do in the Nondi text), it may be assumed that the number of S4 pronoun tokens is atypically high in the present text. It is possible



that a noun phrase would have been used in many of these clauses, had the subject been other than third person plural/dual.

Four of these tokens (Clauses 250, 253, 268, 289) involve a special change of plural subject in which some, but not all, of the subject referents in one clause continue as the subjects of the following clause. For example, in Clause 249 the zero-encoded subject includes the twins plus the creature, whereas the subject of 250 (encoded with the form /*rua*/) is the twins, alone.

Eight of the remaining nine pronoun tokens (those in Clauses 39, 113, 114, 115, 117, 587, 588, and 629) may be explained by their presence in a contrastive construction. These are constructions in which the actions of one participant are set as being in contrast with those of another participant (see discussion of Clause 36 in the section describing the S1 context). Subjects in these constructions often are encoded with pronouns, even though referential identity is not a problem. A good example is found in Clauses 113-117, where each of the twins acts in contrast to the other (one takes his bow, the other takes a different bow; one stands next to one fortification, the other stands next to a different fortification). It is noted that some of these tokens could have been assigned to the S1A context, although the use of a pronoun does seem to typify contrastive clause constructions.

The final remaining pronoun token, in Clause 5, occurs in the opening section, where the creature is being introduced and clearly described. This environment already has been hypothesized to be a unique one in terms of participant reference.

Thus the default level of noun phrase encoding is supported by a close analysis of the apparently deviant tokens. The VIP hypothesis accounts for most of these tokens; the referent is a key participant, and under-coding is a reflection of this relative importance.





being viewed both as separate from, yet fully representing, the participant, the referent is manifested by zero, and the body part is introduced as an extension of its possessor.

In light of the extremely short referential distance involved in these clauses (essentially, a referential distance of zero/same clause), the label **N0** is arbitrarily being assigned to this newly identified context.

An analysis of all non-subject nominals in the text produces the results presented in Table 20 (it is recalled that the "affix" and "zero" columns have been combined, since for non-subjects the selection of these levels of encoding is determined by the number of the referent). As with subject reference, the initial results are presented first, while Table 22 contains modified totals that reflect an adjustment to the definitions of the N1 and N3 contexts for Sio.

A review of Table 20 on the following page indicates that a default level can easily be proposed for four of the five contexts. The N0, N1 and N3 contexts are predominantly encoded with affix/zero anaphora (respectively, 100%, 77.8%, and 83.3%), and the N4 context default level seems to be a noun phrase (71.6%). The only context with an uncertain default level is the N2 context, where the non-subject was the speaker in an immediately preceding segment of reported speech. Each of the three possible levels of encoding is represented among the five tokens. Since two of the tokens (40%) are encoded with a noun phrase, and a similar number are encoded with affix/zero anaphora, the determination of a default level requires some additional analysis. For reasons to be explained shortly, affix/zero anaphora has been selected as the default level for this context.

In the previous discussion of the S4 context (see the Subject Reference section of this chapter), a new context, S1A, was proposed. This context was for tokens where the referent was present within the preceding two to three clauses, but for which no potentially confusing referent could be found in the intervening clauses. A similar situation is found for non-subject referents, but a different solution is being proposed, as both the N1 and the N3 contexts could easily be

TABLE 20: Coding Material used for Non-Subjects in Different Contexts: Initial Analysis

AMOUNT OF CODING MATERIAL				
CONTEXT	Noun Phrase	Pronoun	Affix/Zero	TOTAL
N0	--	--	7*	7
N1	8	2	35	45
N2	2	1	2	5
N3	4	--	20	24
N4	88	2	33	123

\* = Actually, Affix/Zero + NP (body part)

widened to accommodate a number of the N4 tokens. Rather than creating two new contexts (N1A and N3A), the definition for these two contexts is simply being expanded (as indicated by the italicized portions of the following definitions):

N1 -- The referent occupies the same non-subject role as in the previous *one to three* clauses, with no potentially confusing referent in the intervening clauses.

N3 -- The referent was present in the previous *one to three* clauses in a different role than that covered by N2, with no potentially confusing referent in the intervening clauses.

With these new definitions, a total of 28 Affix/Zero tokens are moved out of N4 and into a different context. Eighteen are moved into N1 (see, for example, Clauses 167, 225, etc.), and 10 are moved into N3 (e.g., Clauses 58, 171, 469, etc.). The remaining five tokens (plus two pronoun tokens) are deviant, and will be discussed shortly.

The default levels proposed for the non-subject contexts are presented in Table 21 on the following page.



TABLE 21: Default Encoding Levels for Non-Subjects, by Context

CONTEXT	DEFAULT ENCODING
N0	Affix/Zero + NP (body part)
N1	Affix/Zero Anaphora
N2	Affix/Zero Anaphora
N3	Affix/Zero Anaphora
N4	Noun Phrase

The following table (Table 22) displays the revised totals for non-subject tokens, by context and encoding level. Deviant tokens are highlighted with an asterisk, and default values are shown in bold face.

TABLE 22: Coding Material used for Non-Subjects in Different Contexts: Final Analysis

AMOUNT OF CODING MATERIAL				
CONTEXT	Noun Phrase	Pronoun	Affix/Zero	TOTAL
N0	--	--	7	7
N1	8*	2*	53	63
N2	2*	1*	2	5
N3	4*	--	30	34
N4	<b>88</b>	2*	5*	95

The following paragraphs explore, by context, the possible explanations for tokens with a level of encoding that is different from the default level for that particular context. As with subject reference, Levinsohn's proposed motivations for under- and over-coding will serve as the

organizing categories, where possible. Since there are no deviant tokens for the N0 context, the discussion begins with the N1 context.

#### A. N1 Context

An N1 classification (as revised) indicates that the referent was mentioned in the same role in the preceding one to three clauses, with no other plausible referent in the intervening text. Affix/zero anaphora is the default level, with an affix being used for plural referents, and zero encoding for singular referents.

The over-coding for each of the eight noun phrase tokens seems to be explained by highlighting of one type or another. Four tokens (Clauses 14, 137, 457, and 496) are instances of repetition of portions of the preceding two to three clauses to highlight the referent (either the inclusiveness of the referent, as in 14 and 457, or the significance of the event, as in 137 and 496).

The remaining four noun phrase tokens have different explanations. In Clause 31, the motivation for repeating "canoe", which also appeared in the preceding clause, probably reflects the contra-expectation that the canoe could not carry their weight. In Clause 277, the repetition of the word "water" may be to highlight the improbability of spray being turned back into drinking water. In Clause 313, the repetition of the "cover" of the hole may be to slow down the story as the climax of the creature's death approaches, and in Clause 316, the mention of the creature's name appears to be a final slow-down before the same climax.

Two examples of pronoun encoding occur in the N1 context, and again the motivation seems to be highlighting of the referent. Clause 165 (with 166) is an example of duplication of the previous two clauses, perhaps to highlight the importance of the event. Clause 530, too, is almost identical in content to 528, which was in a direct speech quote. This repetition may be to highlight this as a climactic event -- the birds obtain possession of the drum. (It is noted that

similar duplication of clauses is observed with noun phrases, too. The principle is that, when a clause is duplicated, the form of the N1 referent is reproduced.)

Therefore the counterexamples to the default encoding level of zero anaphora for N1 tokens are accounted for.

### B. N2 Context

The comparable post-speech subject context (S2) already has been discussed, and in that context it was determined that the default level was an agreement affix, with deviant tokens explained either by the length of the preceding segment of reported speech, or by some other factor. However, since three of the five non-subject tokens are encoded with a higher level of encoding than the proposed default level, a detailed analysis is needed.

Focusing first on the tokens encoded with a noun phrase, it once again is apparent that the length of the segment of reported speech that precedes the clause has some impact. The number of preceding speech clauses for the two N2 tokens encoded with a noun phrase are: Clause 607 (four clauses), and Clause 613 (five clauses). In contrast, the other four tokens, which have less coding material, average only two clauses of preceding reported speech.

The one N2 clause that involves a pronoun (Clause 60), uses the third person plural/dual form, /**rua**/ "two", which already has been shown to be rather idiomatic in Sio. This is the only one of the five N2 tokens in which the twins were the speaker in the preceding speech segment, so the use of a pronoun here may be due primarily to this factor.

Thus it is hypothesized that more coding material than the default level is used when the segment of preceding reported speech is long enough (more than three clauses) that the identity of the N2 referent is potentially ambiguous.

Additional support for establishing an N2 default level of affix/zero anaphora is provided by analyzing the two examples thus identified. Clauses 402 and 409 occur in the midst of a series of quick conversational turns between the older brother (who is the speaker in 396, 399-401, and



408), and the mother (who is the speaker in 397-398, 403-407, and 410-411). For this reason, the identity of the participants is obvious, and the lowest level of encoding is sufficient.

In summary, the default level for N2, affix/zero anaphora, is only deviated from when the identity of the referent is less obvious due to the length of the preceding speech segment.

#### C. N3 Context

Four of 34 N3 tokens represent over-coding from the default level for this expanded context, which is affix/zero anaphora. These tokens are encoded with noun phrases. Two of these four (Clauses 454 and 456) represent instances of repetition of significant portions of the preceding two clauses as the speaker slows the story down (the section from 452-462 is marked by substantial duplication of material). The other two tokens seem to mark clear points of discontinuity: Clause 281 moves the action from the rest break back into the battle, and Clause 568 (which was addressed earlier in the discussion of the S3 context) is a transition between the feast for the birds, and the preparations for the older brother's return.

With these explanations, the default level for the N3 context, affix/zero anaphora, is confirmed.

#### D. N4 Context

Finally, there are seven deviant tokens in the N4 context (which, after redefinition of the N1 and N3 contexts as discussed earlier, includes non-subject participants that have not appeared in the preceding three clauses). A noun phrase is the encoding level for almost 93% of the tokens in this context, so the under-coding of these seven counterexamples requires explanation.

There are two pronoun N4 tokens. In Clause 163, which contains the idiomatic /*rua*/, the twins seem clearly to be the Local VIPs, and this would account for the under-coding. The second token, in Clause 33, seems to be an example of using a pronoun to simplify description of the referent. Since the subject of this clause is the woman's husband along with her oldest son, there are multiple noun phrases that might have been used for the woman (who is in the non-

subject position): "his wife" (which would have excluded the son), "his mother" (which would have excluded the husband), or "the woman" (which might have sounded strange in light of the fact that kinship terms are being employed in this section). Thus the most appropriate solution was to just refer to the non-subject with the pronoun /i/ "she". (In a sense, the woman is the "local VIP", having been introduced first, so this also could explain the under-coding.)

The five tokens that are highly under-coded with affix/zero anaphora also can be accounted for by the VIP hypothesis. In Clause 270 (which is just four clauses removed from the previous mention of the participant), the referent is Nɔndi, who certainly represents a key participant in the text, and whose identity is also clear from the "expectancy chain" of the preceding context. The referent in the other three clauses (417, 428, and 465) is the drum which, although inanimate, is the central point of most of the action in the last half of the text. In fact, the drum and the skin from which it was made serve as the thematic prop for the last half of the story, and as such it is under-coded.

The final under-coded N4 token is in Clause 540. In this clause, the referents (the people of the island) were last mentioned 10 clauses earlier; and since they do not seem to be carrying VIP status in the text, the use of such little coding material is unexpected. However, since the construction /**ndeka ku(ku)**/ "rejoice with" was previously used only once in the text (in 501, where the scene is the dance and the subject and non-subject are identical to the participants in 540), there is little doubt as to the referents' identity. It also is possible that the referent is meant to be indefinite rather than specific: the third person plural referent "they" was last explicitly identified in Clause 493.

This concludes the analysis of the text using Levinsohn's methodology. In Chapter 5, the results of this analysis will be compared with those obtained using the methods proposed by Givon and Tomlin, as presented in Chapter 3.



## CHAPTER 5

### DISCUSSION OF FINDINGS

The purpose of this study, as stated in the opening paragraphs of Chapter 1, was to assess three different methodologies that have been developed to measure participant reference in narrative discourse. This was accomplished by describing participant reference in a Sio narrative text using each of the three different approaches. In this concluding chapter, a comparison will be made of the relative effectiveness of each of the methods.

