Participant Reference and Tracking in San Francisco Ozolotepec Zapotec

Jennifer Lynn Córdova

SIL International
2009
Fair Use Policy
Books published in the SIL e-Books (SILEB) series are intended for scholarly research and educational use. You may make copies of these publications for research or instructional purposes free of charge (within fair use guidelines) and without further permission. Republication or commercial use of SILEB or the documents contained therein is expressly prohibited without the written consent of the copyright holder(s).

Series Editor
Mary Ruth Wise

Volume Editor
Mickey Brussow

Managing Editor
Bonnie Brown

Compositor
Margaret González
Contents

List of Tables
List of Figures
Acknowledgments
Abbreviations
Abstract

1 Introduction
1.1 Rationale for the study
1.2 Theoretical framework
1.3 San Francisco Ozolotepec Zapotec
1.4 Scope of the study
1.5 Elements to be studied
1.6 Definition of terms
1.7 Organization of this study

2 Overview of San Francisco Ozolotepec Zapotec
2.1 Phoneme inventory and orthography
2.2 Grammar sketch
   2.2.1 Word order
   2.2.2 Nouns
   2.2.3 Number
   2.2.4 Demonstrative adjectives
   2.2.5 Pronouns
   2.2.6 Null reference
   2.2.7 Verbal morphology
   2.2.8 Embedded clause structure
   2.2.9 Sentence boundaries

3 Literature Review
3.2 Ariel (1988, 2001) – Accessibility theory
3.3 Longacre (1995)
3.4 Dooley and Levinsohn (2001)
3.5 Anderson, Garrod, and Sanford (1983) – Episode shifts
3.6 Previous studies of participant reference in other Zapotec languages
3.7 Summary

4 Methodology Used
4.1 Selection of corpus of texts
4.2 Division of texts into clauses
4.3 Color-coding participants
4.4 Inventory of encoding devices
4.5 Referential distance
4.6 Potential interference
4.7 Persistence
4.8 Tracking chart
4.9 Labeling contexts of each reference
4.10 Prediction of defaults
4.11 Identification of underencoding and overencoding
4.12 Proposal of motivations for special cases

5 Results
5.1 Inventory of encoding devices
5.2 Referential distance
5.3 Potential interference
5.4 Persistence
5.5 Proposed default encodings
5.6 Other factors
  5.6.1 Pronoun restriction
  5.6.2 Introductions
  5.6.3 Episode boundaries
  5.6.4 Participant ranking
  5.6.5 Very Important Participants (VIPs)
  5.6.6 Grammar constraints
5.7 Marked encodings (underencoding and overencoding)
5.8 Proposed motivations for nondefault encodings

6 A Closer Look At One Text

7 Conclusions
7.1 Default encodings
7.2 Null reference
7.3 Nouns
7.4 Other encoding devices
7.5 Implications for theory
7.6 Implications for translation
7.7 Strengths and weaknesses of the study
7.8 Recommendations for further research

Appendix: Corpus of Texts
  Me'dz ichop mgiy
  Tsi¹ bur
  San Bartolo

References
List of Tables

Table 1. SFOZ vowels
Table 2. SFOZ consonants
Table 3. Nouns (unmarked for number)
Table 4. Demonstrative adjectives
Table 5. SFOZ pronouns
Table 6. Three tasks of a scheme of reference
Table 7. Corpus of texts
Table 8. Context categories
Table 9. MM chart for clauses 1 through 11
Table 10. Inventory of encoding devices
Table 11. Referential distance calculations
Table 12. Potential interference
Table 13. Potential interference values across all three texts
Table 14. Persistence across all three texts
Table 15. Persistence values across all three texts
Table 16. Persistence values in TB
Table 17. Percentage of each participant’s references encoded by given device
Table 18. Persistence values in MM
Table 19. Percentage of encoding device references used to refer to each participant
Table 20. Default encoding for subject contexts
Table 21. Default encoding for nonsubject contexts
Table 22. Episode boundaries in MM text
Table 23. Participant ranking in MM
Table 24. Distribution of marked references
Table 25. Revised default encodings
Table 26. Referential distance values in MM
Table 27. Potential interference values in MM
Table 28. Setting (Clauses 1–7)
Table 29. Episode one (Clauses 8–21)
Table 30. Episode two (Clauses 22–34)
Table 31. Episode three (Clauses 35–44)
Table 32. Episode four (Clauses 45–56)
Table 33. Episode five (Clauses 57–69)
Table 34. Episode six (Peak) (Clauses 70–89)
Table 35. Episode seven (Clauses 90–98)
Table 36. Conclusion (Clauses 99–100)
Table 37. SFOZ Default encodings

List of Figures

Figure 1. Verbal structure
Acknowledgments

Many thanks to Shin Ja Hwang, my committee chairperson, who walked me through the thesis experience, encouraging me along the way. Her encouragement and expertise in discourse analysis has meant a lot to me. I also thank my other committee members, Bill Merrifield for his persistent editing and Les Bruce for his helpful suggestions and comments.

I am indebted to the people of San Francisco Ozolotepec for their patience as I continue to learn their language and ask them to tell me story after story.

I also thank my family for their encouragement, my colleagues of the Mexico branch of SIL from whom I have learned so much, and other students, staff, and faculty here at GIAL (Graduate Institute of Applied Linguistics) that have been willing to help at many times and in many ways throughout this process.

Above all, I would like to thank God, who has been my ever-faithful companion throughout this task and all others.
Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADV</td>
<td>adverb</td>
<td>!</td>
</tr>
<tr>
<td>AFM</td>
<td>aforementioned</td>
<td>?</td>
</tr>
<tr>
<td>CC</td>
<td>complement clause</td>
<td>1E</td>
</tr>
<tr>
<td>COMP</td>
<td>completive</td>
<td>1I</td>
</tr>
<tr>
<td>DO</td>
<td>direct object</td>
<td>1S</td>
</tr>
<tr>
<td>DSTR</td>
<td>distributive</td>
<td>2S</td>
</tr>
<tr>
<td>FOC</td>
<td>focus marker</td>
<td>3IN</td>
</tr>
<tr>
<td>FUT1</td>
<td>future</td>
<td>3S</td>
</tr>
<tr>
<td>FUT2</td>
<td>future</td>
<td></td>
</tr>
<tr>
<td>HAB</td>
<td>habitual</td>
<td></td>
</tr>
<tr>
<td>IO</td>
<td>indirect object</td>
<td></td>
</tr>
<tr>
<td>IMP</td>
<td>imperative</td>
<td></td>
</tr>
<tr>
<td>IMPP</td>
<td>imperative plural</td>
<td></td>
</tr>
<tr>
<td>IPA</td>
<td>International Phonetic Alphabet</td>
<td></td>
</tr>
<tr>
<td>INT</td>
<td>interrogative</td>
<td></td>
</tr>
<tr>
<td>MM</td>
<td>Me’dz ichop mgiy</td>
<td>‘The puma and the man’ (text)</td>
</tr>
<tr>
<td>N</td>
<td>number of items</td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>NOM</td>
<td>nominalizer</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>object</td>
<td></td>
</tr>
<tr>
<td>OBL</td>
<td>oblique</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>persistence</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>potential interference</td>
<td></td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>relative clause</td>
<td></td>
</tr>
<tr>
<td>RD</td>
<td>referential distance</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>subject</td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>San Bartolo</td>
<td>‘San Bartolo’ (text)</td>
</tr>
<tr>
<td>SFO</td>
<td>San Francisco Ozolotepec</td>
<td></td>
</tr>
<tr>
<td>SFOZ</td>
<td>San Francisco Ozolotepec Zapotec</td>
<td></td>
</tr>
<tr>
<td>STAT</td>
<td>stative</td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td>Tsi’ bur</td>
<td>‘Ten donkeys’ (text)</td>
</tr>
<tr>
<td>UNR</td>
<td>unreal</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>vowel or verb</td>
<td></td>
</tr>
<tr>
<td>VSO</td>
<td>verb, subject, object word order</td>
<td></td>
</tr>
</tbody>
</table>
Abstract

This study analyzes the encoding devices that refer to and track participants throughout third-person narrative texts in San Francisco Ozolotepec Zapotec (SFOZ). SFOZ is an Oto-Manguean language of southern Mexico. Without a third-person pronoun (other than that used to show respect), SFOZ relies on null reference, nouns, focused nouns, NPs and focused NPs to refer to participants. Givón’s (1983, 1994) quantitative method and Dooley and Levinsohn’s (2001) charting method for analyzing participant references were applied to three SFOZ texts. Dooley and Levinsohn’s method proved most helpful by showing general default encodings for varied contexts, use of overencoded references following episode boundaries, frequent use of nouns for nonsubject references, as well as distinct referential patterns based on participant rank.


About the author

Jennifer Heise earned a Master’s degree in Applied Linguistics from the Graduate Institute of Applied Linguistics (2003) in Dallas, Texas. She also has a B.A. degree in Spanish and a B.S. degree in Elementary Education from Houghton College (1998) in Houghton, New York. She has done field work under the auspices of SIL International in Mexico studying the Zapotec language of San Francisco Ozolotepec since 2001. She married Jonatán Córdova in 2006 and continues the study of Mexican indigenous languages with her husband.
1
Introduction

1.1 Rationale for the study

To a native English speaker listening to natural San Francisco Ozolotepec Zapotec discourse, constant repetition of nouns stands out as unusual. Why should so many such forms be used to track participants that have already been clearly identified when there are simpler forms that could be used? And yet, nonnative speakers of Zapotec sometimes fail to properly identify multiple participants because they are encoded by the same devices. Since the use of a third-person pronoun is relatively uncommon, what forms are used instead? What are the rules for referring to third-person participants in narrative discourse?

This question is the focus of this study; namely, the rules of participant reference and tracking in San Francisco Ozolotepec Zapotec (SFOZ). The specific questions to be answered are:

1. What are the default encodings for tracking participants in the various contexts of narrative discourse?
2. When and why is null reference used?
3. When and why are nouns used?
4. When and why are other encoding devices (including the rare pronouns) used?

Answering these questions will contribute to the proper understanding of SFOZ narrative discourse. It will also aid in producing more natural and satisfying written literature in SFOZ. Though preliminary research on participant reference has been done in related Zapotec languages, this research will provide additional data that can further comparative Zapotec studies and may shed new light on processes that exist in Zapotec languages that have not yet been studied.

1.2 Theoretical framework

This study assumes that language cannot adequately be explained simply in terms of phrases or even clauses. Rather, the study of entire discourses can also shed important light on our understanding of language. As Longacre (1996:1) says, “language is language only in context.” Thus, looking at the discourse context is important in understanding any language.

There are, indeed, a wide range of discourse theories that can each share a part in aiding the understanding of discourse. Linguists sometimes desire to find one comprehensive theory that accurately deals with participant reference, while viewing any lack of concordance among theories as a weakness. In this study, I draw on the great diversity of perspectives in this field and considered it a strength, rather than a weakness (Schiffrin, Tannen, and Hamilton 2001:5). Though I will primarily be using the methodology employed by Givón (1983, 1994) and Dooley and Levinsohn (2001), I will also draw on other theories as they seem relevant. These theories and methodologies will be presented more fully in chapter three.
1.3 San Francisco Ozolotepec Zapotec

Gordon (2005:259) lists 291 languages spoken today in Mexico. As a member of the Oto-Manguean stock, Zapotec is one of many indigenous language families in Mexico’s southern State of Oaxaca. According to Merrill and Miró (1997:97), Zapotec speakers account for about 7% of Mexico’s indigenous population. SFOZ is but one of fifty-eight varieties of Zapotec (Gordon 2005:277–281). There are approximately two thousand speakers of this Zapotec language, including speakers from Santa Catarina Xanaguía and San José Ozolotepec, as well as San Francisco Ozolotepec (INEGI, 2000). The data presented in this study has been collected solely from speakers who live in San Francisco Ozolotepec (SFO).

The town of San Francisco Ozolotepec is located in the Sierra Sur region of Oaxaca, about eighty-seven miles southeast of Oaxaca City. It is located on the side of a mountain, surrounded by a large canyon. Language vitality is high, as evidenced by language use by all sectors of the community, from children to the oldest living generation.

Contact with this community began in April, 2001, when I moved to San Francisco Ozolotepec to study SFOZ. The data presented in this study was collected over the following two and a half years, during which time the orthography was revised and literacy efforts were begun.

1.4 Scope of the study

I have delimited this study by language, genre, and types of participant references to be studied, and restricted this investigation solely to the Zapotec language of San Francisco Ozolotepec. I have chosen to focus only on monologue narrative discourse. Participant reference is most relevant to the narrative genre, so this is a natural starting point for a study of this kind. I also have delimited the type of participant references that will be examined. Though some of the texts investigated have first-person participants, I have focused primarily on references to third-person animate participants, i.e., those animate participants besides the narrator/speaker and hearer. Since there are more options of how to refer to third-person participants than for first- and second-person participants, more insights can be drawn from third-person references.

As a nonnative speaker, I have not mastered all the intricacies of SFOZ grammar. Ideally, a native speaker of SFOZ, with linguistic training, would have collaborated in my efforts. However, since I know of no such speakers, my restricted knowledge of SFOZ is a limiting factor for this study. Moreover, in writing this thesis I did not have access to native speakers for checking data as questions arose. This limitation has left questions for further research that might otherwise have been resolved.