#### Givon's Recency/Distance Method

Givon's work marked one of the first attempts to systematically describe participant reference using quantitative methods. With his Iconicity Principle he was able to explain the presence of many of the referring expressions in narrative discourse that seem to have more coding material than required for hearer identification. His three measures (referential distance, potential interference, and persistence) were tested on several different languages, and were found to be useful (to differing extents) in explaining the usage of differing amounts of coding material.

The most important of Givon's factors is referential distance, with the underlying assumption that the further back into the preceding discourse that a hearer must search for the identity of a referent, the more coding material is likely to be used. This was demonstrated to be true in Sio for both animate and inanimate subject referents; the average "look-back" value is highest when a full noun phrase is used, and lowest when zero anaphora is used (with pronouns and agreement affixes falling predictably in between these extremes). The results for non-subject reference were generally consistent with this finding, although certain morphosyntactic



constraints or restrictions of the Sio language affected the findings for the object position (there is no choice between an agreement affix and zero anaphora for these tokens: plural referents cannot be encoded with zero, and singular referents cannot be encoded with an affix).

The potential interference factor helps to account for over-coding that seems to result from the presence in the immediately preceding discourse of multiple plausible referents, a situation that may produce referential ambiguity. In the present study, however, this factor did not prove to be very useful. A slightly higher value was observed in both subject and non-subject positions when a noun phrase was used as the referring expression; but when a lower level of encoding was used, there was no significant relationship between potential interference and which of the three lower levels of encoding the speaker used.

The use of the persistence factor was even less productive with the Sio text. The number of successive clauses in which a referent remains "on stage" appears to have no correlation with the specific amount of encoding material that is used. It was only when animate and inanimate participants were separated that any generalizations about persistence could be made, but this involved simply whether or not the participant persisted, not the amount of encoding material used in the referring expression.

In addition to the relative non-productivity of two of his three factors, the problems with Givon's method relate to what it does not explain. Since there is no attempt to account for the internal thematic structure of the text (other than positing over-coding as being partly due to a topic being discontinuous), the analysis of participant encoding as related to episodes and larger "chunks" of discourse is ignored. The Sio text certainly is more than a linear sequence of clauses, and the fact must be taken into account that certain linguistic devices (such as intonation contours, repetition of material, the continuative marker, etc.) seem to mark thematic sections that in turn may account for the use of certain referring expressions.

Furthermore, Givon's method does not attempt to identify whether or not the relative importance of participants in the text affects the form of references to them. Finally, the encoding of participants with regard to their relationship to reported speech segments is ignored, with instructions being given to include such sentences in the counting only when the referent is somehow present in the speech segment.

### Tomlin's Episode/Paragraph Method

Addressing one of the shortcomings of Givon's methodology as described above, Tomlin turns his focus on the thematic structuring of narrative texts. His main objective is to identify the thematic episodes of a text, and particularly to mark the boundaries of those episodes. Once the boundaries are determined, the use of referring expressions is related to these boundaries, with a higher encoding level expected for the initial reference to a participant in a new episode, and lower levels for maintaining reference within the episode.

Unfortunately, the precise identification of episode boundaries on which this method depends is not a simple task. A subjective grouping of events into paragraphs by the researcher, without reference to linguistic marking, is unreliable and totally ignores the speaker's own internal organization of the text. Even when linguistic evidence such as intonation and other formal devices is included in determining episode boundaries, the actual existence of these boundaries is not proven, and can at best only be inferred.

Therefore the results of the analysis of the Sio text using Tomlin's method are inherently suspect. Other researchers could easily disagree with the location of certain episode boundaries, or could propose a smaller or greater number of them, with convincing arguments. In fact, this may account for the poor results reported in Chapter 3 (e.g., only 25% of the over-coded subject tokens were accounted for by episode boundaries). Would different boundaries have produced



better results in terms of "hits"? The moving of boundaries to "fit" the location of over-coded tokens would make the explanation of their presence circular.

Of course, Tomlin avoided this ambiguity by designing his study in such a way that the researcher controlled the precise location of episode boundaries. It may well be that a replication of his study in a similarly controlled environment, rather than simply applying his measuring method to a naturally-produced oral text, would have resulted in more hits. Therefore the results obtained for the Sio text may simply demonstrate that Tomlin's method is best used when episode boundaries are pre-determined by the researcher.

By focusing so exclusively on the thematic structure of the text, as important as this is, Tomlin's method, like Givon's, presents only a partial explanation of the over- and under-coding of participants in a text. He makes no attempt to identify the relative importance of participants in the text, and he treats direct speech segments in the same cursory way that Givon does. Thus, while his method correctly expands the study of participant reference to include discourse-level features, a full accounting of why certain expressions are used at certain places is still missing.

#### Levinsohn's Default/Marked Method

As an alternative to both of these methods, Levinsohn approaches the analysis of participant reference systems in terms of default and marked encoding. Default encoding values are identified for various situations which do not contain discontinuity or "surprise", and any tokens that contain encoding other than these default values are considered to be marked. These tokens are then studied to determine the specific motivation for their marked status. When more coding material is used than is expected, it is seen as signalling either a major discontinuity (i.e., the start of a narrative unit, as indicated by a change in time, place, action, participants) or a highlighting of the event or action being described. Under-coded tokens are explained by an



absence of other participants, an intentional repetition of a cycle of events, or the special prominence of the referent.

In a sense, Levinsohn's method contains components of both of the other two methods. Givon's referential distance factor is reflected in the definitions of the various subject and non-subject contexts; his potential interference factor is reflected in the recognition of over-coding for disambiguation; and his persistence factor is reflected in the VIP categories. Tomlin's focus on episode boundaries is carried through in Levinsohn's extensive treatment of thematic continuity.

However, his method goes beyond these other two, resulting in the most comprehensive of all the strategies identified. His differentiation of the various S- and N-Contexts enables the researcher to precisely define what encoding levels are to be expected, and his VIP strategy provides an explanation for much of the otherwise unexplained encoding found in a text. Also, his recognition that segments of reported speech constitute a significant context for analysis allows him to incorporate more of the text into his research.

The utility of Levinsohn's approach was clearly demonstrated in the analysis of the Sio text. Default levels for each of the contexts were easily identified, and once the deviant tokens were isolated, the explanations for marked encoding closely paralleled those presented by Levinsohn. Practically every over-coded token was shown to exist at a point of discontinuity and/or to highlight the action, and under-coding was found to be totally explained by repetition or by participant prominence. Thus not only was Givon's Iconicity Principle supported, but specific motivations were identified for every apparent violation of this principle.

#### Confirmation of the Usefulness of the Default/Marked Method

Since the Default/Marked method proved so useful as a framework for analyzing participant reference in Sio, it was determined to partially analyze one additional Sio narrative text using this same methodology. The purpose of this step was to validate the utility of the

method, and to determine whether any results significantly different from those described in this paper would be obtained.

This second text, to which reference already has been made in earlier chapters, was told by the same narrator, Melikisede, and was tape-recorded in Lambutina village in May, 1987. The text, entitled "The Kulambi Man", tells of a handsome man from a neighboring language area who tricks his way into the heart of a Sio woman, then eventually is murdered by her jealous sister. It is much shorter than the first text (237 analyzed clauses as compared with 490). For this analysis, only the results for subject referents were calculated. Further, since (like the other text) all of the zero-encoded subject tokens occurred in sentence fragments, these tokens have been removed from the analysis. The INTRO context also has been eliminated in this analysis, as the main focus is on reference to activated participants, rather than the form in which new participants are introduced. It is further noted that the S4 level contains a few tokens which would have been reclassified as S1A in the other text.

Table 23 (see following page) presents the results of an analysis of the coding material used for subjects in different contexts in The Kulambi Man. It will be compared to the results presented in Table 19 in Chapter 4.

With regard to the S2 context, while a simple statistical count would suggest that a noun phrase is the default value, an analysis of these tokens reveals that a noun phrase only is used when the preceding quote is lengthy, or to show discontinuities or highlighting. Thus a better case can be made for establishing an agreement affix as the default level, as was done in the other text, and considering a noun phrase as an over-coded token with the explanation as shown above.

Therefore the default levels for each of these subject contexts would be as shown on Table 24, immediately following Table 23. These results will be compared to Table 18 (see Chapter 4). It is noted that the default levels are identical for both Sio texts.

TABLE 23: Coding Material used for Subjects in Different Contexts: The Kulambi Man

AMOUNT OF CODING MATERIAL				
CONTEXT	Noun Phrase	Pronoun	Affix	TOTAL
S1	3	--	144	147
S2	7	1	5	13
S3	5	1	27	33
S4	36	1	7	44

TABLE 24: Default Encoding Levels for Subjects, by Context: The Kulambi Man

CONTEXT	DEFAULT ENCODING
S1	Agreement Affix
S2	Agreement Affix
S3	Agreement Affix
S4	Noun Phrase

A specific analysis of the deviant tokens represented in Table 23 above will not be presented here. Rather, the following summary is offered: each of the over-coded tokens in S1 and S3 may be explained through either discontinuities or highlighting, and five of the seven under-coded S4 tokens are explained by the participant's status as a local VIP.

The consistency of these findings with those presented in Chapter 4 underscores the usefulness of the Default/Marked method in thoroughly explaining the speaker's selection of referring expressions.

#### Concluding Remarks

This study has shown that Levinsohn's method of identifying and explaining participant reference is much more comprehensive and precise than either of the other two methods being



reviewed here. Far from being too heavily oriented toward the VIP strategy, as suggested by Fox in a criticism written during the early stages of the method's development (1987:160), this method simply incorporates participant prominence as one of several explanatory hypotheses that go significantly beyond the variables studied in the other methods.

There are a few areas in which the method might be improved. The S4/N4 contexts are very broad, essentially incorporating "everything else", and it may be that one or two additional, specific contexts could be identified that would reduce the number of tokens in these general contexts. Thus a possible restructuring of the S1 context, as was done for this present study, might be as follows:

- S1     The subject was the same as in the previous clause
- S1A    The subject was the same as the subject two or three clauses earlier, with no potentially confusing referent being mentioned since that occurrence.

Of course, Levinsohn's provision that the researcher may modify contexts based on language-specific features identified does allow for flexibility, and this may be preferable to increasing the rigidity of context definitions.

The new N0 context that was proposed in this study may be one that should be incorporated into the model. Although the problem in Sio arose because of body parts representing participants, other languages may have different factors that motivate the need for such a context. For example, in English reflexive constructions such as "John saw himself" (= "John saw John"), the classification of the object might be problematical with the present category structure. It would not necessarily fit into the N1 context, and it actually represents the shortest possible referential distance: already mentioned in the same clause.

Another possible improvement would be the formalization of categories of marked encoding. For example, in the area of over-coding, the general categories of "highlighting" and

"discontinuities" could be subdivided and labeled. Thus there could be highlighting subcategories such as:

H1 "significant information" (as in Clauses 6 and 7 of Nondi)

H2 "countering or contra-expectation" (Clauses 31, 248)

H3 "pre-climax slow-down" (Clauses 157, 313)

Discontinuity subcategories might also be proposed, such as:

D1 "time" (Clauses 111, 141)

D2 "place" (Clause 425)

D3 "participants" (Clauses 39, 370)

D4 "action" (Clauses 3, 266)

The existence of what Levinsohn calls "points of departure" (1999:33) could be useful in identifying formal marking associated with discontinuities. Points of departure include sentence-initial adverbials and, under certain circumstances, sentence-initial nominals.

The specificity provided by such sub-categorization of these broad terms would enable the researcher to identify the most common types of highlighting or discontinuity in that particular language, and it might also point out slight differences in encoding strategy for one or more subcategory.

The combining together of all non-subject nominals into a single category may obliterate differences that might exist in some languages between direct objects, indirect objects, adjuncts, etc. The results of the text analysis using Givon's methodology (Chapter 3) did indicate some minor differences in Sio when direct objects were separated out from other non-subject positions, and it would seem that the researcher might at least consider analyzing them separately. Some possible sub-categories of N (substituting "non-direct object" for what was referred to in Chapter 3 as "adjunct") might include:

N1 = Direct object of previous clause remains direct object

N1X = Non-subject non-direct object of previous clause becomes direct object

N1Y = Direct object of previous clause becomes non-subject non-direct object

As was shown when the two Sio texts were compared, the S2 context may be highly influenced by text-specific features. It was demonstrated that a different default value was initially obtained for a text characterized by multiple participants who engage in extended dialogues, than for a text in which few conversational turns are reported. As a minimum, the researcher needs to be careful not to generalize a default S2 value to all texts until several different ones have been studied.