1.5 Elements to be studied

In order to answer the questions posed regarding participant reference and tracking in SFOZ, numerous elements must be examined. First, an inventory of possible encoding devices for third-

---

1 INEGI 2000 census lists total population for the three towns as 2,007. Limited to just SFO, the population is 910.
2 In the *Ethnologue* (Gordon 2005:281), the language spoken in SFO is actually listed under the language name of Zapoteco de Xanaguía (ZTG). However, due to minor differences in the dialects, including pronouns and pronoun usage, this study is limited to the particular dialect spoken in SFO.
3 The orthography was revised from a previous orthography that had been developed in Santa Catarina Xanaguía by Mary Hopkins and Julie Olive, SIL linguists, who had worked in that town from 1985–1995.
person participants must be collected. Primary focus will be given to the forms found in three texts chosen as a basis for this study: *Me’dz ichop mgíy* ‘The puma and the man’, *Tsi’l bur* ‘Ten donkeys’, and *San Bartolo* ‘San Bartolo’. Second, the varying contexts in which each reference is found will be delineated in order to compare and contrast their usage, with specific reference to whether the participant (1) was involved in the previous clause or not, (2) was speaking or being addressed during a dialogue when the reference was used, (3) has been on stage, or (4) is being newly activated in the narrative. Third, grammatical restrictions that would automatically restrict referential choice will be explored.

### 1.6 Definition of terms

Many of the terms used in this thesis will already be familiar to readers. Some terms, however, have been used by linguists in different ways. In order to avoid confusion, I list the following terms, with the definitions that apply to them in this study.

**PARTICIPANT** – an animate being that is mentioned within a narrative.

**ENCODING DEVICE** – any word or phrase used to refer to a particular participant. Pronouns, nouns, noun phrases, and null reference are examples of encoding devices used in SFOZ.

**REFERENCE** – the linguistic form, such as a pronoun or noun phrase, used to refer to a participant within a text.

**REFERENT** – the entity (in this case a participant) that is being referred to.

**ENCLITIC PRONOUN** – a linguistic form that refers to a participant and is phonologically bound to a verb, noun, or other element and cannot stand alone.

**INDEPENDENT PRONOUN** – a linguistic form that refers to a participant and can phonologically stand alone.

**NULL REFERENCE** – a subject or object that is not stated explicitly in any form. Also referred to as a zero or null.

**REFERENTIAL DISTANCE** – the number of clauses found between the present reference to a participant and the preceding reference to the same participant.

**POTENTIAL INTERFERENCE** – other participants on stage that could potentially cause ambiguity in participant identification by being on stage.

**PERSISTENCE** – measure of the longevity of a topic.

---

4 I will refer to these texts with the following abbreviations:

- **MM** = *Me’dz ichop mgíy* ‘The puma and the man’
- **TB** = *Tsi’l bur* ‘Ten donkeys’
- **SB** = *San Bartolo* ‘San Bartolo’
OVERENCODING – use of more encoding material than the default.

UNDERENCODING – use of less encoding material than the default.

1.7 Organization of this study

This section introduces the topic of participant reference and the purpose of studying it in San Francisco Ozolotepec Zapotec. The remainder of this study is structured as follows: Chapter 2 gives a framework for understanding the grammar of SFOZ which is referred to in this study. Chapter 3 presents a review of the literature about participant reference and relevant studies done in other Zapotec languages. Chapter 4 presents the design and methodology of the research presented in this thesis. Chapter 5 presents the results of applying Givón’s (1983, 1994) and Dooley and Levinsohn’s (2001) methodologies to the SFOZ texts. Chapter 6 analyzes the MM text in detail. Chapter 7 presents the conclusions drawn from analyzing the findings in SFOZ and the implications for translation work and future study.
Overview of San Francisco Ozolotepec Zapotec

2.1 Phoneme inventory and orthography

In order to read the Zapotec examples used throughout this study, the reader should become familiar with the SFOZ phoneme inventory and orthography. These are presented in this section.

Vowels. SFOZ has six vowels, as displayed in table 1. Orthographic representation is given in parentheses where different from the International Phonetic Alphabet (IPA).

Table 1. SFOZ vowels

<table>
<thead>
<tr>
<th></th>
<th>front</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>mid</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>low</td>
<td>æ (ã)</td>
<td>a</td>
</tr>
</tbody>
</table>

Each of these six vowels can also be glottalized, resulting either in a checked vowel [Vʔ] or fully rearticulated vowel [VʔV]. Since these two varying phonetic outcomes are not contrastive, all glottalized vowels are simply written as a vowel followed by glottal, as in niʔb ‘boy’.

Consonants. Table 2 displays the consonants of SFOZ. The orthographic conventions used in examples are given in parentheses where these differ from the IPA.

---

5 In my analysis, glottalization in SFOZ is considered a vowel feature rather than a consonant.
Table 2. SFOZ consonants

<table>
<thead>
<tr>
<th>Stops</th>
<th>Bilabial</th>
<th>Alveolar</th>
<th>Palato-Alveolar</th>
<th>Velar</th>
<th>Palato-Velar</th>
<th>Labio-Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>fortis</td>
<td>p</td>
<td>t</td>
<td>k</td>
<td>k*</td>
<td>k*</td>
<td>gi</td>
</tr>
<tr>
<td>lenis</td>
<td>b</td>
<td>d</td>
<td></td>
<td>g</td>
<td>g'</td>
<td></td>
</tr>
<tr>
<td>Affricates</td>
<td>fortis</td>
<td>ts</td>
<td>tʃ (ch)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lenis</td>
<td>dz</td>
<td>dʒ (ch)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives(^a)</td>
<td>fortis</td>
<td>s</td>
<td>s̪ (x)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lenis</td>
<td>z</td>
<td>ʒ (x)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>fortis(^b)</td>
<td>n: (nn)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lenis</td>
<td>m</td>
<td>n</td>
<td>ñ (ń)</td>
<td></td>
<td>n̈ (uñg)</td>
<td></td>
</tr>
<tr>
<td>Laterals</td>
<td>fortis</td>
<td>l̃: (l̃)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lenis</td>
<td>l̃̃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td></td>
<td>j (y)</td>
<td></td>
<td>w</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Additional consonants occur in words borrowed from Spanish. These additional consonants are written with the orthographic conventions of Spanish. Alveolar flap r and trill rr, voiceless velar fricative j [x], and voiceless labiovelar fricative f [s].

\(^b\) It is convenient to talk of fortis/lenis pairs in SFOZ even when the actual phonetic correlates that distinguish one from the other are not known. In the case of the voiced alveolar nasal and voiced alveolar lateral, I hypothesize that length is what distinguishes fortis from lenis.

\(^c\) The fortis and lenis laterals are not distinguished in the orthography.

2.2 Grammar sketch

Although the focus of this study is at the discourse level, it is important to understand some of the syntactic features of SFOZ. I briefly describe those grammatical features relevant to this research in the sections that follow.

2.2.1 Word order

SFOZ basic word order is VSO (verb, subject, object). The verb is followed immediately by the subject, and then by the direct object when the clause is transitive. The following clauses demonstrate the VSO word order. The intransitive clause in (1) shows VS ordering, while the transitive clause in (2) shows the complete VSO ordering.

(1) \(Na-na \ uy-a'n \ xin \ minn.\)

and-now COMP-stay.behind child person

Now her children have remained. [SB38]^6

\(^6\) Examples with information in brackets can be found in the appendix. The letters indicate the text and the number indicates the clause.
When an indirect object is also present in the clause, it generally follows the direct object.

Obliques generally occur following all three arguments of a clause. The following example shows a clause including an oblique, where the oblique ‘at eight thirty’ follows all arguments (in this case only two arguments are present).

Obliques can be fronted to a preverbal position to serve as discourse markers of transition, as in example (5).

2.2.2 Nouns

Nouns are often used to encode participants in SFOZ. They can refer to an entire class in a generic sense or be used to refer to a specific identifiable participant. As mentioned in the introduction, it is quite common in SFOZ to use nouns to refer to specific participants. Since these nouns will be a relevant part of this study, it fits to list some of the most common ones.
Table 3. Nouns (unmarked for number)

<table>
<thead>
<tr>
<th>noun</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>minn</td>
<td>person</td>
</tr>
<tr>
<td>mgiy</td>
<td>man</td>
</tr>
<tr>
<td>dad</td>
<td>gentleman</td>
</tr>
<tr>
<td>ni'b</td>
<td>boy</td>
</tr>
<tr>
<td>una'</td>
<td>woman</td>
</tr>
<tr>
<td>nay</td>
<td>lady</td>
</tr>
<tr>
<td>ndzap</td>
<td>girl</td>
</tr>
<tr>
<td>man</td>
<td>animal</td>
</tr>
<tr>
<td>gu'n</td>
<td>thing</td>
</tr>
</tbody>
</table>

2.2.3 Number

SFOZ nouns are unmarked for number; there is no plural marker. That is to say, bich can refer to ‘one cat’, ‘two cats’, or ‘multiple cats’. The lack of plural marking can cause ambiguity when it comes to participant reference. The hearer may be left wondering how many participants are being referred to by a particular expression. In example (6), two different nouns are used to refer to a male participant in the story: mgiy ‘man’ and minn ‘person’. In the context of the discourse, the referent is clearly singular. As a stand-alone phrase, however, the free translation of these nouns could be singular, as in (6a), or plural, as in (6b). This demonstrates the ambiguity in participant reference created by lack of plural marking.

(6) Lël mgiy, minn ine'ch lo- nu,
FOC man/men person/people FUT2-go-ahead face-1E

   a. The man, the person who would guide us [SB8]
   b. The men, the people who would guide us

References to plural participants are not always ambiguous, however. Their plurality can be indicated by modifying nouns with quantifier adjectives. When used, quantifiers precede the noun they modify (as opposed to qualitative adjectives, which follow the noun). In the following example, chop ‘two’ modifies mgiy ‘man’ to avoid ambiguity.

(7) Ts-u chop mgiy.
HAB-stand two man
Two men were standing. [SB6]

2.2.4 Demonstrative adjectives

Another means of encoding participants in SFOZ is the use of anaphoric demonstrative adjectives as noun modifiers. These demonstrative adjectives refer back to participants that were already mentioned. They follow the nouns they modify. Table 4 lists the four demonstrative adjectives available in SFOZ.
Table 4. Demonstrative adjectives

<table>
<thead>
<tr>
<th>Demonstrative adjective</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>che'n</td>
<td>this (when visible)</td>
</tr>
<tr>
<td>ki</td>
<td>that (when visible)</td>
</tr>
<tr>
<td>nu</td>
<td>that (when not visible)</td>
</tr>
<tr>
<td>knu</td>
<td>exactly that (when not visible)</td>
</tr>
</tbody>
</table>

Example (8) shows how a demonstrative adjective can modify a noun to refer to a participant. In this case, *ni'b* ‘boy’ had been introduced two clauses earlier, and is now being referred to again.

(8) Lëñi'b  k-nu sob lo tib siy.
FOC boy exactly-that sit face one chair
That boy was sitting on a chair. [SB76]

2.2.5 Pronouns

Pronouns are one of the encoding devices used for participant reference. Pronouns in SFOZ take one of two forms: ENCLITIC or INDEPENDENT. There are likely grammatical restrictions on when each form may or may not be used. Some pronouns have both forms, and others have only one form. Of most interest in studying participant reference are the third-person pronouns. The inanimate pronoun is used for inanimate objects, such as rocks or trucks. The impersonal human pronoun is used to refer to people in general, as in the use of ‘one’ in “To make tortillas, one must shape the dough into a ball.” The respect pronoun is used in limited situations to show respect for the referent or to the addressee of the speech, regardless of how much respect is to be shown to the referent. To date, no SFOZ pronoun for general third-person participants has been found. That is, if a specific person has been mentioned and the speaker doesn’t desire to show respect, there is no pronoun available to refer to that participant subsequently. Instead, nouns are often used. Table 5 lists the known enclitic and independent pronoun forms of SFOZ.

7 In normal speed discourse, the k- morpheme ‘exactly’ is sometimes not heard and so, in some cases, it is hard to distinguish between the usage of nu ‘that’ and knu ‘exactly that’.
8 Further study is required to determine whether there is a true absence of these pronouns in SFOZ or whether they simply reflect that more data needs to be analyzed to find these forms.
9 These nouns are sometimes called “generic nouns” by linguists because the words chosen are often quite generic (e.g., ‘person’ instead of ‘man’ or ‘woman’). However, if specification is needed for disambiguation, more specific nouns are available. Black (2000:37) discusses whether these nouns could actually be considered pronouns. If considered pronouns, they would fill the holes in table 5 for the independent third-person pronouns in addition to the gap for a general third-person pronoun that could be used widely.
Table 5. SFOZ pronouns

<table>
<thead>
<tr>
<th></th>
<th>Enclitics</th>
<th>Independent forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 singular</td>
<td>-n</td>
<td>na'</td>
</tr>
<tr>
<td>1 plural inclusive</td>
<td>-na, -n*</td>
<td></td>
</tr>
<tr>
<td>1 plural exclusive</td>
<td>-nu</td>
<td>nu'</td>
</tr>
<tr>
<td>2 singular</td>
<td>-l, -la'</td>
<td>lu'</td>
</tr>
<tr>
<td>2 plural</td>
<td>-ta</td>
<td>lu'ta</td>
</tr>
<tr>
<td>3 inanimate</td>
<td>-u, -wa**</td>
<td></td>
</tr>
<tr>
<td>3 impersonal human</td>
<td>-ma</td>
<td></td>
</tr>
<tr>
<td>3 respect</td>
<td>-xi</td>
<td></td>
</tr>
</tbody>
</table>

* The reader will note that there is a form of 1st person that appears to be identical between the singular and plural inclusive forms. There is at least a tonal difference between the two forms. This needs to be studied further.

** The –wa form is used when this clitic immediately follows a back vowel.

The most common use of the pronouns is to refer to a subject or object of a clause. The enclitics are not agreement markers and cannot occur with other nominals or independent pronouns that fill the same argument position in the clause. To refer to the subject or object, they are attached to the end of a verb, following any adverbial enclitics. Like other Zapotecan languages, the enclitic pronouns in SFOZ must preserve the VSO order when they are attached to a verb, since word order is the only way to identify to which argument the pronoun refers (Black 2000:35). (See section 2.2.7 for a fuller discussion of verbal structure.) Example (9) demonstrates the use of an enclitic pronoun as subject.

(9) ¿Be s-ne'dz-l di'dz…
INT FUT1-give-2S word
Will you give permission…? [SB11]

Independent pronouns can stand alone and, therefore, do not attach to another word. Example (10) shows how a first-person independent pronoun is used in the TB text.

(10) Lë' na' ku tib kwent…
FOC 1S tell one story
I’m going to tell a story …[TB1]

2.2.6 Null reference

The term NULL REFERENCE refers to the places in SFOZ clauses where there is an absence of any overt form, but a reference is implied. Most commonly in this investigation, I will refer to null references that are subjects of clauses. Though the use of null reference is one of the foci of

---

10 Enclitic pronouns may also serve other functions, such as marking possession, by attaching to other words in the clause, such as nouns and focus markers.
11 When a clause exhibits unmarked word-order (VSO), it is often unclear whether the pronoun is an enclitic or independent. The clearest examples of independent pronouns, therefore, are found in preverbal positions, as in (10).
this study and will thus be explained further in chapter 5, there are some general observations about use of null reference that should be made at this point.

Null reference is required in positive imperative constructions (i.e., Do something!). In other words, an explicit subject may not be present in a positive imperative. The lack of explicit subject is illustrated in (11). It does not make the second singular subject explicit, but rather, uses a null reference.

(11) *K-ne’-Ø-u lo-nu.*
IMP-say-Ø-3IN face-1E
(You) tell it to us! [SB95]

In contrast, negative imperative constructions (i.e., Do not do something!) do require an explicit subject. Example (12), therefore, necessarily includes an explicit pronoun for the second singular subject.

(12) *G-unn-t-l xki.*
FUT2-do-NEG-2S like.that
Do not do that!

2.2.7 Verbal morphology

All SFOZ verbs are minimally composed of a root and an aspectual prefix. Roots are not found in isolation. Following the root, various optional adverbial enclitics can attach to modify the verb. Finally, subject and object enclitics can be attached. The basic morphological structure of SFOZ verbs is illustrated in figure 1.

![Figure 1. Verbal structure](image)

The following examples show the structure of some of the verbs used in the San Bartolo text.

(13) *G-ak-d-u*
FUT2-able-NEG-3IN
It can’t be. [SB15]

---

12 Although I claim that a root cannot occur without an aspectual prefix, there are some verbs in the text (see the appendix) that have not been glossed for an aspect because I am lacking sufficient data (i.e., a full verb paradigm) to determine where the aspect marker ends and where the root begins.
Embedded clauses often affect participant reference in that they are closely linked to matrix clauses and may allow lesser encoding of participants that were already mentioned in the matrix clause. Before discussing constraints on participant reference in embedded clauses, a brief description of the verb forms used in these clauses is needed. All finite verbs in SFOZ are marked for aspect and have grammatical subjects (whether or not stated explicitly). Black (2000:37) indicates that some Zapotec languages have nonfinite verbs (though only minimally attested) in addition to finite verbs. In SFOZ, however, I am not aware of any nonfinite forms. Embedded clauses in SFOZ employ finite verbs rather than nonfinite verbs.

Complement clauses. By definition, complement clauses occur embedded as arguments within other clauses (Bickford 1998:318). In SFOZ, no extra grammatical material is used to indicate the embedding of these clauses; the entire embedded clause simply fills the appropriate argument position within the matrix clause. In (15), a complement clause (in brackets) functions as the direct object argument of the matrix clause.

(15)  
G-wiy-n  [x-kē-l  karel.]  
FUT2-see-1S PROG-make-2S run  
I see you are running (can run).

When the subjects of the two clauses are different from one another, as in the previous example, both must be referred to explicitly. Even when the subjects are the same in both clauses, explicit subject reference in both clauses is common.

In (16), minn ‘person’ is used twice as the explicit subject of each of the clauses.

(16)  
Xla'nun  minn  [nlo'x  minn  Nay  lba'.]  
PROG-want person UNR-finish person woman heaven  
The person was wanting to destroy the Virgin (but didn’t).

This repetition, rather than gapping, of the subject references in complement clauses is one of the areas of SFOZ participant reference that stands out as unusual.

Relative clauses. In contrast to complement clauses, SFOZ relative clauses often employ a gapping strategy. That is, the head of the relative clause is not represented in the relative clause itself. The head precedes the relative clause. Like complement clauses, relative clauses are not

---

13 This verb root, which is glossed as a passive, appears to be identical to the active form. However, in this case, there is no overt agent and no previous introduction of an agent that might lead to a null reference. Further study of SFOZ passives is needed.

14 Further study is needed to determine if this is always the case.

15 In this example, the unreal aspect on the verb nlo'x ‘finish’ carries the idea that the action did not come to pass.
indicated by any extra grammatical material, such as relativizers. Example (17) illustrates the lack of relativizer and shows the relative clause (in brackets) modifying muchil ‘bag’.

\[
\text{HEAD [RC ]}
\]

(17) \text{Gënd muchil yu xab-n.}  
there.be.NEG bag be clothes-1S 
The bag that my clothes are in is not there.

Headed relative clauses are often used as appositive NPs, which are functional equivalents of relative clauses. The head of the relative clause is often a generic noun that is juxtaposed with a more specific noun. The head of the relative clause in the following example is the generic word gu’n ‘thing’ and stands in apposition to the more specific word tiend ‘store’.

\[
\text{specific HEAD (generic) [RC ]}
\]

(18) \text{Sob s-tib tiend gu’n n-chob-\text{-chi}}  
sit FUT1\text{-one store thing STAT-grow-more}… 
There is another store that is bigger. [SB18]

SFOZ also utilizes headless relative clauses. Payne (1997:328) describes these as clauses which themselves refer to the noun that they modify. Several examples can be found in MM, where det could be translated as ‘where…’, with the semantic equivalence of ‘the place where…’. Thus, even though there is no explicit head, the concept of the head of the relative clause is expressed within the clause itself, as in (19).

\[
\text{[RC ]}
\]

(19) \text{…u-dzinn det zu ti gway.}  
COMP-arrive where stand one mule 
…and arrived at where a mule was. [MM9]

2.2.9 Sentence boundaries

Sentence boundaries in SFOZ are not always clear. Because clauses are often juxtaposed to one another, many can be strung together to form one long sentence. However, at times, there are indicators of sentence boundaries. For instance, sometimes listening for pauses in a native speaker’s phrasing or intonation patterns can indicate where a sentence break should be placed. One example of where this proved useful was with the combination of the two verbs sìa ‘go’ and udzinn ‘arrived’. In the following clauses, native speakers do not pause at all between these two

---

16 One case of an ambiguous relative clause that might employ a relativizer was found. It functioned similarly to Black’s (2000:141–143) description of relativizers in Quiérogolani Zapotec.

17 Though at first glance the head of this relative clause may appear to be a relative pronoun or relativizer, it is most likely not. This is assumed due to the fact that different forms are used depending on characteristics of the other noun in apposition (e.g., whether it is animate or not).

18 Aspect can be marked on some quantifiers, including the numbers one through four. For a description of the function of aspect markers on quantifiers, see Black (2000:265–268).

19 Other examples of headless relative clauses can be found in the appendix in MM (23), (53), and (69).
verbs. Therefore, they are considered part of the same sentence, and are marked with a sentence-final period.

\[(20) \text{Gadzekna lë' me'dz sia}\]
\[\text{then FOC puma go}\]
\[\text{Then the puma went (MM22)}\]

\[\text{u-dzinn det zu ti ngon.}\]
\[\text{COMP-arrive where stand one cow}
\text{and arrived where a cow was. (MM23)}\]

Dependent clauses and embedded clauses are always kept with their matrix clauses within a sentence. Connecting words between clauses tend to indicate a new sentence. With these guidelines and intuition, judgments were made about where one sentence ends and the next one begins.

This chapter has given a framework for understanding the grammar of SFOZ which is referred to in this study. Chapter 3 presents research that others have done on the topic of participant reference and tracking.

\[\text{The initial 's' in sia could potentially be analyzed as a future aspect marker. Further study is needed before a decision can be made about how it should be analyzed.}\]
Participant reference and tracking are by no means new topics of investigation. In fact, they have been discussed from various angles and perspectives by linguists, psychologists, philosophers, and others for years.

“One of the more interesting facts about human language is that we can use different forms to refer to the same thing, and the same form can be used to refer to many different things. Yet people somehow manage to understand one another” (Gundel et al. 1993:274). This correlation between form (the reference) and meaning (the referent) is important to both the speaker and the hearer of an utterance. The hearer wants to be able to understand who did what and to whom, while the speaker wants to be able to make that information known to the hearer (Dooley and Levinsohn 2001).

The forms that refer to participants in narratives are chosen in order to aid communication between the speaker and hearer. The study of participant reference and tracking aims to identify a speaker’s motivation to choose one form over another to refer to a participant. It has often been assumed that there are hard-and-fast rules that govern a speaker’s choice of referential forms. In other words, the speaker would have only one “right” choice of a form for referring to a participant in the narrative, based on a variety of factors related to the preceding context of the reference in question. This reasoning would lead one to make rules such as “After a participant has been introduced, the next reference to that participant must be a pronoun.” Researchers (Bolinger 1979, Li and Thompson 1979, Clancy 1980, DuBois 1980, Staley 1995), however, have shown that this is not necessarily the case. Clancy (1980) suggests that there are few, if any, hard-and-fast rules, but that speakers’ choices of referential devices show tendencies with varying degrees of exceptions.

Speakers, in fact, choose from a wide range of possible forms when they refer to participants throughout a narrative discourse. Their choices can be seen as signaling devices to achieve a variety of discourse objectives. How speakers maneuver the use of reference in order to meet their own discourse objectives demonstrates the rich complexity of human language in general and of “referential management” (Staley 1995:188).

Assuming, then, that the speaker does have a choice in how to refer to a participant, what factors help the speaker decide which form to use? Dooley and Levinsohn (2001) describe three basic tasks of reference that speakers must consider when choosing referential forms. The three kinds of tasks included are semantic, discourse-pragmatic, and processing. They are explained in table 6, replicated from Dooley and Levinsohn (2001:112).

---

21 This is merely a hypothetical example of the kind of thinking that could result from seeing referential choice as being determined by conditions and context only, rather than as a choice.
Table 6. Three tasks of a scheme of reference

<table>
<thead>
<tr>
<th>Semantic</th>
<th>Identify the referents unambiguously, distinguishing them from other possible ones.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse-pragmatic</td>
<td>Signal the activation status and prominence of the referents or the actions they perform.</td>
</tr>
<tr>
<td>Processing</td>
<td>Overcome disruptions in the flow of information.</td>
</tr>
</tbody>
</table>

Clancy (1980) discusses a similar list of factors in her work on referential choice in English and Japanese narratives.

Though all languages have to accomplish the previously mentioned tasks, they do not necessarily go about it the same way. While cross-linguistic studies show some universal constraints on referential choice, they also reveal language-specific factors which govern it (Clancy 1980:196). Therefore, SFOZ probably has its own factors that won’t be learned without a specific study of SFOZ.

Now that some of the general concerns of participant reference have been mentioned, I will review several theories that may help in understanding participant reference in SFOZ. I will discuss Givón’s (1983, 1994) theory of topic continuity, Ariel’s (1988, 2001) accessibility theory, Longacre’s (1995) ordered triplets device, Dooley and Levinsohn’s (2001) method for analyzing participant reference, and Anderson, Garrod, and Sanford’s (1983) focus on episode shifts.


In his volume on topic continuity, Givón (1983) discusses the correlation between topic or participant continuity and choice of referential forms. Continuity, as he uses the term, is the means by which various organizational levels of discourse are brought together for the hearer to be able to comprehend the discourse. Though Givón (1983) refers to several types of continuity, his primary focus is on topic or participant continuity. Specifically, he has studied the correlation between participant continuity and encoding devices. In other words, what devices encode continuous and accessible participants versus discontinuous and inaccessible participants?

(Participants that have been mentioned recently or are topics of entire paragraphs are highly accessible, while participants that are being newly introduced or reintroduced after a significant period “off stage” are less accessible.)

Givón notes that there are four major factors which determine the difficulty or ease a hearer will have in identifying a topic or participant. These four factors are: (1) length of absence from the register, (2) potential interference from other topics, (3) availability of semantic information, and (4) availability of thematic information. Only the first two factors are considered in the cross-language studies found in his volume on topic continuity. The third and fourth factors are less quantifiable and, thus, are not measured. He has developed quantitative measures to assess the first two major factors (Givón 1983:11).