Finally, it might be a useful expansion of the methodology to formally propose thematic paragraph boundaries once the analysis is complete. Having already identified points of discontinuity in the process of explaining over-coding, this information could be combined with linguistic clues (especially intonation contours in oral texts) to determine exactly where these boundaries occur. Over-coding would be a confirmation of previously proposed episode boundaries, while under-coding would call those boundaries into question. (Of course, verification of these breaks by the actual speaker of the text would lend strong support to their precise location.)

In conclusion, Levinsohn's Default/Marked approach has been shown to be an excellent tool for explaining speaker motivation in using participant encoding strategies in narrative texts. It demonstrates that the amount of coding material used in referring expressions is not determined simply by the distance from the previous reference, nor by location within a thematic paragraph. Rather, there is a complex interaction of multiple variables upon which the speaker draws in order to help the hearer determine not only the identity of the referent, but also the structure of the text and the importance of that participant within it.



APPENDICES

## APPENDIX A: LIST OF ABBREVIATIONS USED

Abbreviation	Meaning
=	copula
0	null
1,2,3	1 <sup>st</sup> , 2 <sup>nd</sup> , 3 <sup>rd</sup> person
AMPL	amplifier
CNTR	contrastive
CONT	continuative aspect marker
DUB	dubitative mood
EMPH	emphatic
EX	exclusive
EXCL	exclamatory
HAB	habitual aspect marker
IN	inclusive
INST	instrument
INTNS	intensifier
IRR	irrealis mood
LMTR	limiter
NEG	negative
NOM	nominalizer
p	plural
PERF	perfective aspect marker
POSS, poss	possessive
PURP	purpose
REDUP	reduplication
RFX	reflexive
RSTR	restrictor
s	singular
sbl	sibling
SPEC, spec	specifier
TOP	topic marker
TR	transitivizer
ygr	younger

## APPENDIX B: THEMATIC PARAGRAPHS (EPISODES) IN TEXT

Episode Number	Action	First Clause	Last Clause	Total No. Clauses
1	Introduction; Nɔ̃ndi's description	1	14	14
2	Peoples' reaction to Nɔ̃ndi	15	23	9
3	Nɔ̃ndi arrives at Sio	24	28	5
4	The mother is abandoned	29	38	10
5	The twin sons are born	39	46	8
6	Life in the banyon tree	47	55	9
7	Mother explains events to twins	56	89	34
8	The twins grow	90	98	9
9	The mother warns the twins	99	106	8
10	The twins develop their fighting skills	107	126	20
11	Weapons and fortifications are prepared	127	140	14
12	A rat steals Nɔ̃ndi's armpit skin covering	141	168	28
13	The mother's second warning to the twins	169	188	20
14	The rat carries the twins' challenge to Nɔ̃ndi	189	211	23
15	Nɔ̃ndi searches for missing armpit cover	212	226	15
16	Nɔ̃ndi arrives for the battle	227	236	10
17	The battle begins	237	248	12
18	A rest break is proposed	249	259	11
19	Details of the rest break	260	280	21
20	The battle resumes	281	294	14
21	The twins reach their last fortification	295	303	9
22	Nɔ̃ndi is killed	304	324	21
23	The body is put in a cave	325	338	14
24	The people return to their villages	339	345	7
25	The skin is tested for a drumhead	346	369	24
26	Older brother hears the drum	370	381	12
27	Older brother arrives, confronts mother	382	411	30
28	Mother gives the drum to him	412	434	23
29	Mother defends her action to the twins	435	448	14
30	Birds are sent to retrieve drum, but fail	449	472	24
31	Nzɔ̃ŋi birds arrive at Kavɔ̃	473	491	19
32	The singing (dance) begins	492	502	11
33	The birds' leader wants a better drum	503	520	18
34	He is given the Nɔ̃ndi drum	521	538	18
35	The drum is snatched, taken back to Sio	539	567	29
36	A feast; preparations for older brother	568	577	10
37	Older brother arrives and issues challenge	578	598	21
38	Older brother confronts mother	599	614	16
39	Older brother grabs drum, is killed	615	632	18
40	Body is sent back to Kavɔ̃	633	648	16



#	Comment	Conjunc	Time Phrase	Subject	Verb	Object	Adjunct
36		Aku and		kinzi 3p	si-kai 3p-get	wɔŋga canoe	
37					si-kɔwa 3p-flee		pa Kavɔe to Kavɔe
38					si-lo. 3p-go		
39				I 3s	nde =		kapwa tora lo abdomen hanging PERF
40		ku and			0		tia NEG
41		ku and			i-roto 3s-search	nia area	lee CONT
42		ku and			i-kɔki 3s-go.up		lo amba tonge kulu. in tree.name a atop
43					l-veta 3s-make	nia area	lo PERF
44		ku and			i-mo 3s-be		amba-ni kulu lee tree.name-SPEC atop CONT
45				pagugu-ŋa ne zo give.birth-NOM 3sPOSS time	i-pɔŋga. 3s-arrive		
46					l-pagugu 3s-give.birth	natu rua child two	ŋosi ku wia rua. left and right two
47					Si-mo 3p-be		kuku with
48					i-katona-nzi 3s-care.for-3p		lee lee CONT CONT
49		ku and	zo ŋana time about		ma i-tu IRR 3s-want		

## APPENDIX D: FREE ENGLISH TRANSLATION OF TEXT

1) I am telling the story about the legend of Nondi. 2) That man Nondi, all of his skin was made of quartz. 3) Quartz is a rock 4) that is found in the rivers. 5) He was a person 6) whose skin was extremely hard. 7) That man's skin was just like quartz. 8) He was a very strong man, 9) whose teeth were like husks, 10) and he would move around, 11) going from village to village, 12) killing the inhabitants of each village 13) and eating them. 14) He would kill every person in each village.

15) All the women and men were terrified of him. 16) When they heard a report 17) that Nondi had arrived at the village, 18) then the people of that village would flee 19) and go away. 20) The people of some villages would hide, 21) while people of other villages would go into the rainforest 22) and hide. 23) This is what they would do.

24) Time went by, 25) and one day he came 26) and arrived at Sio 27) with the intent of killing the Sio people. 28) So all of the Sio women and men fled in every direction.

29) One woman was pregnant, 30) and (she), her husband and her son got into a canoe. 31) But (their weight) overwhelmed the canoe, 32) and it started to sink. 33) (It wasn't going to work) so they told her, 34) "OK, you go back to shore 35) and stay." 36) And they took the canoe 37) and fled to Kavoe (Siassi Island), 38) and left.

39) She was very pregnant (literally, "her belly was hanging"), 40) and (she had no other choice), 41) so she searched and searched for a place (to hide), 42) and climbed up to the top of a banyon tree. 43) She fixed a hiding place, 44) and remained atop the banyon tree 45) until her delivery time arrived. 46) And she gave birth to twins, a left-handed one and a right-handed one.

47) They stayed with her, 48) and she kept caring for them, 49) and whenever she wanted 50) to go get food, 51) when those times came, 52) she would wait until night, 53) then she would cook 54) and eat 55) and feed her two children.

56) They kept living like this, 57) and the two children grew. 58) And they asked, 59) "Mother, where is our father?" 60) She told the two of them, 61) "Your skin is filthy! (a rebuke) 62) That's what you two are like, 63) (so) you ask (me that question). 64) As for your older brother and your father, Nondi arrived 65) and killed all of us, 66) eating them 67) and chasing them, 68) and they fled 69) and went away, 70) leaving me 71) to live here like this. 72) So I lived all alone 73) until I gave birth to the two of you 74) and you arrived. 75) So I have taken care of you two as you grew, 76) and now (this is the way things are), 77) and we must (be careful), and 78) during the daytime we absolutely cannot cook 79) or bake. 80) If we cook 81) or bake, 82) then Nondi will see our smoke 83) as it rises, 84) and will come 85) and kill us. 86) So we must wait until night, 87) then I can bake 88) and cook for the two of you, 89) and you can eat."

90) This is how they did things, 91) and time went by, 92) and the two of them strung their bows. 93) Their mother taught them about bow-stringing and arrow-sharpening and spear-sharpening, 94) and the two of them did it. 95) Then every day they would go down to the beach 96) and shoot fish, 97) and would go into the grasslands 98) and shoot birds, too.



99) Time went by, 100) and their mother told them, 101) "Don't you two go very far away. 102) If you go far away, 103) Nōndi will find you 104) and will kill you. 105) You must only stay under our dwelling place, 106) and remain here.

107) Time went by, 108) and the two of them on their own began to see themselves 109) as being strong men; 110) they had become strong. 111) And one day, on their own, the two of them started to practice against each other 112) and fight. 113) One took his bow while 114) the other took his bow, 115) and one went 116) and stood next to one fortification 117) while the other stood next to another fortification -- 118) they had buried logs as fortifications 119) so they would be able 120) to fight 121) and practice together. 122) They kept fighting 123) and practicing together, 124) and the right-handed one told the left-handed one, 125) "Oh, friend, that's enough! 126) The two of us are going to kill Nōndi!"

127) They said this, 128) and they sharpened their bows. 129) They sharpened their bows and spears and arrows, 130) that's what they did. 131) They kept doing this, 132) and fastened them into bundles. 133) They dried them over a fire, 134) then they took them 135) and went down 136) and started erecting fortifications. 137) They erected fortifications, 138) and kept doing this, 139) then they fastened their bows to the edges. 140) And then they were all done.

141) Then one day a rat came. 142) Then the two of them sent it off 143) and told it, 144) "Go and see. 145) The low tide has fully arrived over there. 146) Nōndi has taken off that strong skin 147) and put it down 148) and it is lying there. 149) Start crawling 150) and go 151) and get Nōndi's armpit skin, 152) take it, 153) snatch it, 154) bring it 155) and come." 156) And the rat went. 157) The rat crawled, 158) moving like this 159) as it went, 160) and it reached out its hand 161) and took Nōndi's armpit skin, 162) it took it off 163) and got it for the two of them, 164) and came. 165) It got it for the two of them 166) and came 167) and the two hid it 168) and it lay there.

169) That morning the two of them finished eating, 170) their mother cooked 171) and baked for them 172) and got their drinking water 173) and came 174) and set it down, 175) and she told the two of them, 176) "My two children, you have said 177) that you are going to kill Nōndi. 178) Oh, you're just deceiving yourselves 179) that you can kill him! 180) And he's going to kill both of you 181) and kill me 182) as well! 183) You two aren't able to do this. 184) That man is extremely strong." 185) But the two of them struck a tree (act of defiance) 186) and said, 187) "We are able, 188) the two of us will kill him!"

189) And they sent the rat off again. 190) "Go 191) and tell Nōndi, 192) who's over there in the low tide, 193) tell him 194) to come 195) and we'll fight him." 196) And the rat went 197) and told Nōndi, 198) and Nōndi responded, 199) "What?! Where is the man 200) who has come 201) wanting 202) to kill me? 203) I look for people just like that 204) so I can kill them 205) and eat them! 206) This being true, then where are they 207) that they should come 208) to fight with me? 209) OK, tell the two of them 210) to wait for me, 211) and I'll come later."

212) And he moved along the low tide 213) and went ashore 214) and got all of that quartz skin 215) and stuck it to his body. 216) He got dressed 217) and kept sticking the skin to his body 218) and then he was done. 219) Then he started searching for that one piece of armpit skin. 220) He searched and searched, 221) but he couldn't find it. 222) Therefore he reached out his hand 223, and took a leaf from a breadfruit tree 224) and came 225) and folded it 226) and stuck it on his armpit



227) And he went 228) and said, 229) "Oh, you two youngsters, just where have you come from, 230) coming 231) and wanting 232) to fight with me? 233) I go looking for battles exactly like this, 234) but you want 235) to fight with me! 236) OK then, wait for me!"