Corresponding to the first factor is referential distance. REFERENTIAL DISTANCE refers to the number of clauses found between the present reference to a participant and the preceding reference to the same participant. The shorter the gap between references to the same participant, the easier it is to identify the participant. A short referential distance implies high accessibility of the participant. To calculate referential distance, Givón looks back to the clauses prior to the
given reference and counts each clause up to and including the clause in which the participant was most recently mentioned. In other words, in (21), the referential distance of “He” is ‘one’ because the previous reference to that participant occurs one clause earlier, encoded as “John”.

(21) John is a linguist.
    He studies the Zapotec languages of Mexico.

Details of Givón’s method of calculating and analyzing referential distance are discussed further in chapter 4.

The second of the major factors is POTENTIAL INTERFERENCE. It is a binary feature that evaluates whether or not there are participants on stage that could potentially cause ambiguity in participant identification. Participants are most easily identified when no other participants have been present in the preceding context. Givón measures this by looking back arbitrarily, at up to five clauses previous to the given reference. If one or more other referents are present that could potentially cause interference, assign a ‘two’ as the potential interference value. If there is none, assign a ‘one’. So, the value does not identify how many potential interferences are found, merely whether or not there are any (Givón 1983:11, 14).

The cross-language studies presented in Givón’s (1983) volume on topic continuity include a third measurement: persistence. PERSISTENCE looks at the following context, rather than what precedes the reference to measure the longevity of a topic. The measure indicates how many of the following clauses (occurring after the reference in the natural flow of discourse) include references to the given participant (no matter how the participant is encoded) before there is an interruption of that continuity (Givón 1983:15).

Applying the aforementioned measurements to various encoding devices of a language, patterns result that suggest certain encoding devices refer to more accessible topics, while others more commonly refer to less accessible topics. By identifying which devices refer to more accessible topics and which refer to less accessible topics, Givón has identified a complex continuum. Encoding devices found to be common in a cross-linguistic study are ordered along the continuum, according to whether they refer to more or less accessible topics (Givón 1983:17). His continuum is replicated in (22).

---

22 In his more recent volume, Givón (1994) only refers to referential distance and persistence, not potential interference.
Following the cross-linguistic studies, Givón realized, however, that breaking the continuum into various continua, which each reflect some specific syntactic means of encoding, would facilitate more relevant predictions about topic continuity. In other words, one scale might refer to the phonological size of referential devices while another scale might refer to word-order shifts as correlates of topic accessibility. These various continua, then, alone or in various combinations, account for the various encoding devices available in a given language. The principle underlying Givón’s continuum is the principle of iconicity, stated in example (23).

\[(23) \text{“The more disruptive, surprising, discontinuous, or hard to process a topic is, the more coding material must be assigned to it” (Givón 1983:18).}\]

If true, this principle alone would imply that null reference would be used only for topics that are quite continuous. The principle, however, along with Givón’s quantitative method, do not specify the ways discourse is disrupted, such as through episode boundaries, that would cause more encoding material to be used to refer to a participant.

3.2 Ariel (1988, 2001) – Accessibility theory

In accessibility theory, Ariel (2001:31) claims that references to participants do not actually point to specific entities, but rather give instructions to the hearer on “how to retrieve the appropriate mental representation in terms of degree of mental accessibility.”

Ariel also asserts that factors similar to Givón’s (1983) affect processing of referential expressions. These factors are listed below.

\[(24) \text{Factors affecting accessibility of antecedents (Ariel 1988)} \]
\[a. \text{Distance between antecedent and anaphor} \]
\[b. \text{Number of competitors for the role of antecedent} \]
\[c. \text{Importance of topicality in antecedent assignments} \]
\[d. \text{Role of frames in identifying antecedents} \]

---

23 Y-moved NPs often involve fronting of the contrasted topic. “Normative expectations are most typically created by listing various members of a group (type, genus) who are expected to display similar behavior, or receive similar treatment” (Givón 1990:706).
Factors a. and b. roughly correspond to Givón’s (1983, 1994) ideas of referential distance and potential interference. Ariel (1988, 2001) argues that there is no direct correlation between referential form and from where the stored information must be retrieved (e.g., from the physical context versus from general encyclopedic knowledge). Rather, the same referring expression can refer to information stored anywhere by informing the hearer of the accessibility level of the participant (Ariel 2001:31). She refers to an accessibility-marking scale, which orders referential forms in a continuum from low accessibility markers (full names + modifiers) to high accessibility markers (zeros). Although Ariel presents her ideas in a slightly different framework than Givón, she is extending many of his ideas (Ariel 2001).

Accessibility theory applies to participant reference discussion and makes claims about participant reference forms. Ariel predicts that “the more informative, rigid, and unattenuated an expression is, the lower the degree of accessibility it codes, and vice versa, the less informative and rigid and the more attenuated the form is, the higher the accessibility it codes” (Ariel 2001:32). The following serve as two examples of claims that accessibility theory makes. Definite descriptions, such as “The President,” are better retrievers when the antecedent is not highly accessible (e.g., when there is a large referential distance between an anaphor and its antecedent). The theory also predicts that pronouns are better retrievers when an antecedent is highly accessible (Ariel 1988:72). These generalizations, however, make no claim to account for all referential choices. Rather, Ariel admits that accessibility theory can only account for default referential choices where the speaker has no special motive for choosing a referential form, beyond helping the hearer identify a referent (Ariel 2001:40).

3.3 Longacre (1995)

Longacre (1995) refers to effects on participant reference as an interplay of three variables: (1) referential form (noun, pronoun, null reference, etc.), (2) participant ranking, and (3) the operations related to participant reference. He notes that many linguists have overlooked these interrelated concerns. In order to prevent these factors from being overlooked in the future, he proposes using “ordered triplets” that refer to each of those areas as a way of describing participant reference data and thereby making analysis easier.

Longacre (1995:698–702) has developed standard categories under each of these three variables and labeled each with a number or letter. These standard categories can then be used to assign the appropriate triplet to any given participant reference in a text. The options available in each of the three categories are listed in (25), (26), and (27).


1. Nouns (including proper names) accompanied by qualifiers
2. Nouns (including proper names) without qualifiers
3. Surrogate nouns, especially for kinship, occupation, and role
   (a) As substitutes for (1) and (2)
   (b) As the highest level of participant reference within a given narrative
4. Pronominal elements
   (a) Pronouns proper
   (b) Deictics
5. Affixal elements
   (a) Subject affixes on verbs
(b) Object affixes (or subject-object combinations) on verbs
(c) Switch reference affixes on verbs
(d) Possessor affixes on nouns
6. Null reference

(26) Ranking of Participants (Longacre 1995:700–701)

A. Major participants are relevant, for the most part, to the entire story
   1. Central (protagonist)
   2. Non-central
      a. Antagonist
      b. Helpers of 1 or 2 above that figure prominently in the whole story
B. Minor participants, whose role is restricted and do not act throughout the story
C. Props
   1. Human
   2. Animate nonhuman
   3. Inanimate
   4. Natural forces

(27) The Operations of Participant Reference (Longacre 1995:702)

M: First MENTION within a story
I: INTEGRATION into the story as central to the whole narrative or to some embedded narrative
T: Routine TRACKING (characteristic of the storyline but not confined to it)
R: REINSTATEMENT/RESTAGING of a participant that has been offstage
C: CONFRONTATION and/or role change
L: Marking of LOCALLY CONTRASTIVE/thematic status
E: An intrusive narrator EVALUATION or comment

Identifying the appropriate triplet for any given reference involves locating the appropriate category label under each of the three lists. For example, if a participant is referred to with a qualified noun, the first number in the triplet would be a ‘one’ [see participant reference resources (25)]. If the participant is a major participant and the protagonist, the second item in the triplet would be ‘A1’ [see ranking of participants (26)]. Finally, if the reference is used when first mentioning a participant, the third triplet item would be ‘M’ [see operations of participant reference (27)]. That triplet, ‘1-A1-M’, can then be compared with all other triplets starting with a ‘one’ to see the various functions of qualified nouns used in participant reference. Similarly, it can be compared with other triplets to make further generalizations about participant reference patterns in a given language.

Another concern of Longacre’s (1995) is the function of word-order shifts in participant reference. He is not alone in recognizing word-order shifts as a feature relevant to participant reference. Givón recognizes that word-order shifts are relevant to participant reference and therefore, has included various dislocations of noun phrases (e.g., right-dislocation or left-dislocation) at different points in his scale of continuity; see (22). He considers that there is a
word-order scale which relates to continuity. His prediction is that R-dislocation would encode more continuous topics than neutral word-ordering and that L-dislocation would encode more discontinuous topics (Givón 1983:19).

In VSO languages, like Zapotec, a variety of constituents can be placed before the verb for special purposes. Some of these special purposes, such as contrast or reinstatement, seem to be related to one of the operations of participant reference mentioned by Longacre: marking of contrastive/thematic status (1995:711).

3.4 Dooley and Levinsohn (2001)

Dooley and Levinsohn propose that there are two types of strategies used for participant reference: sequential and very important participant (VIP). Languages may use either system or a combination of both. Sequential strategies share three characteristics. First, they aim to identify what is being referred to by a given reference (especially when lesser in form than a noun phrase). Second, they look at what or who was previously mentioned as a key factor in identifying the referent. Finally, they do not consider effects of discourse organization (Dooley and Levinsohn 2001:117). Sequential strategies can be considered two-dimensional in that they mainly look at participants and references to them on a time line, without considering other factors that may affect their ordering and usage.

On the other hand, VIP strategies consider that not all participants are of equal importance. Those that are more important are VIPs, either at a local level (for a given section of a discourse) or at a global level (for an entire discourse) (Dooley and Levinsohn 2001:119). Languages often use different participant reference patterns, depending on whether or not the participant is a VIP. Identifying which participants are VIPs can then help to separate default reference patterns from special cases.

One reason participant reference systems are so complex is that they combine the vast resources of any given language with one, or even both, participant reference strategies. Because of this complexity, when analyzing participant reference, it is quite helpful to identify default cases, as well as special cases. Default cases follow the rules for participant reference when there is relative continuity without any surprises. Special cases are the references that seem to break the rules because of discontinuities or other complexities (Dooley and Levinsohn 2001:124).

In their book on analyzing discourse, Dooley and Levinsohn (2001:127–134) propose the eight-step method in (28) for analyzing reference patterns:

(28) Eight-step method for analyzing reference patterns

1. Identify the inventory of encoding devices used for participant reference.
2. Prepare a chart of participant encodings within a text.
3. Track the participants.
4. Identify the contexts in which each reference occurs.
5. Propose default encoding for each context.
6. Inspect the text for other than default encoding.
7. Make any necessary modifications to the proposals of step #5.
8. Generalize the motivations for deviances from default encoding.

These steps will be discussed in more detail in chapter 4 and when they are applied to the SFOZ texts in chapters 5 and 6.
3.5 Anderson, Garrod, and Sanford (1983) – Episode shifts

Research by Anderson, Garrod, and Sanford (1983) has shown that simple linear accounts of participant reference strategies are not sufficient to account for why certain encodings are chosen over others. Their research implies that sequential strategies referred to by Dooley and Levinsohn (2001) and Givon’s (1983, 1994) measure of referential distance cannot adequately describe the full complexity of referential choice. Although quantitative data about referential distance may show some basis for referential choice, it is not enough to reveal the whole picture. Results of two research experiments have shown that changes in scene and time (indicators of episode shifts) in a discourse can also affect participant reference (Anderson et al 1983). Results corroborate the thought that “statements indicating a shift in temporal setting tend to reduce the prominence of a scenario-dependent character in relation to the main character” (Anderson et al 1983:433). The reduced prominence of these scenario-dependent characters means that mental representations of them are less available to the hearer, thus calling for use of fuller-form referents. In contrast, main characters do not seem to suffer the same reduction of prominence at episode boundaries (Anderson et al 1983).

The prediction made by Anderson, Garrod, and Sanford (1983:429) was that “if a change in episode has occurred, then a new model should be in current use, and entities mapped into a previous model should not be directly accessible by means of pronouns, in which case the pronoun could only be resolved through a time-consuming, long-term memory search.” In the experiments, research subjects often chose pronouns to encode main characters in spite of episode shifts, but in the same context, scenario-dependent characters were not encoded nearly as often with pronouns. This indicates that pronouns seem to fulfill a specific function in the discourse of signaling the maintenance of the current thematic subject, which is usually the main character (Anderson et al 1983:437). This helps to explain why pronouns are used more often for central characters than non-central ones. This phenomenon has been documented in Yalálag Zapotec by Newberg (1987). He claims that “no marked construction occurs at the beginning of new episodes to reintroduce the primary participant. A pronoun is used at each new episode” (Newberg 1987:21). These findings have implications on two important factors involved in participant reference: differentiation of various ranks of participants (referred to in Dooley and Levinsohn as VIP strategy) and effect of episode boundaries on anaphoric reference.

3.6 Previous studies of participant reference in other Zapotec languages

A few linguists have studied participant reference in other Zapotec languages, including Kreutz (1980), Newberg (1987), and Lyman (1977). I will note a few of their observations to compare with my own findings and conclusions presented in chapters 5, 6, and 7. Kreutz has studied participant reference in Guevea de Humboldt Zapotec. She found that overt pronouns mark higher prominence in paragraphs, while nouns and null pronominal allomorphs mark less
prominency (Kreutz 1980:1). This observation was noted by contrasting primary-participant (more prominent) references with secondary-participant (less prominent) references. In doing this, Kreutz found that:

The primary participant of a text is referred to by an overt pronoun whenever possible, and as other participants interact with the primary participant, they tend to be referred to nominally, and occasionally by the null pronominal allomorph in clauses which immediately follow its nominal reference (Kreutz 1980:6).

She notes, however, that this generalization is subject to restrictions imposed by grammatical and discourse rules. The preference to encode primary participants with overt pronouns rather than allowing null reference, as with secondary participants, is somewhat surprising. Givón’s (1983) continuum of accessibility would predict that a more-continuous topic (often the primary participant) be encoded by null reference rather than a less-prominent participant that is often less continuous in a given context.

Kreutz found some special functions of nominal references. First mention of participants in a given discourse are often encoded by noun phrases (Kreutz 1980:6). Nominal references are used to restage a participant, particularly a nonprimary one that is returning to the story, after not being a participant for an interval of time. When there are two participants in one clause, one is referred to with a nominal and the other with an overt pronoun or null reference to disambiguate them. The topic of a paragraph is generally introduced by a nominal, if it is a nonprimary participant. However, nominals are not generally used to introduce the primary participant, unless it is necessary for disambiguation (Kreutz 1980:9). Although a variety of nominal forms may be used to refer to a single participant throughout a story, when the participant leaves the story, the same noun that was used to introduce him/her is generally used as the final reference as well (Kreutz 1980:19). So, if a participant is introduced as ‘one man’ and later referred to as ‘person’ and ‘man’, the final reference will use ‘man’, since that was the noun used in the introduction.

Now turn to Kreutz’s interpretation of overt versus null pronoun usage. There are times when two participants are on stage and both are referred to pronominally (either overtly or by a null pronominal), rather than nominally. She claims that “when two participants are referred to pronominally in a paragraph, it is the higher-ranking one (or topic) for that paragraph which is referred to by an overt pronoun, and the nontopic participant, which is referred to by Ø” (Kreutz 1980:13). Again, this usage of null reference for a less-continuous participant does not fit with Givón’s (1983) continuum.

Null pronominals are also used to “refer back to a noun phrase in a closely related preceding clause” (Kreutz 1980:14). In fact, when the topic is introduced as a nominal subject, a null reference can be used in multiple subsequent clauses to refer back to that participant, provided that no other participants are referred to nominally or pronominally as a subject of the intervening clauses.

Participant references can be modified by quantifiers, focus constructions, and demonstrative adjectives, in order to fulfill specific functions. Sets of indefinite and definite quantifiers may modify nominal references to participants, depending on whether the information referred to is

---

25 I use Kreutz’s (1980) terminology here. She uses the term “null pronominal allomorph” to refer to what I have been calling “null reference.” At points in her paper, she generalizes about pronouns and includes both overt and null pronouns in that category.
new or old. Fronted focus constructions occur which function at a discourse level to contrast, highlight a topic, present an argument, and mark conclusion. Finally, demonstrative adjectives can also be used with nouns to reference participants that have been previously mentioned in the discourse (Kreutz 1980:1).

Another linguist, Newberg (1987), has studied participant accessibility in Yalálag Zapotec. He has applied Givón’s quantitative method to a text in order to compare the Yalálag Zapotec order of encoding devices to the order proposed by Givón. The resulting order corroborates Givón’s (1983) proposed continuum and is presented in (29).

(29) MOST CONTINUOUS/ACCESSIBLE
   clitic pronouns
   free pronouns
   generic pronoun constructions
   neutral ordered definite NPs
   fronted definite NPs
   indefinite NPs

MOST DISCONTINUOUS/INACCESSIBLE

Newberg (1987:17) has found that after being introduced, a participant is generally referred to by enclitic pronouns until there is sufficient ambiguity to warrant a more marked encoding device. In other words, after first mention, a participant is considered highly accessible and can be encoded using a device high on the continuity continuum.

In addition to applying Givón’s (1983) method, Newberg also considers the difference in reference strategies for primary versus secondary participants. He notes two differences: (1) Primary participants are encoded as more accessible topics, and allow pronoun encodings in relatively more ambiguous contexts than secondary participants. (2) Generic pronouns (Newberg’s term for pronouns that refer to a generic class, such as people or animals) are often used to refer to secondary participants, but not to primary participants. Therefore, a generic pronoun could aid hearers/readers in disambiguating the reference by leading them to search for a secondary participant rather than a primary participant as the referent. Yalálag Zapotec generic pronouns, unlike “regular pronouns,” refer predictably to secondary participants as opposed to primary participants (Newberg 1987:21).

3.7 Summary

As seen from the variety of factors mentioned above, participant reference systems are complex. Referential distance and linear views of participant reference will not be sufficient to show patterns of participant reference in specific contexts. Other factors which affect referential choice include the types of participants in a discourse and their ranking, potential interference from other participants on stage, episode breaks affecting the continuity of the discourse as a whole, and the desired discourse function to be achieved by the reference. These factors, then, are considered in addition to the linear approach in the analysis of SFOZ participant reference. Chapter 4 discusses the methodology designed for this study.

---

26 Newberg (1987) does not include zero anaphora in his scale because there were no instances of it in the text he analyzed. However, using outside data, he proposes that zero anaphora would fit on the continuum as the highest degree of participant accessibility.
4
Methodology Used

I now describe the methodology I have used in order to study SFOZ participant reference. First, I prepared for analysis by selecting three texts, dividing them by clauses, and color-coding the third-person references. Then, I inventoried the range of encoding devices used to track participants. Using the prepared texts, I applied Givón’s (1983) method to calculate referential distance, potential interference, and persistence values for each of the encoding devices. Next, I applied Dooley and Levinsohn’s (2001) method by preparing a tracking chart and labeling the contexts of each reference. Examining these charts, I proposed default encoding for each context. Finally, I examined the use of underencoded and overencoded references and other discourse characteristics to propose motivations for special cases. Each of these steps is examined in more detail below.

4.1 Selection of corpus of texts

In selecting texts for this study, I narrowed the focus to narrative texts only, as mentioned in chapter 1. Within the narrative genre, I selected three texts ranging in length from 28 to 110 clauses. The three texts are titled as shown in table 7.

Table 7. Corpus of texts

<table>
<thead>
<tr>
<th>Zapotec title</th>
<th>English title</th>
<th>Clauses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Me'dz ichop mgiy (MM)</td>
<td>The puma and the man</td>
<td>100</td>
</tr>
<tr>
<td>Tsi' bur (TB)</td>
<td>Ten donkeys</td>
<td>28</td>
</tr>
<tr>
<td>San Bartolo (SB)</td>
<td>San Bartolo</td>
<td>110</td>
</tr>
</tbody>
</table>

All three texts were narrated by native SFOZ speakers. MM was narrated by a young man in his early twenties, TB by a man in his mid thirties, and SB by a teenage girl. The texts were transcribed by Julie Nelson or me from a recording or as the narrator was telling the story. Later, the texts were glossed with the assistance of the narrators and other native SFOZ speakers.

MM is an animal folktale told from a third-person perspective. Four participants are introduced throughout the story. The central character, a puma, is looking for a creature that is stronger than he is. He has heard that man might be stronger, so he sets out to find man and challenge him. He meets a mule and a cow along the way and they direct him to a man. Then, the puma challenges the man to a screaming contest. While the puma looks for a good place to have the contest, the man returns home and gets a gun. When it is the man’s turn to scream, he makes the puma look away and instead of screaming, he shoots the puma.

TB is a humorous story about a man who takes his ten donkeys out to the field. He realizes that he might as well ride one, and so he mounts one. Then he counts his donkeys and realizes that one is missing. After spending a lot of time looking for his missing donkey, he realizes that he is sitting on it. One main reason for choosing to study this text is that it is the only text studied that shows use of the somewhat rare third-person pronoun, xî.

27 Julie Nelson is an SIL linguist who worked together with me in SFO from April 2001 through May 2002.
SB is a first-person account of a group of students who visited a town where pottery is made. The narrator relates the group’s interactions with several store owners and then explains how the pots were made. Though told from a first-person perspective, many third-person participants are involved. The study of the references in this text is limited to third-person references.

4.2 Division of texts into clauses

Before Givón’s (1983, 1994) quantitative analysis could be applied to the SFOZ texts, text division into clauses was necessary. In general, each verb, together with its arguments, was placed on a different line as a separate clause. However, certain verbs raised red flags during this process and called for careful consideration of how to divide the given clauses. For instance, speech-act verbs such as *chép* ‘say’ are often repeated within the same utterance to refer to one speech act. These could be counted as two separate clauses, but they seem to function as a unit, so I considered them a single clause. Example (30) illustrates two identical speech-act verbs that are counted as a single clause.

(30) Ch-ëp               mînn,\(^\text{28}\) ch-ëp,
    PROG-answer  person  PROG-answer
    He says, [SB55]

In contrast, when the speech-act verbs are not identical, I counted each verb as a separate clause, claiming that they are not joined as closely as the previous example. In (31), two different speech verbs are used: *unabdii’dz* ‘ask’ and *chép* ‘say’, so I counted them as two separate clauses (see clause numbers in brackets).

(31) u-nab-di’dz       Ø lo mgiy k-nu,
    COMP-request-word  face  man exactly-that
    and (he) asked that man [SB9]

          ch-ëp,
    PROG-answer
          saying, [SB10]

Another verb that could be considered a separate clause or combined with the following is the inchoative verb *uyuslo* ‘began’. I have kept this verb and the following verb together in one clause, since they work as a single unit to express one event, sharing one set of arguments. Thus, in (32), the verbs *uyuslo* ‘began’ and *xgib* ‘seek’ are counted together in one clause, rather than two separate clauses.

(32) Lë’ xi uy-uslo x-gib s-tib bur.
    FOC 3S COMP-begin PROG-seek FUT1-one donkey
    He began to look for the other donkey. [TB18]

\(^{28}\) The comma has been used as an indicator of a brief pause in natural speech.
If discourse were completely linear, the task of dividing clauses would be relatively easy. However, discourse is composed of various levels of embedding which affect how clauses are divided. The various types of embedded clauses are another area where decisions are needed.

Quoted material does not necessarily include mainline events in a narrative, but nonetheless continues to refer to participants and aids a hearer in tracking participants throughout the narrative. Because of this important role in a text, and the long length of some of the quotes, it seems advisable to list out the various clauses of direct speech rather than consider the entire speech part of the matrix clause with the speech-act verb. Therefore, the mule’s speech in MM was counted as nine separate clauses (see MM 13-21 in the appendix), rather than being considered part of clause twelve.

Complement clauses were also counted as separate clauses, rather than being combined with their matrix clauses. Though they are embedded grammatically, functioning as arguments of the matrix clause, structurally they are no different from independent clauses. Neither are they linked to the matrix clause by a complementizer. In (33), a matrix clause and complement clause in TB were counted as two separate clauses.

\[
\begin{align*}
\text{[Matrix clause ]} \\
(33) & \quad Uw-iy \quad xi \\
\text{COMP-see} & \quad 3S \\
\text{He saw [TB14]} \\
\text{[CC]} & \quad gënd \quad s-tib \quad bur. \\
\text{there.is.not FUT1-one donkey} \\
\text{the other donkey wasn't there. [TB15]}
\end{align*}
\]

Relative clauses, in contrast to complement clauses, are more dependent upon the matrix clauses in which they are embedded. The dependence is shown by the gapping strategy in headed relative clauses (the gap implies a heavier dependence upon the matrix clause to supply the missing information) and the fact that headless relative clauses use words that function as relativizers and, thus, have a different structure than independent clauses and could not stand alone. For these reasons, I counted relative clauses as a part of their matrix clauses. Example (34) from MM shows how a relative clause was kept with its matrix clause when counting clauses.

\[
\begin{align*}
\text{[Matrix clause ]} & \quad [\text{headless RC }] \\
(34) & \quad u-dzinn \quad det \quad zu \quad ti \quad gway. \\
\text{COMP-arrive where stand one mule} \\
\text{and arrived at where a mule was. [MM9]}
\end{align*}
\]

As can be seen from the above discussion, my decisions about clause-counting were based primarily on the level of embedding that seemed to be evident in the various clause types. A deeply embedded clause, such as a relative clause, was counted together with its matrix clause as a single clause, whereas a less embedded clause, such as a complement clause or direct quote, was counted separately from its matrix clause.\(^{29}\)

\(^{29}\) The level of embedding refers to tree diagrams. A relative clause which modifies a noun would be more deeply embedded than a clause which occupies an argument position, such as a direct object.
4.3 Color-coding participants

Once the clauses were divided, I color-coded each of the references to third-person participants, assigning a different color to each participant. Even if a participant was referred to with a first- or second-person pronoun in a quote, it was color-coded, provided the story generally referred to that participant in the third person. (However, when actually calculating referential distance for pronouns, only third-person pronouns were used.)

4.4 Inventory of encoding devices

Because of the small amount of data actually studied, it was hard to make generalizations about different categories of encoding devices, unless they were grouped together under general categories. For example, rather than listing NPs that are nouns + adjectives as a separate category from NPs that are nouns + relative clauses, I classified them both under the general NP category. Focusing of arguments by placing them prior to the verb and/or using the focus marker lë' commonly affects references to participants. I have classified these focused references together under each general category: focused pronouns, focused nouns, and focused NPs. I did not distinguish the way the focusing was accomplished because the majority of the time if a reference is focused it is both preverbal and takes the focus marker. Example (35) shows a focused NP that is preverbal, but does not use the focus marker lë'. Example (36) shows a focused NP that is both preverbal and uses the focus marker. Both of these were simply considered focused NPs in the inventory of encoding devices.