237) And that was it. 238) He moved like this, 239) slashing his teeth like this, 240) slashing his teeth like that, 241) and he chopped down the fortification into a pile right at the feet of the two of them. 242) And the two of them started shooting. 243) They shot, 244) but his skin was like quartz, 245) and nothing happened. 246) Their spears and arrows flew, 247) and struck Nɔ̃ndi, 248) but those spears and arrows just splintered on his skin.

249) They just kept doing this. 250) The two of them fought with him. 251) They fought, 252) and kept fighting, 253) then they told him, 254) "Let's eat!" 255) So Nɔ̃ndi returned, 256) going 257) to rest. 258) The two of them went 259) and ate.

260) They were eating, 261) and as they threw away the skins of the vegetables, 262) the skins went 263) and Nɔ̃ndi did magic 264) and the skins became like vegetables again 265) and he ate them. 266) Nɔ̃ndi didn't have any food. 267) This is what they did. 268) They kept eating 269) and throwing the crumbs down 270), which went toward him. 271) They threw the skins down, 272) which went toward him, 273) and Nɔ̃ndi did magic 274) and vegetables appeared 275) and he ate. 276) They kept drinking water, 277) and then would spit water out in the same way, 278) and Nɔ̃ndi did magic 279) and water appeared 280) and he drank.

281) That's what the two of them did for Nɔ̃ndi, 282) then they told him, 283) "OK, let's start fighting again!" 284) And the two went down 285) and stood next to their fortifications again, 286) and they fought again. 287) They fought 288) and fought, 289) and the two of them kept shooting Nɔ̃ndi, 290) but nothing happened. 291) Nɔ̃ndi's teeth were like husks, 292) and he slashed like this, 293a) and the logs 294) which the two of them had erected 293b) splintered at their feet.

295) That's what the two of them did. 296) They kept doing this, 297) and two of the fortifications were destroyed, 298) and they arrived at the last fortification. 299) And their mother's skin was shaking with fear, 300) and she said, 301) "Oh no, the two of them completely deceived themselves, 302) and now Nɔ̃ndi is going to eat them 303) and eat me, too!"

304) And the two of them struggled on, 305) fighting with him, 306) fighting on and on with him, 307) and they arrived at that last fortification, 308) and Nɔ̃ndi went down 309) to knock that fortification into a pile next to them. 310) Then the left-hander ran 311) and went 312) and kicked open the covering (of a hole he had dug) 313) and lay down inside the opening. 314) Then he took aim at that armpit, 315) so that he could shoot the armpit apart. 316) And he shot Nɔ̃ndi, 317) and his arms and legs flew out 318) and he struck the ground 319) and lay there. 320) And Nɔ̃ndi went (a short distance) 321) and took a deep gasp, 322) and lay there 323) and died, 324) and there he lay.

325) The two of them came 326) and whooped a victory cry at his head, 327) and their mother was thrilled, 328) and ran 329) and went down 330) and shook their hands. 331) She told them, 332) "Wow, you two are so strong 333) that you have killed Nɔ̃ndi!" 334) And they took (his body) 335) and



moved like this 336) and went down inside a cave in the rocks 337) and shoved the body inside, 338) and there it lay.

339) News about this traveled 340) and went to all the villages 341) and they rejoiced greatly 342) and moved like this 343) and came 344) and reclaimed their villages and land again, 345) and settled in.

346) As for that quartz skin, those two youngsters put it inside the cave, 347) and it lay there for a time, 348) and it became rotten. 349) And they went 350) and struck it (like a drum). 351) They wanted 352) to strike the leg 353) and try it out, 354) but it didn't make a sound. 355) They struck the abdomen, 356) but it didn't make a sound. 357) They struck the head, 358) but it didn't make a sound. 359) They struck the back, 360) but it didn't make a sound. 361) Then they wanted 362) to strike the neck, 363) and the neck resounded. 364) The noise of that drum was incredible. 365) They struck it on the beach at Sio, 366) and the sound went 367) and reached Kavoe, 368) and they heard it, too. 369) Wow, that drum resounded awesomely!

370) And that older brother of theirs heard the drum 371) as it resounded, 372) and he saw the smoke 373) as it arose. 374) And he said, 375) "Oh, I think I'll meander over 376) and see Mom. 377) (That fire) is my Mom's, 378) she must be alive, 379) and has made her smoke 380) which is rising. 381) Perhaps she has given birth to my younger brothers."

382) And he pulled up the sail, 383) prepared food, 384) fixed a taro meal, 385) and put it on top of the canoe, 386) and raced on, 387) going toward shore, 388) coming. 389) He kept going toward shore, 390) coming, 391) and he saw his mother, 392) and he spoke to her, 393) and he gave her food and bread, 394) and she ate. 395) And he asked, 396) "Oh, and where are my younger brothers?" 397) "Oh, your two younger brothers went to the grasslands 398) to work in their garden." 399) "Well, I heard my drum 400) when it resounded, 401) so where is my drum?" 402) But she told him, 403) "Oh, you don't have a drum. 404) That drum belongs to your two younger brothers. 405) They killed Nōndi, 406) and then (they took) his neck 407) and struck it." 408) "OK, show it to me." 409) But she told him, 410) "Oh, I'm not going to show it to you. 411) It belongs to your two younger brothers."

412) He kept doing this, 413) going on and on, 414) and he wanted 415) to stab his mother with a spear. 416) Well, their mother was afraid for her life, 417) and so she showed it to him. 418) She showed it to him, 419) and he reached out his hand 420) and took it. 421) he reached out his hand 422) and took it, 423) and he went down 424) and struck it on the path, 425) and those two youngsters heard it from the grasslands. 426) The two started running. 427) He went down, 428) put it on top of the canoe, 429) pulled up his sail, 430) and fled like this, 431) snatching it 432) and going toward Kavoe, 433) heading up 434) as he went.

435) The two youngsters came 436) but couldn't do a thing, 437) and they wanted 438) to shoot their mother. 439) But their mother told them, 440) "Oh man, I did try to hide it, 441) but that older brother of yours took a spear 442) to stab me, 443) wanting 444) to kill me, 445) so I was afraid for my life, 446) so I showed it to him, 447) so he took it 448) and left!"

449) This statement greatly disturbed the two of them, 450) and they spoke animatedly, 451) greatly angered (literally, "with very hot hearts"). 452) So they called out to the birds in their nesting places 453) to come. 454) They called out to the birds in their nesting places, 455) and they all came. 456) and they sent off all the birds. 457) They sent off all the birds, 458) and they went. 459) to go to



Kavœ 460) in order to get that drum for them. 461) They called out to the birds 462) to come. 463) But all the different birds tried 464) to go 465) and get (the drum), 466) but they didn't succeed. 467) They went, 468) but the wind was blowing hard, 469) and it pushed them, 470) and the waves rose up, 471) and they all returned 472) and came back.

473) Then they sent the nzoŋi birds. 474) "OK, you go." 475) Then the nzoŋi birds went. 476) The birds called nzoŋi birds went 477) and flew, 478) going, 479) and when the waves rose 480) they would go up, 481) when the waves went down 482) they would go down. 483) When the wind came 484) to push them, 485) they went down 486) and traveled just above the ocean. 487) This is what they did. 488) They went, 489) and kept going on and on, 490) and they went 491) and sat down on top of the coconut trees.

492) And then it was night, 493) the time when the people of Kavœ were going to dance a singsing. 494) Night came, 495) and they danced a singsing. 496) They danced a singsing, 497) and all the nzoŋi birds went down with them 498) and went into their midst 499) and danced the singsing with them. 500) At night, those very same little birds went 501) and rejoiced with them 502) and danced the singsing with them.

503) They did this, 504) and they had given one nzoŋi bird a bad drum, 505) and he was upset. 506) "I'm not going to take this drum. 507) I am a top singsing person, 508) so give me a good drum." 509) And all of his companions told them, 510) "Hey, this guy is like the composer of this singsing. 511) If we give him a bad drum, 512) then he won't dance the singsing well. 513) If we give him a good drum, 514) he'll dance the singsing awesomely." 515) And they wanted 516) them to give (him) a good drum, 517) but they wouldn't, 518) so he was upset 519) and disgusted (literally, his skin was wrinkled). 520) "I am disgusted about this drum."

521) They did this, 522) and kept doing it, 523) and he heard Nōndi's neck drum, 524) they struck it 525) and (the sound) rang out. 526) It resounded awesomely. 527) And they told them, 528) "Yes, if you give him that drum, 529) that would be good." 530) And they gave him that drum, 531) he took it, 532) and deceived them, 533) dancing, 534) going about the perimeter (of the singsing), 535) going 536) and coming. 537) He was being deceitful, 538) dancing intensely.

539) Those very birds, the small ones, danced 540) and rejoiced with them. 541) They danced 542) and kept dancing, 543) and then, just at first light, they let go of all the drums, 544) and took Nōndi's neck, 545) quickly snatching it, 546) and went up, 547) going to the top of the coconut branches, 548) and sat down. 549) And he yelled, 550) "Hey, bring my drum to me 551) and come, 552) don't do this!" 553) He called to them, 554) "Bring it to me 555) and come!" 556) But they let go of all the stuff, 557) and it crashed down, 558) and they came together 559) and took that drum 560) and snatched it, 561) taking it, 562) going toward shore, 563) coming. 564) They took it on and on, 565) and went ashore 566) and came 567) and gave it to the left-hander and the right-hander.

568) Then the two of them made a huge feast for the little nzoŋi birds. 569) They finished eating 570) and drinking, 571) then they said, 572) "Very good, now we will keep this drum of ours in our possession, 573) then we will wait for this man 574) to go ashore 575) and come." 576) Then the two of them stayed there, 577) waiting for their older brother.



578) The two stayed only in their village, 579) and then they looked up toward the middle of the ocean 580) and saw the glistening sail of their older brother's canoe 581) as it came. 582) "OK, come on!" 583) And he kept moving just like this 584) and he came ashore, 585) and the two of them went 586) and hid. 587) He took his bow, 588) and he (the other) took his bow. 589) The two of them went 590) and hid 591) and he moved just like this 592) to go ashore on the beach, 593) and he waved with his spear 594) and stabbed the ground 595) and (the sand) sprayed up. 596) And he said, 597) "This is what my skin is like, 598) and I'm going to kill someone!"

599) He hurried, 600) moving like this, 601) and went ashore 602) and asked his mother, 603) "Those two bad people took my drum, 604) they took it 605) and came. 606) Where are they?!" 607) But she told that oldest child of hers, 608) "Hey, I told you 609) to stop, 610) this thing belongs to those two alone, 611) and if you do this 612) you'll destroy yourself!" 613) But he told his mother, 614) "Those two bad people aren't able to do anything!" (loose translation of an idiom)

615) He said this, 616) and that was it. 617) He reached out his hand 618) and grabbed Nondi's neck, that throat drum, 619) and took it 620) and hurried 621) and went down 622) and came. 623) He went down 624) and came, 625) and he got close, 626) then the left-hander pulled back his bowstring 627) and the right-hander pulled back his bowstring. 628) He shot him in one armpit, 629) and he (the other) shot him in the other armpit, 630) and their older brother staggered around, 631) hit the ground, 632) and lay there.

633) And the two of them killed him, 634) and he died, 635) and the two of them took him 636) and went 637) and threw him on top of the canoe 638) and pulled his sail up, 639) fastening the paddle to the canoe brace, 640) and it was all over. 641) They shoved it, 642) and it went down 643) and went into the opening in the reef 644) and it drifted 645) and went up toward his village, Kavoe. 646) And the two of them returned 647) and came back, 648) and stayed in their village.

649) The end.