(35) Xïn-le\textsuperscript{30} mïn në xto-wa.
child-? person do PROG-sell-3IN
Her children do the selling. [SB40]

(36) lë' tib ndza'p sob xki
FOC one girl sit like.this
a girl sitting there like this [SB72]

A number of the null references found in these texts are grammatically constrained to be null because the verbs are imperatives (see chapter 2). Since they don’t represent a choice of reference, I left them out of the count and did not analyze them in this quantitative study.

Also, as mentioned above, pronouns counted in this inventory did not include first- and second-person pronouns used by participants during dialogue to refer to themselves. For example, in MM (75), where the man is telling the puma to go first, he uses a second-person pronoun to refer to the puma. This pronoun is not counted in the inventory of pronouns. Since these are not ambiguous references and do not represent a choice made by the speaker, they were not analyzed. If they were analyzed together with the third-person pronouns, they would skew the statistics and make it appear that pronouns are used quite commonly, when in reality third-person pronouns are used in very limited contexts.

\textsuperscript{30} Though this particle appears similar to the focus particle, it is distinct. The meaning of the particle remains unknown to me. Nonetheless, the NP is focused by its preverbal position.
4.5 Referential distance

As mentioned in chapter 3, Givón (1983, 1994) has developed a method for calculating the referential distance between participant references. For the most part, I have used the method as outlined in his volume on Topic Continuity (1983:13–14). For each encoding device, I counted how many clauses separated the given reference from the most recent preceding reference to the same participant. If the immediately preceding clause contained a reference to the same participant, the referential distance (RD) value was ‘one’. If the reference was not mentioned in the previous clause, but the one preceding that, the RD value was ‘two’. The maximum value that was assigned was ‘twenty’. Most references with an RD value of ‘twenty’ are introducing participants for the first time.

In example (37), from MM (clauses 35–40), I illustrate how a few RD values were calculated.

\[(37) \text{Konde } s\text{-tib } \text{rat } lë' \text{ me'dz}_3 \text{ uy-ann} \]
\[\text{when } FUT1\text{-one moment FOC puma COMP-see} \]
\[\text{A little later the puma saw [MM35]} \]
\[lë' \text{ mgiy}_1 \text{ z-id-le} \]
\[\text{FOC man FUT1-come-already} \]
\[\text{man already coming, [MM36]} \]
\[n-ey \quad \emptyset_1 \quad \text{gix } g-ow \quad \text{ngon}_2. \]
\[\text{STAT-carry (man) plant FUT2-eat cow} \]
\[\text{carrying plants for the cow to eat. [MM37]} \]
\[\text{Konde } u\text{-dzinn } \text{mgiy}_1 \]
\[\text{when COMP-arrive man} \]
\[\text{When the man arrived [MM38]} \]
\[u\text{-gix} \quad \emptyset_1 \quad \text{gix } lo \quad \text{ngon}_2. \]
\[\text{COMP-put.down (man) plant face cow} \]
\[\text{he put down the plants for the cow. [MM39]} \]
\[Lë' \text{ me'dz}_3 \text{ u-dzinn} \]
\[\text{FOC puma COMP-arrive} \]
\[\text{The puma arrived [MM40]} \]

To begin, calculate the RD value for the null subject reference in clause (37), which refers to the man. (All references to a given participant are marked with the same subscript number.) Looking back (in the preceding clauses), the most recent reference to the man is in clause (36), so the RD value is ‘one’. Similarly, the RD value for the nominal reference to the cow in clause (39) is calculated as ‘two’ because the most recent preceding reference to the cow is found in clause

---

This method of calculating RD is in line with Givón 1983, but does not follow his more recent way of reporting the RD values, as discussed in Givón 1994. There he proposes reporting the values as a trinary distinction (1, 2/3, or >3). I have chosen to list the actual values because it seems to represent the data more precisely.
(37), which is two clauses earlier. A final example can be drawn by calculating the RD value for the focused noun in clause (40) that refers to the puma. The last time the puma had been referred to was in clause (35). The RD value, therefore, is ‘five’.

In considering what counts as the most recent reference, references to groups of participants can be difficult to analyze. I considered any reference to a group as a reference to each of the individuals involved. So, the first-person plural inclusive pronoun in MM (48) is considered a reference to the puma as well as a reference to the man. This has implications on calculating referential distance because with the way I counted it, the RD value for the focused noun lë’mgiy ‘man’ in clause (49) is ‘one’ as opposed to ‘five’ (what it would be if I counted only references to man in the singular sense). In SB, I made a slight variation to this practice, since the story is told from a first-person perspective and the narrator frequently refers to a group, including herself, and the truck driver. When the narrator is speaking, the driver does not seem to be in focus at all, and, therefore I did not consider the narrator’s references to the group (see SB clauses 1–4) as a reference to the driver. However, when the driver is speaking and referring to the group (see SB clauses 12–13), I did consider that to be a reference to the driver, since his participation is in focus as he is speaking.

Following the above method, I then calculated the mean and the median RD value for each encoding device. The results for each text and the combined total results are presented in chapter 5.

4.6 Potential interference

Calculations of potential interference (PI) are not as straight-forward as referential distance. The general idea is to determine whether or not there are other participants in the context preceding a given encoding that could potentially interfere with the hearer’s ability to positively identify the referent at hand. Following Givón’s (1983) practice, I considered the “preceding context” to be the three clauses preceding the reference. Upon examining that context, a binary value was assigned. If there are no other participants in those three clauses, the PI value is ‘one’. If there are other participants involved in the preceding context, the PI value is ‘two’. Also, if the participant is being mentioned for the first time, a PI value of ‘two’ is assigned. This value is assigned because the hearer will likely attempt to access a previously-mentioned participant rather than anticipate a new participant, thereby causing interference.

Givón explains another consideration in measuring PI in the following quote:

This measurement was further mitigated by the factor of semantic compatibility with the predicate of the relevant clause: An interfering topic was counted only if it was just as semantically compatible (most commonly in terms of animacy, humanity, agentivity, or semantic plausibility as object or subject) with the predicate of the clause as the topic under consideration (Givón 1983:14).

How this factor applies can be illustrated by looking at example (38), which are TB clauses (3–5). To calculate the PI for tsi’ bur ‘ten donkeys’ in clause (5) the preceding clauses are examined for other participants. (Only two preceding clauses are actually examined here because the other clause is in the narrator’s introduction to the story, which is a major break from the story itself.)
One day a gentleman has ten donkeys. [TB3]

He left the field [TB4]

accompanying ten donkeys. [TB5]

There is another participant, the man (subscript one), involved in the context preceding clause (5). However, looking carefully at the semantics of the given clause, it becomes clear that the man could not fill the role of direct object in clause (5) because the man is the subject of the clause and the verb does not allow the subject and direct object to be the same. Therefore, that participant does not interfere with identifying the proper referent, ten donkeys, and the PI value is ‘one’, which indicates no potential interference.

The mean and median PI values for the encoding devices in each of the three SFOZ texts are given in chapter 5.

4.7 Persistence

The third and final calculation performed on the texts measures the persistence (P) of participants. Persistence measurements attempt to show how long a participant’s presence remains uninterrupted as a semantic argument in the clauses following the given reference. The P value equals the number of clauses to the right of a reference in which the participant remains uninterrupted as a semantic argument. If the clause to the immediate right of the reference does not involve the participant, the P value is zero (the minimum). There is no maximum persistence value. I considered the speaker and addressee of quotes as present throughout the speech, even if not directly involved as a semantic argument of the clause in the quoted material.

By looking at example (39), I consider the persistence value for mgiy ‘man’ in MM clause (74). In clauses (75–78), there are no further explicit references to the ‘man’. This would lead one to assign a P value of zero. However, it should be noted that in clause (75), the ‘man’ is speaking and is thus maintaining an uninterrupted presence. Clause (76) is a quote formula and although it doesn’t specifically give reference to the ‘man’, the ‘man’ is known to be the addressee of what will be spoken in clause (77). Therefore, throughout both of these clauses the ‘man’ is considered persistent. Finally, clause (78) does not refer to the man explicitly or implicitly and, thus, is an interruption of persistence. Therefore, rather than assigning P=0, as was proposed earlier, P=3 is assigned.

This is the verb of the clause.
(39) \[ \text{Ch-ëp} \quad mgiy_1 \]
PROG-answer man
The man says [MM74]

\[ \text{“Lu'2 g-ne'ch.”} \]
2S FUT2-go.ahead
“You will go first.” [MM75]

\[ \text{Gadzekna lë' me'dz}_2 \text{ ch-ëp,} \]
then FOC puma PROG-answer
Then the puma says, [MM76]

\[ \text{“Z-ak-u.”} \]
FUT1-be.able-3IN
“OK.” [MM77]

\[ \text{Lë' me'dz}_2 \text{ u-kë' chedz.} \]
FOC puma COMP-make scream
The puma roared. [MM78]

The calculations of persistence for each of the encoding devices in the three texts are found in chapter 5.

4.8 Tracking chart

Leaving Givón’s (1983, 1994) methodology to one side, I move to Dooley and Levinsohn’s method. After the encoding devices are inventoried (step one in their method), I combine steps two and three of Dooley and Levinsohn (2001:128–130) to prepare charts for each text, showing the participants and labeling each participant with a unique number to track references throughout the text. The participant references are divided into two categories: subjects of clauses and nonsubjects of clauses. The chart enables easy observation of the order of references to any given participant and the encoding devices used. I used word-processing tables to make the charts, thus allowing me to sort by various features. For instance, I could group all the null references together to quickly see whether they are used for the central character only, or whether they are used to encode a variety of participants.

4.9 Labeling contexts of each reference

On the same chart that was created to track the participants, a brief context was shown, including any connecting expressions between clauses that might indicate episode shifts, and a brief summary of the content of each clause. For each participant reference listed in the chart, I then categorized the given context according to the categories for subjects and nonsubjects as defined by Dooley and Levinsohn. The lists in table 8 are replicated from Dooley and Levinsohn (2001:130–131).
Table 8. Context categories

<table>
<thead>
<tr>
<th><strong>SUBJECT CONTEXTS</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>The subject is the same as in the previous clause or sentence.</td>
</tr>
<tr>
<td>S2</td>
<td>The subject was the addressee of a speech reported in the previous sentence.</td>
</tr>
</tbody>
</table>
| S3                   | The subject was involved in the previous sentence in a nonsubject role other than in a closed conversation.  
| S4                   | Other changes of subject than those covered by S2 and S3. |

<table>
<thead>
<tr>
<th><strong>NONSUBJECT CONTEXTS</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>The referent occupies the same nonsubject role as in the previous clause or sentence.</td>
</tr>
<tr>
<td>N2</td>
<td>The addressee of a reported speech was the subject (speaker) of a speech reported in the previous sentence.</td>
</tr>
<tr>
<td>N3</td>
<td>The referent was involved in the previous sentence in a different role than that covered by N2.</td>
</tr>
<tr>
<td>N4</td>
<td>Other nonsubject references than those covered by N1–N3.</td>
</tr>
</tbody>
</table>

In example (40), each context is illustrated with examples from the texts. The underlined reference in each context is the one being examined.

(40)

S1: But the puma didn’t know man. So the puma went [MM7–8]  
S2: (The puma) says to the man, “Are you man?” The man says, . . [MM41–43].  
S3: He’s looking everywhere for the donkey. The other donkey is not there. [TB19–20]  
S4: There was an echo in the mountain. The puma says…[MM79–80]

N1: He began to look for the other donkey…He’s looking everywhere for the donkey. [TB18–19]  
N2: The puma says to the man, “Scream!”…The man says (to the puma)…[MM72–74]  
N3: “…the place we (puma and man) will scream that was found. Then the puma went with the man. [MM69–70]  
N4: It fired. It hit the puma’s stomach. [MM88–89]

The first mention of a participant in a text is not relevant to this categorization scheme and is simply labeled as ‘introduction’.

A small section from the chart for MM is presented as table 9. The first column shows the sentence number. Column two shows the clause number. Column three gives any connecting material between clauses. Column four shows the encoding device used for the subject of each clause, with the number assigned to each referent in brackets (1=puma, 2=man, 3=mule). The

---

Dooley and Levinsohn (2001:97) define “closed conversation” as a conversation when “each new speaker and addressee is drawn from the speakers and addressees of previous speeches of that conversation.” If a speaker was not previously involved in the conversation, the conversation is not closed.
fifth column, ‘S Value’, shows the context of the subject of the clause according to Dooley and Levinsohn’s classification (see table 8). The sixth column shows participants in a nonsubject role, with bracketed numbers to track the references (as in column four). The seventh column, ‘N Value’, shows the nonsubject context (see table 8). The eighth column gives a literal translation of the remaining content of the clause.

Table 9. MM chart for clauses 1 through 11

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>(one time)³⁵</td>
<td>Intr</td>
<td></td>
<td></td>
<td>be</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>one puma [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>be</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>Ø [1]</td>
<td></td>
<td></td>
<td></td>
<td>want</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>if</td>
<td>Ø [1]</td>
<td>S1</td>
<td>puma [1]</td>
<td>N4</td>
<td>do test</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>one man [2]</td>
<td>S1</td>
<td>Intr</td>
<td></td>
<td>see</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ø [1]</td>
<td></td>
<td></td>
<td></td>
<td>if man be stronger than puma.</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>And now</td>
<td>puma [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>go</td>
</tr>
<tr>
<td>9a</td>
<td></td>
<td>where</td>
<td>Ø [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>arrive</td>
</tr>
<tr>
<td>9b</td>
<td></td>
<td>where</td>
<td>one mule [3]</td>
<td>Intr</td>
<td></td>
<td></td>
<td>where be mule.</td>
</tr>
</tbody>
</table>

Some sentences include multiple embedded clauses, which complicate the task of identifying the context, so the embedded clauses are shaded in the chart. When examining a reference that follows embedded clauses (shaded areas), I looked beyond the embedded clauses and compared the given clause with the preceding matrix clause, in order to determine the context. For example, the subject in clause (7) is in a same-subject context (S1), because it has the same subject as the preceding matrix clause (3).

Following Dooley and Levinsohn’s suggestion, reported speech is included in the free translation column and not analyzed separately, since it is embedded discourse (2001:128). Though this is a change from how the clauses were analyzed for Givón’s (1983, 1984) calculations, it does not skew the data. Rather, it limits the analysis of the references to a smaller subset. A further study could be done, including the reported speech as separate clauses and analyzing all participant references within them.

³⁴ Information included in the translation column is a rather literal translation of the context of each clause. In general, the subject is not repeated, except where confusion might occur without making it explicit in the translation column.

³⁵ Parentheses are used to indicate subjects of clauses that are not third-person participants.
4.10 Prediction of defaults

Using the charts prepared for each text, I looked at the most common encodings used in each context and established these as my prediction of default encodings. In other words, I made assumptions that a subject would be encoded a certain way when appearing in context S1. I assigned default predictions for each of the subject and nonsubject contexts as well as for introductions.

4.11 Identification of underencoding and overencoding

Some encodings did not fit neatly into the default encodings predicted. These encodings were identified. All of these nondefault encodings were then classified as instances of (1) OVERENCODING, where more encoding material than expected was used or (2) UNDERENCODING, where less encoding material than expected was used.

4.12 Proposal of motivations for special cases

In order to make proposals for why “special cases” are encoded as they are, it was necessary to draw together all of the information gathered in the above steps. Observations are also made based on research that pointed to other factors in participant reference (episode breaks, VIP strategies, etc.). Examining the “special cases” in depth, I then propose motivations for the encoding of these participants.
5

Results

Applying the methodology described in chapter 4 yielded the results presented here. I first present the inventory of encoding devices, followed by the results of Givón’s method (1983, 1994). I then present the results of Dooley and Levinsohn’s method (2001) and discuss the variety of factors that influence referential choice in SFOZ. Chapter 6 gives a detailed view of the application of the methodology to one text.

5.1 Inventory of encoding devices

In order to identify patterns of reference in SFOZ, it is first necessary to identify the encoding devices available in the language. Table 10 shows the quantity and percentage of each encoding device within each text, as well as the totals across all three texts.

<table>
<thead>
<tr>
<th>SB</th>
<th>MM</th>
<th>TB</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>12</td>
<td>24%</td>
<td>21</td>
</tr>
<tr>
<td>pronoun</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>noun</td>
<td>9</td>
<td>18%</td>
<td>33</td>
</tr>
<tr>
<td>NP</td>
<td>13</td>
<td>27%</td>
<td>4</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>foc noun</td>
<td>1</td>
<td>2%</td>
<td>20</td>
</tr>
<tr>
<td>foc NP</td>
<td>14</td>
<td>29%</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>100%</td>
<td>78</td>
</tr>
</tbody>
</table>

* Though there was one null reference in TB, it was not included in the count because it appears to be anomalous. It occurs in a clause where the speaker appears to have begun a sentence and then backtracked and begun again. Thus, the ellipsis of the subject is probably not significant here and can be ignored.

Nouns and null references are the most frequently used encoding devices. Pronouns (counting only third-person pronouns, not first- or second-person pronouns used in speech) only occur in one text, TB. But there, they account for over half the references, since the speaker is showing respect and the donkey owner is a global VIP. (This is explained further in sections 5.6.1 and 5.6.5.)

5.2 Referential distance

Referential distance, which measures the distance between references to the same participant, is useful in this study to predict which encoding devices are used for more continuous topics and which are used for discontinuous topics. Given Givón’s (1983) continuum presented in example (22), it is expected that the encoding devices in SFOZ would line up as shown in example (41):
The focused NPs in SFOZ correspond to Givón’s (1983, 1990) L-dislocated DEF-NPs. These NPs are typically dislocated to the left of their neutral position and are typically used to mark topics that have been out of the focus of the hearer for some time, but are then being brought back into focus. In SFOZ, this refocusing of the topic is often accomplished through left dislocation, as well as the use of the focus marker lëß.

It is not immediately obvious where focused pronouns and focused nouns would be expected to fall, according to Givón’s predictions. They might all group closely with the L-dislocated NPs, since they are all left dislocations. Or, they might each stay closer to their neutral-ordered counterparts. In other words, focused pronouns would encode slightly more discontinuous referents than neutral-ordered pronouns, focused nouns would encode slightly more discontinuous referents than neutral-ordered nouns, etc.

The referential distance calculations for each of the three texts, as well as the overall values for the three texts combined, are shown in table 11.

Table 11. Referential distance calculations

<table>
<thead>
<tr>
<th>SB N*</th>
<th>SB mean</th>
<th>SB med</th>
<th>MM N</th>
<th>MM mean</th>
<th>MM med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>12</td>
<td>24%</td>
<td>1.08</td>
<td>1</td>
<td>1.08</td>
</tr>
<tr>
<td>noun</td>
<td>9</td>
<td>18%</td>
<td>2.44</td>
<td>2</td>
<td>2.41</td>
</tr>
<tr>
<td>NP</td>
<td>13</td>
<td>27%</td>
<td>7.92</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>foc noun</td>
<td>1</td>
<td>2%</td>
<td>12</td>
<td>12</td>
<td>20</td>
</tr>
<tr>
<td>foc NP</td>
<td>14</td>
<td>29%</td>
<td>9.07</td>
<td>3.5</td>
<td>15</td>
</tr>
</tbody>
</table>

49 100% 78 100%

* N refers to the number of references involved in the calculations. In other words, there were twelve null references in the San Bartolo text.

<table>
<thead>
<tr>
<th>TB N</th>
<th>TB mean</th>
<th>TB med</th>
<th>Totals N</th>
<th>Total mean</th>
<th>Total med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>33</td>
<td>21%</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>52%</td>
<td>1.29</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>noun</td>
<td>4</td>
<td>12%</td>
<td>6.25</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>NP</td>
<td>8</td>
<td>24%</td>
<td>4.25</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>6%</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>foc noun</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>21</td>
<td>13%</td>
</tr>
<tr>
<td>foc NP</td>
<td>2</td>
<td>6%</td>
<td>14</td>
<td>14</td>
<td>16</td>
</tr>
</tbody>
</table>

33 100% 160 101%*

* Rounding percentages to the nearest whole percentage yields a total of 101%.
Looking at the mean referential distance values (given in parentheses) across all three texts, the following continuum is suggested.

(42) MOST CONTINUOUS/ACCESSIBLE

- focused pronouns (1)
  - Ø (1.06)
  - pronouns (1.29)
  - focused nouns (2.57)
  - nouns (2.7)
  - NPs (8.68)
  - focused NPs (9.69)

MOST DISCONTINUOUS/INACCESSIBLE

Although, when looking at the absolute order, the placement of focused pronouns and focused nouns seems unexpected, the numerical difference in the means is not large enough to be significant. Essentially, the continuum could be written with various encoding devices grouped together, as in (43). (The combined RD values for each group of devices are shown in parentheses.)

(43) MOST CONTINUOUS/ACCESSIBLE

- Ø, pronouns, focused pronouns (1.13)
  - nouns, focused nouns (2.65)
  - NPs, focused NPs (9.07)

MOST DISCONTINUOUS/INACCESSIBLE

The results do seem to confirm that the focused encoding devices group with their neutral-ordered counterparts on the continuity continuum, rather than all being used to encode highly discontinuous topics.

The resulting continuum in example (43) is used in determining whether marked encodings of participants are underencoded (encoded by a device listed as more continuous than the default device) or overencoded (encoded by a device listed as less continuous than the default device).

Some predictions that come from studying the RD values in these texts are shown to hold true by using Dooley and Levinsohn’s method for studying participant reference. For example, given the position on the continuity scale in example (43), null references should be found in texts when the topic is highly continuous, as in an S1 context, where the subject is the same as the subject of the previous clause. This can be seen in example (44) taken from MM, where a null reference in clause [100] is used to refer to the puma, after the puma is already the established topic in clause [99].

(44) Axta or k-nu ñ-obiche-chi me'dz
  from hour exactly-that UNR-return-more puma
  From that time on the puma didn't go back [MM99]

36 The small number of focused pronouns in the data (only two) is another reason why its placement in the continuum is not a cause for concern.
Another prediction that can be made from the continuity scale in (43) is that NPs will be used to encode highly discontinuous topics. This prediction holds true in SFOZ texts. An introduction of a participant is certainly a discontinuous topic and it is encoded by an NP. The following example shows the use of an NP to introduce the cow in the MM text.

\[(45) \quad \text{n-dzinn det zu tì ngon.}\]

COMP-arrive where stand one cow

(The puma) arrived where a cow was. [MM23]

5.3 Potential interference

Potential interference refers to whether or not there are other participants on stage that could interfere with correctly identifying the given referent. Potential interference measurements predict how encoding devices will be used when multiple participants are on stage. The values for potential interference in each of the three texts and the overall values for the three texts combined are found in table 12.

<table>
<thead>
<tr>
<th></th>
<th>SB N</th>
<th>SB mean</th>
<th>SB med</th>
<th>MM N</th>
<th>MM mean</th>
<th>MM med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>12</td>
<td>1.33</td>
<td>1</td>
<td>21</td>
<td>1.48</td>
<td>1</td>
</tr>
<tr>
<td>noun</td>
<td>9</td>
<td>1.56</td>
<td>2</td>
<td>33</td>
<td>1.82</td>
<td>2</td>
</tr>
<tr>
<td>NP</td>
<td>13</td>
<td>1.46</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>foc noun</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>1.75</td>
<td>2</td>
</tr>
<tr>
<td>foc NP</td>
<td>14</td>
<td>1.64</td>
<td>2</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TB N</th>
<th>TB mean</th>
<th>TB med</th>
<th>Total N</th>
<th>Total mean</th>
<th>Total med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>33</td>
<td>1.42</td>
<td>1</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>1.65</td>
<td>2</td>
<td>17</td>
<td>1.65</td>
<td>2</td>
</tr>
<tr>
<td>noun</td>
<td>4</td>
<td>1.75</td>
<td>2</td>
<td>46</td>
<td>1.76</td>
<td>2</td>
</tr>
<tr>
<td>NP</td>
<td>8</td>
<td>1.75</td>
<td>2</td>
<td>25</td>
<td>1.64</td>
<td>2</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>1.5</td>
<td>1.5</td>
<td>2</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>foc noun</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>21</td>
<td>1.76</td>
<td>2</td>
</tr>
<tr>
<td>foc NP</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>1.69</td>
<td>2</td>
</tr>
</tbody>
</table>

Potentially, some devices would be ruled out as possibilities for encoding participants when there is potential interference. This would be indicated, if all references of that encoding device had a PI value of one. However, this is not the case for any SFOZ encoding device in these texts, as indicated by the mean values in table 13.
Table 13. Potential interference values across all three texts

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>MEAN</th>
<th>MEDIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>33</td>
<td>1.42</td>
<td>1</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>1.50</td>
<td>1.5</td>
</tr>
<tr>
<td>NP</td>
<td>25</td>
<td>1.64</td>
<td>2</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>1.65</td>
<td>2</td>
</tr>
<tr>
<td>foc NP</td>
<td>16</td>
<td>1.69</td>
<td>2</td>
</tr>
<tr>
<td>noun</td>
<td>46</td>
<td>1.76</td>
<td>2</td>
</tr>
<tr>
<td>foc noun</td>
<td>21</td>
<td>1.76</td>
<td>2</td>
</tr>
</tbody>
</table>

Even null references (which have the lowest PI value overall) are occasionally used when there is potential interference. Though the median PI value for null references is ‘one’, the mean value is 1.42, indicating that at times null reference is used when there is potential interference. Example (46) shows the use of a null reference in MM [67] for the puma, in spite of the fact that the man is on stage and creates potential interference. The null reference in MM [67] is used in this context because the subject is remaining the same from the previous clause.

(46)  
U-dzinn Ø det zu ngon.  
COMP-arrive (man) where stand cow  
(The man) arrived to where the cow was. [MM65]

Lë' me'dz u-dzinn  
FOC puma COMP-arrive  
The puma arrived there [MM66]

ch-ëp Ø  
PROG-answer (puma)  
and (the puma) says, [MM67]

As a whole, the results of the PI calculations don’t seem to shed much light on participant tracking strategies in SFOZ. However, the lack of clear patterns, in and of itself, may be an indication that other discourse factors carry more weight than PI in deciding how participant references are encoded.