## APPENDIX E: INTERLINEARIZED VERSION OF SIO TEXT

- 1 A-tapori ɲana usi Nɔndi. 2 Tamɔta ɲinde Nɔndi nde tini ndoni nde nɔndi nɔ.  
*1s-tell.story about legend Nɔndi person that Nɔndi TOP skin all = quartz LMTR*
- 3 Nɔndi ɲinde nde mira tonɲe 4 uru i-keno loo ilo ɲinde. 5 I tamɔta 6 karae ɲinde  
*Nɔndi that = rock a HAB 3-be water inside that 3s person skin that*
- kaika ndo 7 aku tamɔta ɲinde nde tini i-kura nɔndi. 8 I tamɔta kaika ndo 9 niɲo saɲiri.  
*strong AMPL and person that TOP skin 3s-be.equal.to quartz. 3s person strong AMPL tooth husk*
- 10 I-yoka 11 i-kura lawea lawea 12 i-pu-nzi lawea lawea 13 i-ka-nzi. 14 I-pu-nzi lawea lawea  
*3s-move 3s-cover village village 3s-hit-3p village village 3s-eat-3p 3s-hit-3p village village*
- tamɔta ndoni.  
*person all*
- 15 Taine wa tamɔne rɔrɔni si-ruru ndo ɲana. 16 Ma si-loɲo pɔri 17 tu Nɔndi i-pɔɲga lawea  
*woman and man all 3s-fear AMPL about IRR 3p-hear story thus Nɔndi arrive village*
- ɲinde 18 ande lawea ɲinde si-kɔwa 19 si-lo. 20 Lawea pinde si-muna 21 pinde si-lo dugu  
*that and.then village that 3p-flee 3p-go village some 3p-hide some 3p-go bush*
- ilo-lo-lo 22 si-muna. 23 Si-veta mine.  
*inside-REDUP-REDUP 3p-hide 3p-do like.this*
- 24 Si-mo lee lee 25 ku zo tonɲe ku i-mɔ 26 ku i-pɔɲga Sio 27 ɲana ma i-pu-nzi Sio kɔ.  
*3p-be CONT CONT and time a and 3s-come and 3s-arrive Sio PURP IRR 3s-hit-3p Sio RSTR*
- 28 Aku Sio taine-tamɔta rɔrɔni si-kɔwa pwa-pwataki.  
*And Sio woman-person all 3p-flee REDUP-split*
- 29 Taine tonɲe kapwa 30 ku kaiwa wa natu wa si-kɔki wɔɲga kulu 31 andetia si-tawa wɔɲga  
*woman a abdomen and spouse and child and 3p-go.up canoe atop but 3p-push canoe*
- 32 i-mbwatuke. 33 Tia ku si-pai i 34 tu "Ara noko ku-toa 35 ku-mo." 36 Aku kinzi si-kai  
*3s-submerge NEG and 3p-tell 3s thus good 2s 2s-go.ashore 2s-be and 3p 3p-get*
- wɔɲga 37 si-kɔwa pa Kavɔe 38 si-lo.  
*canoe 3p-flee to Kavɔe 3p-go*
- 39 I nde kapwa tɔra lo 40 ku tia 41 ku i-roto nia lee 42 ku i-kɔki lo amba tonɲe  
*3s = abdomen hanging PERF and NEG and 3s-search area CONT and 3s-go.up in tree.name a*



kulu. 43 I-veta nia lo 44 ku i-mo amba-ni kulu lee 45 pagugu-ŋa ne zo  
*atop 3s-make area PERF and 3s-be tree.name-SPEC atop CONT give.birth-NOM 3sPOSS time*

i-pɔŋga. 46 I-pagugu natu r ɔsi ku wia rua.  
*3s-arrive 3s-give.birth child two left and right two*

47 Si-mo kuku 48 i-katona-nzi lee lee 49 ku zo ɔjana ma i-tu 50 i-kai kɔpwa 51 i-mo  
*3p-be with 3s-care.for-3p CONT CONT and time about IRR 3s-want 3s-get food 3s-come*

52 ande ma i-mo lee mbo 53 ɔineŋga ma i-mbwea 54 aku i-ka 55 i-sua-nzi natu rua.  
*and.then IRR 3s-be CONT night then IRR 3s-cook and 3s-eat 3s-feed-3p child two*

56 Si-mo lee lee 57 ku natu rua ɔinde ɔalae lo. 58 Aku si-kasoŋa 59 "Tina-ma, aŋga  
*3p-be CONT CONT and child two those large PERF and 3p-ask mother-1pEXposs CNTR*

maka tama-ma nde i-mo ndia?" 60 I-pai-nzi rua 61 tu "Tini-mi muso rua! 62 Miki rua  
*1pEX father-1pEXposs TOP 3s-be where 3s-tell-3p two thus skin-2pPOSS dirty two 2p two*

mine 63 ka-kasoŋa. 64 Tua-mi wa tama-mi wa Nɔndi i-pɔŋga 65 ku i-pu maka  
*like.this 2p-ask old.sbl-2pPOSS and father-2pPOSS and Nɔndi 3s-arrive and 3s-hit 1pEX*

rɔrɔni 66 i-ka-nzi 67 wa i-ɔara-nzi 68 ku si-kɔwa marumbu 69 si-lo 70 si-pile naŋa 71 a-mo  
*all 3s-eat-3p and 3-rout-3p and 3p-flee finished 3p-go 3p-leave 1s 1s-be*

mine. 72 Kala naŋa simbo-ŋgu nɔ a-mo lee 73 kala a-pagugu-a miki rua 74 ka-pɔŋga. 75 Kala  
*like.this so 1s alone-1s LMTR 1s-be CONT so 1s-give.birth-TR 2p two 2p-arrive so*

naŋa a-katona miki rua lee ɔalae. 76 Ande ma zo ɔine 77 ande ma kinda ma 78 kari kari ma  
*1s 1s-care.for 2p two CONT large and IRR time this and IRR 1pIN IRR sun sun IRR*

ta-mbwea 79 ta-momo tia ndo. 80 Ma ta-mbwea 81 ta-momo 82 Nɔndi i-mora nenda mundo  
*1pIN-cook 1pIN-bake NEG AMPL IRR 1pIN.cook 1pIN-bake Nɔndi 3s-see 1pINposs smoke*

83 0-kɔki 84 ma i-mo 85 i-pu ninda lo. 86 Kinda ma ta-mo lee lee ku mbo 87 ɔineŋga  
*0-go.up IRR 3s-come 3s-hit 1pIN PERF 1pIN IRR 1pIN-be CONT CONT and night then*

naŋa a-momo 88 a-mbwea pa-mi rua 89 ku ka-ka."  
*1s 1s-bake 1s-cook to-2p two and 2p-eat*

90 Si-veta mine 91 si-mo lee lee 92 ku rua si-tike nenzi temba. 93 Tina-nzi i-panana-nzi  
*3p-do like.this 3p-be CONT CONT and two 3p-string 3pPOSS bow mother-3pPOSS 3s-teach-3p*

ɔjana temba tike-ŋa wa silɔwa nziu-ŋa wa logo nziu-ŋa wa 94 rua si-veta. 95 Aku  
*about bow string-NOM and arrow sharpen-NOM and spear sharpen-NOM and two 3p-do and*

kari kari ande ma si-ndue sɔwa 96 si-pane kanzi iŋa 97 wa s-oka nzo  
*sun sun and.then IRR 3p-go.down beach 3p-shoot 3pPOSS fish and 3p-move plains*



ilo-lo-lo 98 si-pane nenzi sii wa.  
*inside-REDUP-REDUP 3p-shoot 3pPOSS bird and*

99 Si-mo lee lee 100 ku tina-nzi i-pai-nzi 101 tu "Miki rua ma ka-lo malawae tia.  
*3p-be CONT CONT and mother-3pPOSS 3s-tell-3p thus 2p two IRR 2p-go far.away NEG*

102 Ma rua ka-lo malawae 103 ma Nondi i-sonda kulu-mi 104 ma i-pu-mi rua lo. 105 Rua  
*IRR two 2p-go far.away IRR Nondi 3s-find head-2pPOSS IRR 3s-hit-2p two PERF two*

ka-mo nenda lawea kalo no 106 ku ka-mo."  
*2p-be 1pINposs village under LMTR and 2p-be*

107 Si-mo lee lee 108 ku kinzi rua waraka-nzi rua si-pa-mora-i 109 tu tamota kaika 110 rua  
*3p-be CONT CONT and 3p two self-3p two 3p-RFX-see-TR thus person strong two*

si-kaika lo. 111 Aku zo tonge ku kinzi rua waraka-nzi rua si-pa-togo 112 si-para. 113 I i-kai  
*3p-be.strong PERF and time a and 3p two self-3p two 3p-RFX-test 3p-fight 3s 3s-get*

ne temba 114 i i-kai ne temba. 115 Aku i i-lo 116 i-mandi tenggo tonge tini 117 i  
*3sPOSS bow 3s 3s-get 3sPOSS bow and 3s 3s-go 3s-stand wall a next.to 3s*

i-mandi tenggo tonge tini 118 ku rua si-kea nenzi koi togo tenggo mine lo 119 0-kura  
*3s-stand wall a next.to and two 3p-bury 3pPOSS log like wall like.this PERF 0-be.able*

120 si-para 121 si-pa-togo. 122 Si-para 123 si-pa-togo lee lee 124 ku wia i-pai nosi 125 tu  
*3p-fight 3p-RFX-test 3p-fight 3p-RFX-test CONT CONT and right 3s-tell left thus*

"O, nambwe 0-kura lo! 126 Kinda rua ma ta-pu Nondi no!"  
*Oh friend 0-be.able PERF 1pIN two IRR 1pIN-hit Nondi LMTR*

127 Rua si-poro mine lo 128 ku rua si-nziu temba. 129 Si-nziu temba wa logo wa silowa  
*two 3p-say like.this PERF and two 3p-sharpen bow 3p-sharpen bow and spear and arrow*

wa 130 rua si-veta. 131 Si-veta lee lee 132 ku si-pa wora-wora. 133 Rua si-reke lee  
*and two 3p-do 3p-do CONT CONT and 3p-fasten bundle-REDUP two 3p-grill CONT*

lee 134 ku si-kai 135 si-ndue 136 ku rua si-ngunu tenggo. 137 Si-ngunu tenggo 138 si-ngunu lee  
*CONT and 3p-get 3p-go.down and two 3p-erect wall 3p-erect wall 3p-erect CONT*

lee 139 ku si-pa temba lo tini 140 ku marumbu lo.  
*CONT and 3p-fasten bow on skin and finished PERF*

141 ninenga zo tonge koto tonge i-mo. 142 ninenga rua si-supwa 143 ku si-pai 144 tu "Noko  
*then time a rat a 3s-come then two 3p-send and 3p-tell thus 2s*

- ku-mora. 145 Maroti 0-pɔ-pɔŋga lo ndai. 146 Nɔndi i-sowa karae kaika ŋinde piti 147 ku i-o  
*2s-see low.tide 0-REDUP-arrive at there Nɔndi 3s-dress skin strong that off and 3s-put*
- ndue 148 i-keno ndai. 149 Noko ku-ŋgara lee 150 ku-lo 151 ku pw-ai Nɔndi kambasi ŋinde 152 nde  
*down 3s-be there 2s 2s-crawl CONT 2s-go and 2s-get Nɔndi armpit that and*
- pw-ai piti 153 ku pw-awea 154 pw-ai 155 ku-mo." 156 Aku koto i-lo. 157 Koto i-ŋgara 158 i-yoka  
*2s-get off and 2s-snatch 2s-get 2s-come and rat 3s-go rat 3s-crawl 3s-move*
- mine lee 159 i-lo 160 ku mbau i-lo 161 i-kai Nɔndi kambasi kaika 162 nde i-kai piti 163 ku  
*like.this CONT 3s-go and hand 3s-go 3s-get Nɔndi armpit strong and 3s-get off and*
- i-kai pa-nzi rua 164 i-mo. 165 I-kai pa-nzi rua 166 i-mo 167 ku rua si-vea lo 168 ku 0-keno.  
*3s-get to-3p two 3s-come 3s-get to-3p two 3s-come and two 3p-hide PERF and 0-be*
- 169 Kinzi rua si-ka lo mboyo lo, marumbu lo 170 tina-nzi i-mbwea 171 i-momo  
*3p two 3p-eat PERF morning PERF finished PERF mother-3pPOSS 3s-cook 3s-bake*
- pa-nzi lo 172 ku i-kai kanzi lo 173 i-mo 174 i-o ndue 175 ku i-pai-nzi rua 176 "Natu-ŋgu  
*to-3p PERF and 3s-get 3pPOSS water 3s-come 3s-put down and 3s-tell-3p two child-1sPOSS*
- rua, miki rua kala ka-poro 177 tu ma ka-pu Nɔndi. 178 Yoo ma rua ka-pa-loŋe 179 tu ka-pu.  
*two 2p two so 2p-say thus IRR 2p-hit Nɔndi Oh IRR two 2p-RFX-deceive thus 2p-hit*
- 180 Aku ma i-pu-mi rua 181 ku ma i-pu naŋa 182 a-yoka tava lo. 183 Miki rua ka-kura tia.  
*and IRR 3s-hit-2p two and IRR 3s-hit 1s 1s-move with PERF 2p two 2p-be.able NEG*
- 184 Tamota ŋinde tamota kaika ndo." 185 Rua si-pu koi 186 si-poro 187 tu "0-kura 188 maka rua  
*person that person strong AMPL two 3p-hit tree 3p-say thus 0-be.able 1pEX two*
- ma ka-pu!"  
*IRR 1pEX-hit*
- 189 Aku si-supwa koto kilo. 190 "Noko ku-lo 191 ku ku-pai Nɔndi, 192 i-mo maroti kulu ndai,  
*and 3p-send rat again 2s 2s-go and 2s-tell Nɔndi 3s-be low.tide atop there*
- 193 ku ku-pai 194 tu i-mo 195 maka ma ka-para." 196 Aku koto i-lo 197 ku i-pai Nɔndi  
*and 2s-tell thus 3s-come 1pEX IRR 1pEX-fight and rat 3s-go and 3s-tell Nɔndi*
- 198 aku Nɔndi i-tu 199 "A? Tamota ŋinde nde i-mo ndia 200 ŋineŋga i-mo 201 i-tu 202 i-pu  
*and Nɔndi 3s-reply huh person that TOP 3s-be where then 3s-come 3s-want 3s-hit*
- naŋa? 203 Naŋa a-roto ŋana-nzi tamota mine 204 tu na-pu-nzi 205 wa a-ka-nzi ko. 206 Ambo  
*1s 1s 1s-search about-3p person like.this PURP 1s-hit-3p and 1s-eat-3p RSTR if*



ɲine nde si-mo ndia 207 ɲineɲga si-mo 208 tu si-para kuku naɲa? 209 Ara, ku-pai-nzi rua 210 tu  
*this and 3p-be where then 3p-come PURP 3p-fight with 1s good 2s-tell-3p two thus*

si-ndamwa-na 211 naɲa mwaɲga a-mo.”

*3p-await-1s 1s later 1s-come*

212 Aku i-yoka maroti kulu 213 i-toa 214 ku i-kai tini nɔndi ɲinde rɔrɔni 215 i-ponda lo tini.  
*and 3s-move low.tide atop 3s-go.ashore and 3s-get skin quartz that all 3s-stick on skin*

216 I-sowa 217 wa i-ponda lo tini lee lee 218 ku marumbu lo. 219 ɲineɲga i-roto ɲana  
*3s-dress and 3s-stick on skin CONT CONT and finished PERF then 3s-search about*

kambasi taitu ɲinde ko. 220 I-roto lee 221 i-sonda kulu tia. 222 Tia, ku mbau i-lo 223 i-kai  
*armpit one that RSTR 3s-search CONT 3s-find head NEG NEG and hand 3s-go 3s-get*

lawua lau 224 i-mo 225 i-ɲgu 226 ku i-ponda lo kambasi.

*breadfruit leaf 3s-come 3s-fold and 3s-stick on armpit*

227 Aku i-lo 228 ku i-poro 229 "O, lolu rua miki rua ɲineɲga k-oka sinia 230 ka-mo 231 kala  
*and 3s-go and 3s-say Oh youth two 2p two then 2p-move from.where 2p-come so*

ka-tu 232 ka-para kuku naɲa? 233 Naɲa a-roto ɲana temba-ni mine ko 234 aɲga miki ka-tu  
*2p-want 2p-fight with 1s 1s 1s-search about bow-SPEC like.this RSTR CNTR 2p 2p-want*

235 ka-para kuku naɲa. 236 Ara mbo, ka-ndamwa-na.”

*2p-fight with 1s good EMPH 2p-await-1s*

237 Aku marumbu. 238 I-yoka mine lo 239 ku niɲo no i-wasa mine 240 i-wasa mine  
*and finished 3s-move like.this PERF and tooth INST 3s-slash like.this 3s-slash like.this*

241 ku i-so teɲgo bibo lo kinzi rua ke-nzi tini-nzi. 242 Aku kinzi rua si-pane.  
*and 3s-chop wall together at 3p two foot-3pPOSS skin-3pPOSS and 3p two 3p-shoot*

243 Si-pane, 244 ande tini nɔndi, 245 ku tia, 246 ku logo wa silowa wa i-lo 247 ande i-ka  
*3p-shoot CNTR skin quartz and NEG and spear and arrow and 3s-go and 3s-eat*

Nɔndi 248 ku logo wa silowa ɲinde nde mbwaranuki-nuki lo tini.

*Nɔndi and spear and arrow that TOP splintered-REDUP on skin*

249 Si-veta mine no. 250 Rua si-para kuku. 251 Si-para 252 si-para lee lee 253 ku rua  
*3p-do like.this LMTR two 3p-fight with 3p-fight 3p-fight CONT CONT and two*

si-pai 254 tu "Ta-ka ɲga." 255 Aku Nɔndi i-taulo 256 i-lo 257 i-pwarea. 258 Kinzi rua si-lo  
*3p-tell thus 1pIN-eat EXCL and Nɔndi 3s-return 3s-go 3s-rest 3p two 3p-go*



259 ku si-ka.  
and 3p-eat

260 Si-ka 261 ku kaniṅa karae ṅinde nde si-sia 262 i-ḷo 263 ku Nḷndi i-sombe 264 i-pḷṅga  
3p-eat and vegetable skin that TOP 3p-throw 3s-go and Nḷndi 3s-bless 3s-become

togo kaniṅa 265 ku i-ka. 266 Nḷndi nde kḷpwa tia. 267 Si-veta mine. 268 Rua si-ka lee lee  
like vegetables and 3s-eat Nḷndi = food NEG 3p-do like.this two 3p-eat CONT CONT

269 pupunu si-tambira 270 i-ḷo papa. 271 karae si-tambira 272 i-ḷo papa 273 Nḷndi i-sombe  
crumbs 3p-throw.down 3s-go to skin 3p-throw.down 3s-go to Nḷndi 3s-bless

274 ku kaniṅa i-pḷṅga 275 ku i-ka. 276 Si-nu ḷo lee lee 277 nde ḷo si-supwara  
and vegetables 3s-appear and 3s-eat 3p-drink water CONT CONT and water 3p-spit.out

mine ḷo 278 nde Nḷndi i-sombe 279 ḷo i-pḷṅga 280 ku i-nu.  
like.this PERF and Nḷndi 3s-bless water 3s-appear and 3s-drink

281 Rua si-veta mine papa Nḷndi ḷo 282 ku si-pai 283 tu, "Ara ṅine ma ta-para kilo."  
two 3p-do like.this to Nḷndi PERF and 3p-tell thus good this IRR 1pIN-fight again

284 Aku rua si-ndue 285 ku si-mandi nenzi teṅgo tini kilo 286 ku si-para kilo. 287 Si-para  
and two 3p-go.down and 3p-stand 3pPOSS wall next.to again and 3p-fight again 3p-fight

288 si-para 289 rua si-pane Nḷndi 290 andetia tia. 291 Nḷndi nde niṅo saṅiri ḷo 292 i-wasa  
3p-fight two 3p-shoot Nḷndi but NEG Nḷndi = tooth husk LMTR 3s-slash

mine 293(a) ku kḷi ṅinde 294 rua si-ṅgunu ṅinde 293(b) nde mbwaranuki-nuki ḷo kinzi  
like.this and log that two 3p-erect that TOP splintered-REDUP at 3p

ke-nzi tini.  
foot-3pPOSS skin

295 Rua si-veta. 296 Si-veta lee 297 ku teṅgo rua marumbu 298 aku ṅia si-pḷṅga ḷo teṅgo taitu.  
two 3p-do 3p-do CONT and wall two finished and two 3p-arrive at wall one

299 Aku tina-nzi nde tini rudidi 300 i-poro 301 "Opopo, ṅine rua si-pa-ḷḷe pḷta ḷo,  
and mother-3pPOSS = skin trembling 3s-say oh.no this two 3p-RFX-deceive INTNS PERF

302 ma ṅineṅga Nḷndi ma i-ka-nzi rua 303 i-ka naṅa ḷo!"  
IRR then Nḷndi IRR 3s-eat-3p two 3s-eat 1s PERF

304 Aku rua si-la-laṅe 305 si-para kuku 306 si-para kuku lee lee 307 ku si-pḷṅga teṅgo taitu  
and two 3p-REDUP-strive 3p-fight with 3p-fight with CONT CONT and 3p-arrive wall one

- ηινδε, 308 i-ndue 309 ηana i-pu teηgo ηινδε bibo lo rua tini-nzi ko. 310 ηινεηγα ηosi  
*that 3s-go.down PURP 3s-hit wall that together at two skin-3pPOSS RSTR then left*
- i-palilu 311 i-lo 312 i-pale kawa utu 313 ku i-keno ndue kawa ilo. 314 ηινεηγα i-pambia kambasi  
*3s-run 3s-go 3s-kick mouth AMPL and 3s-lie down mouth inside then 3s-aim.at armpit*
- ηινδε 315 ηana i-pane kambasi pwataki 316 ku i-pane Nondi 317 ku kie-mbau piripiri 318 ku  
*that PURP 3s-shoot armpit split and 3s-shoot Nondi and foot-arm floppy and*
- i-pu nia lo 319 i-keno. 320 Aku Nondi i-lo 321 ku i-ηoro lo 322 i-keno 323 i-mote utu lo  
*3s-hit area PERF 3s-lie and Nondi 3s-go and 3s-gasp PERF 3s-lie 3s-die AMPL PERF*
- 324 i-keno.  
*3s-lie*
- 325 Rua si-mo 326 si-kai lolove lo i kulu 327 ku tina-nzi i-ndeka pota 328 i-palilu  
*two 3p-come 3p-get war.cry at 3s head and mother-3pPOSS 3s-rejoice INTNS 3s-run*
- 329 i-ndue 330 i-taηo rua mbau-nzi. 331 I-pai-nzi 332 tu "Opo, miki rua ka-kaika 333 ka-pu  
*3s-go.down 3s-touch two hand-3pPOSS 3s-tell-3p thus oh 2p two 2p-be.strong 2p-hit*
- Nondi lo!" 334 Si-kai 335 s-oka mine lee 336 si-lo mira ηgonza ilo 337 si-lumbwa lo  
*Nondi: PERF 3p-get 3p-move like.this CONT 3p-go rock cave inside 3s-shove.in PERF*
- 338 ku i-keno.  
*and 3s-be*
- 339 Pori ηινδε i-yoka 340 i-kura-nzi lawea lawea ndoni 341 ku si-ndeka pota lo 342 ku  
*story that 3s-move 3s-cover-3p village village all and 3p-rejoice INTNS PERF and*
- s-oka mine 343 si-mo 344 si-kai nenzi lawea wa nia kilo 345 si-mo.  
*3p-move like.this 3p-come 3p-get 3pPOSS village and area again 3p-be*
- 346 Aku nondi ηινε nde lolu rua ηινδε si-o lo ηgonza ilo 347 i-keno lee lee 348 ku sombu.  
*and quartz this TOP youth two that 3p-put in cave inside 3s-be CONT CONT and rotten.*
- 349 Aku si-lo 350 ku si-ra. 351 Rua si-tu 352 si-ra kie 353 si-sama 354 andetia i-ta tia.  
*and 3p-go and 3p-strike two 3p-want 3p-strike foot 3p-test but 3s-cry NEG*
- 355 Si-ra kapwa 356 andetia i-ta tia. 357 Si-ra kulu 358 ande i-ta tia. 359 Si-ra kumbu  
*3p-strike abdomen but 3s-cry NEG 3p-strike head CNTR 3s-cry NEG 3p-strike back*
- 360 ande i-ta tia. 361 ηινεηγα rua si-tu 362 si-ra ηandola 363 ku ηandola i-ta. 364 Wηηgu  
*CNTR 3s-cry NEG then two 3p-want 3p-strike neck and neck 3s-cry drum*
- 0-pwarana nia ηalae ηινδε. 365 Si-ra lo Sio sowa ηινε 366 ku i-lo 367 i-poηga Kavoe ηινδε  
*0-make.noise area large that 3p-strike at Sio beach this and 3s-go 3s-arrive Kavoe that*



368 si-loŋo tava. 369 Opo ŋinde i-ta pɔta ŋinde.  
*3p-hear with wow that 3s-cry INTNS that*

370 Aku tua-nzi tamɔta ŋalae ŋinde i-loŋo wɔŋgu 371 i-ta 372 aŋga i-mora mundo 373  
*and old.sbl-3pPOSS person large that 3s-hear drum 3s-cry CNTR 3s-see smoke*

i-kɔki. 374 Aku i-poro 375 tu "O, naŋa a-pete 376 tu a-mora nana. 377 ŋinde nde naŋa mama  
*3s-go.up and 3s-say thus oh 1s 1s-stroll PURP 1s-see mommy that = 1s mom*

378 i-mo 379 kala i-veta ne mundo 380 i-kɔki. 381 I-kai-nzi tai-ŋgu to."  
*3s-be so 3s-make 3sPOSS smoke 3s-go.up 3s-get-3p ygr.sbl-1sPOSS DUB*

382 Aku i-tapa loa kɔki 383 i-veta kɔpwa 384 i-sowe papatu 385 i-o lo wɔŋga kulu 386 ku  
*and 3s-pull sail up 3s-make food 3s-mix food.nm 3s-p:ʋt on canoe atop and*

i-kɔwa 387 i-toa 388 i-mɔ. 389 I-toa lee 390 i-mɔ 391 ku i-mora tina 392 aku i-poro papa  
*3s-flee 3s-go.ashore 3s-come 3s-go.ashore CONT 3s-come and 3s-see mother and 3s-say to*

393 ku kɔpwa wa puroŋa wa i-lua 394 i-ka. 395 Aku i-kasoŋa 396 "O, ŋga naŋa tai-ŋgu  
*and food and bread and 3s-give 3s-eat and 3s-ask oh CNTR 1s ygr.sbl-1sPOSS*

kinzi si-mo ndia?" 397 "O, tai rua nde si-lo nzo ilo 398 si-veta nenzi tɔno." 399 "Aŋga  
*3p 3p-be where oh ygr.sbl two TOP 3p-go plains inside 3p-make 3pPOSS garden CNTR*

na-loŋo neŋgu wɔŋgu 400 i-ta 401 nde neŋgu wɔŋgu i-keno ndia?" 402 Ande i-pai 403 tu "O,  
*1s-hear 1sPOSS drum 3s-cry and 1sPOSS drum 3s-be where CNTR 3s-tell thus oh*

noko ne wɔŋgu tonŋe tia. 404 Wɔŋgu ŋinde nde tai rua nenzi. 405 Si-pu Nɔndi 406 ande  
*2s 2sPOSS drum a NEG drum that = ygr.sbl two 3pPOSS 3p-hit Nɔndi and*

kala ŋandola 407 ande kala si-ra." 408 "Ara, ku-tula papa-na." 409 Ande i-pai 410 tu "O, ma naŋa  
*so neck and so 3s-strike good 2s-show to-1s CNTR 3s-tell thus oh IRR 1s*

a-tula pa-no tia. 411 ŋinde nde tai rua nenzi."  
*1s-show to-2s NEG that = ygr.sbl two 3pPOSS*

412 I-veta 413 i-veta lee lee 414 ku i-tu 415 i-sowe tina lo ua 416 Tia, tina-nzi  
*3s-do 3s-do CONT CONT and 3s-want 3s-stab mother INSTR spear NEG mother-3pPOSS*

i-ruru ŋana tini, 417 tia ku i-tula papa. 418 I-tula papa 419 ku mbau i-lo 420 i-kai. 421  
*3s-far about skin NEG and 3s-show to 3s-show to and hand 3s-go 3s-g:ʋt*

Mbau i-lo 422 i-kai 423 ku i-ndue 424 ku i-ra lo nzoŋa 425 ku kinzi loŋu rua ŋinde si-loŋo  
*hand 3s-go 3s-get and 3s-go.down and 3s-strike on path and 3p youth two that 3p-hear*

lo nzo ilo. 426 Rua si-palilu. 427 I-ndue 428 i-o !o wɔŋga kulu 429 i-tapa loa kɔki 430 ku  
*in plains inside two 3p-run 3s-go.down 3s-put on canoe atop 3s-pull sail up and*



i-kowa mine-ni 431 i-kawea 432 i-lo pa Kavoe 433 i-koki 434 i-lo.  
*3s-flee like.this-SPEC 3s-snatch 3s-go to Kavoe 3s-go.up 3s.go*

435 Lolu rua si-mo 436 si-veta kura tia 437 ku si-tu 438 si-pane tina-nzi pota. 439  
*youth two 3p-come 3p-do be.able NEG and 3p-want 3p-shoot mother-3pPOSS INTNS*

Andetia tina-nzi i-pai-nzi 440 tu, "Opopo, a-vea ma, 441 ambo taitu nana tua-mi  
*but mother-3pPOSS 3s-tell-3p thus wow 1s-hide EMPH IRR however about old.sbl-2pPOSS*

njinde i-kai ua 442 tu i-sowe-na 443 wa i-tu 444 i-pu-na pota 445 kala a-ruru nana tini-njgu  
*that 3s-get spear PURP 3s-stab-1s and 3s-want 3s-hit-1s INTNS so 1s-fear about skin-1sPOSS*

446 kala a-tula papa 447 kala i-kai 448 i-lo."  
*so 1s-show to so 3s-get 3s-go*

449 I-kura rua ilo-nzi tia ndo. 450 Rua si-poro pota 451 wisi-nzi nana nanae njinde.  
*3s-cover two insides-3pPOSS NEG AMPL two 3p-say INTNS heart-3pPOSS hot large that*

452 Tia ku si-sarawa-nzi sii lawea lawea 453 si-mo. 454 si-sarawa-nzi sii lawea lawea 455  
*NEG and 3p-call.to-3p bird village village 3p-come 3p-call.to-3p bird village village*

roroni si-mo 456 ku si-supwa-nzi sii roroni. 457 Si-supwa-nzi sii roroni 458 si-lo 459 tu  
*all 3p-come and 3p-send-3p bird all 3p-send-3p bird all 3p-go PURP*

si-lo Kavoe 460 nana ma si-kai wonggu njinde pa-nzi. 461 Si-sarawa-nzi 462 si-mo. 463 Andetia  
*3p-go Kavoe PURP IRR 3p-get drum that to-3p 3p-call.to-3p 3p-come but*

sii kiekie roroni si-togo 464 nana si-lo 465 si-kai 466 andeta tia. 467 Si-lo, 468 andetia lawea  
*bird various all 3p-ty about 3p-go 3p-get but NEG 3p-go but wind*

kaika i-yoka 469 i-pu-nzi, 470 kola-lawea i-koki, 471 ku roroni si-taulo 472 si-mo marumbu.  
*strong 3s-move 3s-hit-3p wave-wind 3s-go.up and all 3p-return 3p-come finished*

473 njinejga si-supwa-nzi nzoji. 474 "Ara, miki ma ka-lo." 475 njinejga nzoji si-lo. 476 Sii  
*then 3p-send-3p bird.spec good 2p IRR 2p-go then bird.spec 3p-go bird*

nzoji si-lo 477 si-lou 478 si-lo 479 ku kola i-koki 480 ande si-koki, 481 kola i-ndue 482 ande  
*bird.spec 3p-go 3p-fly 3p-go and wave 3s-go.up and 3p-go.up wave 3s-go.down and*

si-ndue. 483 Lawea i-mo 484 tu i-pu-nzi 485 ande si-ndue 486 s-oka toi kulu. 487 Si-veta  
*3p-go.down wind 3s-come PURP 3s-hit-3p and 3p-go.down 3p-move sea atop 3p-do*

mine no. 488 Si-lo 489 si-lo lee lee 490 ku si-lo 491 si-sajona niu mbau kulu.  
*like.this LMTR 3p-go 3p-go CONT CONT and 3p-go 3p-sit coconut hand atop*

492 Ande mbo 493 ɲana ma Kavɔe si-kina kiniɲa kɔ. 494 Aku mbo 495 ku si-kina kiniɲa.  
*and night about IRR Kavɔe 3p-dance singsing RSTR and night and 3p-dance singsing*

496 Si-kina kiniɲa 497 ku nzoɲi rɔrɔni si-ndue ku-nzi 498 ku si-lɔ ilo-nzi  
*3p-dance singsing and bird.spec all 3p-go.down with-3p and 3p-go insides-3pPOSS*

marumbu 499 aku si-kina kiniɲa ku-nzi. 500 Sii kiri-kiri ndaina mbo si-lɔ 501 ku  
*finished and 3p-dance singsing with-3p bird little-REDUP there night 3p-go and*

si-nde-ndeka pa-nzi 502 wa si-kina kiniɲa ku-nzi.  
*3p-REDUP-rejoice to-3p and 3p-dance singsing with-3p*

503 Si-veta mine 504 ku nzoɲi tɔŋge ande wɔŋgu sakamao si-lua 505 ande i-mboso. 506 "Naɲa  
*3p-do like.this and bird.spec a TOP drum bad 3p-give and 3s-be.upset 1s*

ma ə 'kai wɔŋgu ɲine tia. 507 Naɲa nde tamɔta ɲana kiniɲa kɔ. 508 Wɔŋgu ara aɲga ka-lua  
*IRR 1s-get drum this NEG 1s = person about singsing RSTR drum good CNTR 2p-give*

naɲa." 509 Aku nuwala rɔrɔni si-pai-nzi 510 tu, "Ee, ɲga tamɔta ɲine nde kiniɲa warika. 511  
*1s and agemate all 3p-tell-3p thus hey CNTR person this = singsing foundation*

Ma wɔŋgu sakamao ta-lua 512 ma i-kina kiniɲa ara tia. 513 Wɔŋgu ara ta-lua 514 ma  
*IRR drum bad 1pIN-give IRR 3s-dance singsing good NEG drum good 1pIN-give IRR*

i-kina kiniɲa pɔta lɔ." 515 Aku kinzi si-tu 516 wɔŋgu ara tɔŋge si-lua 517 ande tia 518  
*3s-dance singsing INTNS PERF and 3p 3p-want drum good a 3p-give CNTR NEG*

i-mboso 519 tini pwɔka. 520 "Naɲa tini-ɲgu pwɔka wɔŋgu ɲine."  
*3s-be.upset skin wrinkled 1s skin-1sPOSS wrinkled drum this*

521 Si-veta 522 si-veta lee 523 ku i-lɔŋɔ wɔŋgu-ni Nɔndi ɲandola ɲinde 524 si-ra 525 nde  
*3p-do 3p-do CONT and 3s-hear drum-SPEC Nɔndi neck that 3p-strike and*

i-lɔ. 526 I-ta pɔta lɔ. 527 Aku si-pai-nzi 528 tu, "Etia, wɔŋgu ɲinde ka-lua i 529 aɲga ma  
*3s-go 3s-cry INTNS PERF and 3p-tell-3p thus yes drum that 2p-give 3s CNTR IRR*

ara". 530 Aku wɔŋgu ɲinde si-lua i 531 i-kai 532 ku i-lɔŋɔ nɔ 533 i-ki 534 i-nzare  
*good and drum that 3p-give 3s 3s-get and 3s-deceive LMTR 3s-dance 3s-spread.into*

nia lɔ ɲgaɲe-ɲgarɲe 535 i-lɔ 536 wa i-mɔ wa. 537 I-lɔŋɔ nɔ 538 i-ki pɔta lɔ ɲinde.  
*area at edge-REDUP 3s-go and 3s-come and 3s-deceive LMTR 3s-dance INTNS PERF that*

539 Aku sii-ni kiri-kiri-ni ndaina si-ki 540 wa si-ndeka ku-nzi. 541 Si-ki 542 si-ki  
*and bird-SPEC small-REDUP-SPEC there 3p-dance and 3p-rejoice with-3p 3p-dance 3p-dance*



lee 543 ku ilo-kakana kɔ, ɲineŋga mbau-nzi 0-pareŋge wɔŋgu rɔrɔni 544 ku si-kai Nɔndi  
 CONT and daybreak RSTR then hand-3pPOSS 0-open drum all and 3p-get Nɔndi

ɲandola ɲinde 545 si-kawea tɔɔ nɔ 546 si-kɔki 547 si-lɔ niu mbau kulu 548 si-saŋona.  
 neck that 3p-snatch quickly LMTR 3p-go.up 3p-go coconut hand atop 3p-sit

549 Aku i-tarɔra 550 "Wa! ka-kai neŋgu wɔŋgu pa-na 551 ka-mɔ ɲga tɔ, 552 ka-veta mine  
 and 3s-scream hey 2p-get 1sPOSS drum to-1s 2p-come EXCL DUB 2s-do like.this

tia!" 553 I-sarawa pa-nzi 554 tu, "Ka-kai pa-na 555 ka-mɔ ɲga!" 556 Tia mbau-nzi 0-pareŋge  
 NEG 3s-call to-3p thus 2p-get to-1s 2p-come EXCL NEG hand-3pPOSS 0-open

ɲinde rɔrɔni 557 0-bururu ndue 558 ku si-lɔ taitu 559 ku si-kai wɔŋgu ɲinde 560 ku si-kawea  
 that all 0-crash down and 3p-go one and 3p-get drum that and 3p-snatch

561 si-kai 562 si-toa 563 si-mɔ. 564 Si-kai lee lee 565 ku si-toa 566 si-mɔ 567 si-la-nzi  
 3p-get 3p-go.ashore 3p-come 3p-get CONT CONT and 3p-go.ashore 3p-come 3p-give-3p

ɲɔsi ku wia rua.  
 left and right two

568 ɲineŋga rua si-veta kɔpwa ɲalae ɲinde pa-nzi sii nzoŋi kiri-kiri. 569 Si-ka 570 wa  
 then two 3p-make food large that to-3p bird bird.spec little-REDUP 3p-eat and

si-nu lɔ marumbu 571 ɲineŋga si-poro 572 tu, "Ara mbo, ma ɲineŋga ma maka ka-kai  
 3p-drink PERF finished then 3p-say thus good AMPL IRR then IRR 1pEX 1pEX-get

nema wɔŋgu ɲine lɔ mbau-ma lɔ, 573 ɲineŋga ma ka-o ɲana tamɔta ɲine 574 ma  
 1pEXposs drum this in hand-1pEXposs PERF then IRR 1pEX-wait about person this IRR

i-toa 575 i-mɔ." 576 ɲineŋga rua si-mɔ 577 ɲineŋga si-o ɲana tua-nzi.  
 3s-go.ashore 3s-come then two 3p-be then 3p-wait about old.sbl-3pPOSS

578 Rua si-mo lawea ilo nɔ 579 ku mata-nzi 0-kɔki lɔ tɔi ɲgini 580 nde si-mora  
 two 3p-be village inside LMTR and eye-3pPOSS 0-go.up at sea middle and 3p-see

tua-nzi ne wɔŋga ɲana loa salaga-salaga nɔ 581 ku i-mɔ. 582 "Ara, ku-mɔ." 583 Aku  
 old.sbl-3pPOSS 3sPOSS canoe about sail flash-REDUP LMTR and 3s-come good 2s-come and

i-yoka mine-ni lee 584 ku i-toa 585 ku kinzi rua si-lɔ 586 ku si-muna. 587 I i-kai  
 3s-move like.this-SPEC CONT and 3s-go.ashore and 3p two 3p-go and 3p-hide 3s 3s-get

ne temba 588 i i-kai ne temba. 589 Rua si-lɔ 590 ku si-muna 591 ku i-yoka  
 3sPOSS bow 3s 3s-get 3sPOSS bow two 3p-go and 3p-hide and 3s-move

mine-ni 592 ma i-toa sowa 593 ku ua nɔ i-saru lɔ 594 i-sowe tɔno 595 ku  
 like.this-SPEC IRR 3s-go.ashore beach and spear INST 3s-wave PERF 3s-stab ground and



i-sawati koki lo. 596 Aku i-poro 597 tu, "Naga tini-ngu mine 598 nga ma na-pu tamota!"  
*3s-spray up PERF and 3s-say thus 1s skin-1sPOSS like.this EMPH IRR 1s-hit person*

599 I-mbomba 600 i-yoka mine 601 i-toa lo 602 ku i-kasonga tina, 603 "Tamota sakamao rua  
*3s-hurry 3s-move like.this 3s-go.ashore PERF and 3s-ask mother person bad two*

mine nde kinzi si-kai nanenggu wongu 604 si-kai 605 si-mo. 606 Kinzi si-mo ndia?" 607 Ande  
*like.this TOP 3p 3p-get 1sPOSS drum 3p-get 3p-come 3p 3p-be where CNTR*

i-pai natu ngalae njinde, 608 "Wa, naga a-pai-no 609 tu ku-pile 610 nde waraka-nzi rua nenzi  
*3s-tell child large that hey 1s 1s-tell-2s thus 2s-leave = self-3pPOSS two 3pPOSS*

kelekele, 611 ma ku-veta 612 ma ku-pa-yaula ko." 613 Ande i-pai tina 614 tu, "Kinzi rua  
*thing IRR 2s-do IRR 2s-RFX-damage RSTR CNTR 3s-tell mother thus 3p two*

sakamao rua mine si-kura nenzi so?"  
*bad two like.this 3p-be.able 3pPOSS what*

615 I-poro mine lo 616 ku marumbu, 617 mbau i-lo 618 ku i-kai Nondi ngandola mbwendu  
*3s-say like.this PERF and finished hand 3s-go and 3s-get Nondi neck throat*

wongu njinde 619 ku i-kai 620 ku i-mbomba 621 i-ndue 622 i-mo. 623 I-ndue 624 i-mo 625 ku  
*drum that and 3s-get and 3s-hurry 3s-go.down 3s-come 3s-go.down 3s-come and*

i-pongga laiti 626 ku ngosi i-rai 627 wia i-rai 628 i-pane pa kambasi pinde 629 i i-pane pa kambasi  
*3s-arrive close and left 3s-draw right 3s-draw 3s-shoot to armpit some 3s 3s-shoot to armpit*

pinde 630 ku tua-nzi i-wuwura lo 631 i-pu nia lo 632 i-keno.  
*some and old.sbl-3pPOSS 3s-stagger PERF 3s-hit area PERF 3s-lie*

633 Aku rua si-pu pota 634 i-mote 635 ku rua si-kai 636 i-lo 637 si-tambira lo wongga kulu  
*and two 3p-hit INTNS 3s-die and two 3p-get 3s-go 3p-throw.down on canoe atop*

638 ku si-tapa ne loa koki, 639 poe si-pa tava wongga saporu, 640 ku marumbu. 641  
*and 3p-pull 3sPOSS sail up paddle 3p-fasten with canoe brace and finished*

Si-zuara 642 ku i-ndue 643 ku i-lo mota ilo 644 ku i-sakoko 645 i-koki papa ne lawea  
*3p-shove and 3s-go.down and 3s-go passage inside and 3s-drift 3s-go.up to 3sPOSS village*

Kavoe. 646 Aku kinzi rua si-taulo 647 si-mo 648 ku si-mo nenzi lawea.  
*Kavoe and 3p two 3p-return 3p-come and 3p-be 3pPOSS village*

649 Kala ngaina.  
*so here*

## REFERENCES CITED

- Brown, Gillian, and George Yule. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.
- Carlson, Robert. 1987. Narrative connections in Supyire. In *Coherence and grounding in discourse*, ed. by Russell Tomlin, 1-16. Amsterdam: John Benjamins.
- Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and topic*, ed. by Charles Li, 25-55. New York: Academic Press.
- Clancy, P. 1980. Referential choice in English and Japanese narrative discourse. In *The pear stories: cognitive, cultural and linguistic aspects of narrative production*, ed. by Wallace L. Chafe. Norwood, NJ: Ablex.
- Clark, Dawn. 1994. Phonology of the Sio language. In *Data papers*. Ukarumpa, PNG: Summer Institute of Linguistics.
- Clark, H., and S. Haviland. 1977. Comprehension and the given-new contrast. In *Discourse production and comprehension*, ed. by R. Freedle, 1-40. Hillsboro, NJ: Erlbaum.
- Clark, Stephen, and Dawn S. Clark. 1987. *Sio grammar essentials*. Ukarumpa, PNG: Summer Institute of Linguistics.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: syntax and morphology*. 2<sup>nd</sup> ed. Chicago, IL: University of Chicago Press.
- Cooreman, Ann. 1983. Topic continuity and the voicing system of an ergative language: Chamorro. In *Topic continuity in discourse: a quantitative cross-language study*, ed. by T. Givon, 429-474. Philadelphia, PA: John Benjamins.
- Dooley, Robert A., and Stephen H. Levinsohn. 1999. *Analyzing discourse: basic concepts*. (Prepublication draft). Online at:  
<http://www.und.nodak.edu/dept/linguistics/textbooks/DooleyLevinsohn.pdf>
- Evangelical Lutheran Church of Papua New Guinea. 1985. *The ELC in PNG: the first 100 years*. Madang, PNG: Kristen Pres, Inc.
- Fox, Andrew. 1983. Topic continuity in Biblical Hebrew narrative. In *Topic continuity in discourse: a quantitative cross-language study*, ed. by T. Givon, 219-258. Philadelphia, PA: John Benjamins.
- Fox, Barbara. 1987. Anaphora in popular written English narratives. In *Coherence and grounding in discourse*, ed. by Russell Tomlin, 156-172. Amsterdam: John Benjamins.
- Givon, Talmy, ed. 1983. *Topic continuity in discourse*. Amsterdam: John Benjamins.
- \_\_\_\_\_. 1983. Topic continuity and word-order pragmatics in Ute. In *Topic continuity in discourse*, ed. by T. Givon, 145-195. Amsterdam: John Benjamins.



- \_\_\_\_\_. 1990. *Syntax: a functional-typological approach*. 2 vols. Amsterdam: John Benjamins.
- Green, Georgia. 1996. *Pragmatics and natural language understanding*. Mahwah, NJ: Lawrence Erlbaum.
- Grice, H. Paul. 1975. *Logic and conversation*. In *Syntax and semantics, vol. III: speech acts*, ed. by Peter Cole and Jerry L. Morgan, 41-58. New York: Academic Press.
- Grimes, Joseph E. 1975. *The thread of discourse*. The Hague: Mouton.
- Grimes, Barbara F., ed. 1996. *Ethnologue: the languages of the world*. 11<sup>th</sup> ed. Dallas, TX: Summer Institute of Linguistics.
- Halliday, M.A.K. 1978. *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.
- Harding, Thomas. 1967. *Voyagers of the Vitiaz Strait*. Seattle, WA: U. of Washington Press.
- Hopper, Paul J., and Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60:703-752.
- Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus, and the mental representation of discourse referents*. New York: Cambridge University Press.
- Levinsohn, Stephen H. 2000. *Discourse features of New Testament Greek*. 2<sup>nd</sup> ed. Dallas, TX: SIL International.
- Longacre, Robert E. 1996. *The grammar of discourse*. 2<sup>nd</sup> ed. New York: Plenum.
- Napoli, Donna. 1996. *Linguistics: an introduction*. New York: Oxford University Press.
- Prince, Ellen F. 1981. *Toward a taxonomy of given-new information*. In *Radical pragmatics*, ed. by P. Cole, 223-255. New York: Academic Press.
- \_\_\_\_\_. 1992. The ZPG letter: subjects, definiteness, and information status. In *Discourse description: diverse linguistic analyses of a fund-raising text*, ed. by W. C. Mann and S. A. Thompson, 295-325. Philadelphia, PA: John Benjamins.
- Schiffin, Deborah. 1994. *Approaches to discourse*. Cambridge, MA: Blackwell.
- Tomlin, Russell, ed. 1987. *Coherence and grounding in discourse*. Amsterdam: John Benjamins.
- \_\_\_\_\_. 1987. Linguistic reflections of cognitive events. In *Coherence and grounding in discourse*, ed. by Russell Tomlin, 455-479. Amsterdam: John Benjamins.