5.4 Persistence

Persistence values are the measurement of the longevity of a topic. They typically reveal differences between the encoding of more important participants and less important ones. The persistence values from each of the three texts and the values of the three texts combined are displayed in table 14.
Table 14. Persistence across all three texts

<table>
<thead>
<tr>
<th></th>
<th>SB N</th>
<th>SB mean</th>
<th>SB med</th>
<th>MM N</th>
<th>MM mean</th>
<th>MM med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>12</td>
<td>3.5</td>
<td>1.5</td>
<td>21</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>noun</td>
<td>9</td>
<td>3.78</td>
<td>2</td>
<td>33</td>
<td>6.58</td>
<td>3</td>
</tr>
<tr>
<td>NP</td>
<td>13</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>14.5</td>
<td>11.5</td>
</tr>
<tr>
<td>foc noun</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>4.95</td>
<td>3</td>
</tr>
<tr>
<td>foc NP</td>
<td>14</td>
<td>3.29</td>
<td>2</td>
<td>0</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TB N</th>
<th>TB mean</th>
<th>TB med</th>
<th>Total N</th>
<th>Total mean</th>
<th>Total med</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>7.63</td>
<td>3</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>1.35</td>
<td>1</td>
<td>17</td>
<td>1.35</td>
<td>1</td>
</tr>
<tr>
<td>noun</td>
<td>4</td>
<td>0.75</td>
<td>0.50</td>
<td>46</td>
<td>5.52</td>
<td>2.5</td>
</tr>
<tr>
<td>NP</td>
<td>8</td>
<td>0.50</td>
<td>0</td>
<td>25</td>
<td>4.04</td>
<td>1</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>0.50</td>
<td>0.50</td>
<td>2</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>foc noun</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>21</td>
<td>4.81</td>
<td>3</td>
</tr>
<tr>
<td>foc NP</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>3.13</td>
<td>2</td>
</tr>
</tbody>
</table>

The assumption about persistence values is that the more important participants will have more frequent references and, therefore, a higher probability of persisting longer (Givón 1983:15). Following this assumption and looking at the mean persistence values across all three texts (presented in table 15), it appears that null references should be used frequently for important participants, since they refer to participants with the highest persistence value.

Table 15. Persistence values across all three texts

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>MEAN</th>
<th>MEDIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>1.35</td>
<td>1</td>
</tr>
<tr>
<td>foc NP</td>
<td>16</td>
<td>3.13</td>
<td>2</td>
</tr>
<tr>
<td>NP</td>
<td>25</td>
<td>4.04</td>
<td>1</td>
</tr>
<tr>
<td>foc noun</td>
<td>21</td>
<td>4.81</td>
<td>3</td>
</tr>
<tr>
<td>noun</td>
<td>46</td>
<td>5.52</td>
<td>2.5</td>
</tr>
<tr>
<td>Ø</td>
<td>33</td>
<td>7.63</td>
<td>3</td>
</tr>
</tbody>
</table>

What should be made of the fact that there is not a single null reference found in TB? Are there no important participants in TB? That is definitely not an accurate deduction; the man is a major character, without whom there would be no plot. However, perhaps the encoding used in TB does not really counter expectations at all. Table 16 shows that in TB alone, participants referred to by pronouns have a high persistence value.\(^{37}\) Then, as would be expected, the majority of the references to the major character, the man, are pronouns.

\(^{37}\) Participants referred to by focused NPs have the highest persistence value, but there are only two such references, as opposed to 17 references by pronouns.
Table 16. Persistence values in TB

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>MEAN</th>
<th>MEDIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>8</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td>foc pronoun</td>
<td>2</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>noun</td>
<td>4</td>
<td>0.75</td>
<td>0.50</td>
</tr>
<tr>
<td>pronoun</td>
<td>17</td>
<td>1.35</td>
<td>1</td>
</tr>
<tr>
<td>foc NP</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Since the range of persistence values in each of the three texts is very different, looking at the persistence values across all three texts combined is not very helpful. It may prove most beneficial to look at each text separately, as I have already done with TB. In the SB text, all the encoding devices encode participants with very similar persistence values, ranging from 2 to 3.78. This may be due to the fact that the text is actually told from a first-person perspective, so the references to the narrator (an important participant) are not considered in this study. Therefore, due to the nature of the text itself, the persistence values in SB do not suggest any patterns of importance of participants.

Table 17 shows how references to each participant in MM are encoded. In MM, null reference is used thirteen times to refer to the puma and ten times to refer to the man. As will be discussed later, these are the two major characters in the text. (See section 5.6.4 for further discussion of participant ranking.) In contrast, the mule and cow, which are minor characters, are never referred to by null reference. This evidence, therefore, corroborates the idea that devices that encode highly persistent participants are used more often to refer to important participants.

Table 17. Percentage of each participant's references encoded by given device

<table>
<thead>
<tr>
<th></th>
<th>N=</th>
<th>puma</th>
<th>man</th>
<th>mule</th>
<th>cow</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (P=14.5)</td>
<td>35</td>
<td>3%</td>
<td>3%</td>
<td>50%</td>
<td>20%</td>
</tr>
<tr>
<td>Ø (P=10)</td>
<td>37%</td>
<td>28%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>noun (P=6.58)</td>
<td>23%</td>
<td>58%</td>
<td>50%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>foc noun (P=4.95)</td>
<td>37%</td>
<td>13%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

* Rounding percentages to whole numbers results in a total of 102%.

Looking at the persistence values of MM alone, the NP ranks highest (table 18). Yet, since the only noun phrases used in MM are those that introduce participants, noun phrases are used with equal frequency to refer to each participant (table 19). Because there are no noun phrases other than introductions in this text, it is unfair to expect the persistence value of participants encoded by NPs to be significant in determining patterns of reference in MM. The NPs do not refer more frequently to participants of higher importance, because they serve the special function of introducing participants. The high persistence values are caused because the participants, even when only minor participants, remain on the scene immediately after they are introduced. The lowest persistence value (in MM alone) is associated with participants referred to by focused nouns (see table 18). Contrary to expectations, these are only used to refer to the two major characters (see table 19).
Table 18. Persistence values in MM

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>MEAN</th>
<th>MEDIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>foc noun</td>
<td>20</td>
<td>4.95</td>
<td>3</td>
</tr>
<tr>
<td>noun</td>
<td>33</td>
<td>6.58</td>
<td>3</td>
</tr>
<tr>
<td>Ø</td>
<td>21</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>NP</td>
<td>4</td>
<td>14.5</td>
<td>11.5</td>
</tr>
</tbody>
</table>

Table 19. Percentage of encoding device references used to refer to each participant

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>puma</th>
<th>man</th>
<th>mule</th>
<th>cow</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NP (P=14.5)</td>
<td>4</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>25%</td>
<td>100%</td>
</tr>
<tr>
<td>Ø (P=10)</td>
<td>24</td>
<td>54%</td>
<td>46%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>noun (P=6.58)</td>
<td>36</td>
<td>22%</td>
<td>64%</td>
<td>3%</td>
<td>11%</td>
<td>100%</td>
</tr>
<tr>
<td>foc noun (P=4.95)</td>
<td>18</td>
<td>72%</td>
<td>28%</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Though the persistence values have, in some cases, shown patterns of reference in SFOZ, the differences among the texts and the limited number of references being studied limit how helpful this measurement is in this particular study. Other factors need to be considered to understand participant reference in these texts. The values cannot simply be put in numerical order and be expected to correspond directly with the participants, according to level of importance. Due to differences in usage of encoding devices across the three texts studied, it is difficult to generalize about the overall correspondence between persistence and importance of participants. The lack of compelling evidence for a direct correlation between persistence and the importance of participants in these particular texts, calls for a more detailed look at other discourse factors that may be affecting referential choice. Note: Although referential distance, potential interference, and persistence are important features to consider, they actually are only a few variables in the complex system of participant reference in SFOZ.

5.5 Proposed default encodings

Using the charts prepared according to the Dooley and Levinsohn method, I noted which encodings were most commonly used in each of the subject and nonsubject contexts listed. I also considered which encodings might have motivations for being marked with more or less encoding material than the default. Those encodings that had very few plausible reasons for not being the default were then chosen as the default over those that could reasonably be considered marked occurrences.

Table 20 shows the defaults I propose for each subject context.
Table 20. Default encoding for subject contexts

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>DEFAULT ENCODING</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Ø</td>
</tr>
<tr>
<td>S2</td>
<td>noun</td>
</tr>
<tr>
<td>S3</td>
<td>NP</td>
</tr>
<tr>
<td>S4</td>
<td>NP</td>
</tr>
<tr>
<td>Intro</td>
<td>NP (quant + noun)*</td>
</tr>
</tbody>
</table>

* The typical quantifier used when the participant is singular is *tib* ‘one’ or its allomorph *ti*. If the participant is not singular, then the appropriate quantifying term is used.

Even though these default encodings are used widely, one reason to shift from default encoding to overencoding is to disambiguate referents, when the need arises. This usage of overencoded references is seen in TB, where references need to be overencoded to disambiguate participants. Unlike MM, where nouns are enough to disambiguate between the different animals (because they are all different kinds of animals), in TB, all ten animals are donkeys, so referring to any one of the donkeys as *bur* ‘donkey’ does not identify the specific referent. In the following example, if the default encoding had been used, the noun *bur* ‘donkey’ would replace the noun phrase *stib bur* ‘the other donkey’. However, if the noun was used, the reference would be ambiguous as to whether the group of donkeys, or just the one, was missing as in example (47).

(47) **gënd**  s-tib  **bur**.
    there.is.not  FUT1-one  donkey
    the other donkey wasn't there. [TB15]

Therefore, to avoid this ambiguity, references in TB are encoded with noun phrases, as necessary, to disambiguate between the donkey that is “lost” and the nine donkeys that remain in the group.

When default encodings are used, the same-subject context (S1) appears to be the most clear cut. Looking at the totals of encoding devices used in all three texts, null references occur more frequently than any other encoding device when the encoded subject remains the same as the subject of the previous clause and even the previous sentence. Example (48) illustrates the usage of null reference in MM [9–10] where the subject remains the same as MM [8].

(48) **Ga-na**  sia  **me’dz**
    and-now  go  puma
    So the puma went [MM8]

**u-dzinn**  Ø₁  **det zu ti gway**.
    COMP-arrive (puma) where stand one mule
    and (the puma) arrived at where a mule was. [MM9]

**Ga-na**  **ch-ëp**  Ø₁,
    and-now  PROG-answer (puma)
    He says, [MM10]
Null reference may be found in the same-subject context (S1) even when other participants are on stage. In other words, it is not restricted to the same-subject contexts where there is no potential interference.

There are no instances of null reference in the TB text, in spite of the fact that there are many same-subject contexts. All of the same-subject contexts in TB are encoded with a pronoun. (In one of those instances the pronoun is focused.) This absence of nulls and the presence of pronouns in TB suggests that pronouns might be used instead of null references in a given text. Since observations outside of this study have suggested that third-person pronouns are used when the speaker wants to show respect to the addressee, I deduce that pronouns are likely used instead of null references in same-subject contexts when respect is to be shown. This possibility is discussed further in section 5.6.1.

Null reference is used in the same-subject context in Guevea de Humboldt Zapotec, as well (Kreutz 1980). This is in spite of the fact that Guevea de Humboldt Zapotec uses pronouns more widely than SFOZ. That is, pronoun usage is not restricted to when the speaker desires to show respect, as it may be in SFOZ.

When the subject of a clause was, in the previous clause, the addressee of a speech (S2), it is generally encoded with a noun. In example (49) from MM [41–43], the man is the addressee of the speech reported in [42], but becomes the subject of clause [43]. Therefore, in the previous-addressee context (S2) in MM [43], the man is encoded with the noun _mgiy_ ‘man’.

(49)  
\[
\text{å-ëp} \quad \emptyset \quad \text{lo} \quad \text{mgiy}, \\
\text{PROG-answer} \quad \text{(puma) face man} \\
\text{and (the puma) says to man, [MM41]} \\
\]

\[ "\text{¿Be lu'} \ n-ak \quad \text{mgiy}?" \]

\[
\text{INT} \quad \text{2S STAT-be man} \\
\text{"Are you man?" [MM42]} \\
\]

\[
\text{Ch-ëp} \quad \text{mgiy} \\
\text{PROG-answer man} \\
\text{The man says,…[MM43]} \\
\]

In addition to nouns, there are a good number of focused nouns used in MM. Other encoding devices used at least once in this context are noun phrases and focused noun phrases.

The previous-nonsubject context (S3) is when the subject is involved in the previous clause in a nonsubject role other than in a closed conversation (Dooley and Levinsohn 2001:130). Throughout the three texts, this context occurs six times. In the six occurrences of this context, the subject is encoded with a pronoun once, a noun once, and a noun phrase three times.39

---

38 Dooley and Levinsohn (2001:97) define “closed conversation” as a conversation when “each new speaker and addressee is drawn from the speakers and addressees of previous speeches of that conversation.” If a speaker was not previously involved in the conversation, the conversation is not closed.

39 Two of the noun phrases are focused, but are classified together with the other noun phrase. Focusing may be used in the previous-nonsubject context (S3) to promote the participant to the subject role, and should be investigated further.
Therefore, I propose that a noun phrase is the default encoding device in this context. I discuss the marked encodings in section 5.7.

An example of the default encoding used in the previous-nonssubject context (S3) is the reference to the missing donkey from TB. As seen in example (50), the donkey referenced in clause [20] is encoded with a noun phrase after it has been the direct object in the previous clause. The noun bur ‘donkey’ is modified by the quantifier stib ‘FUT1-one’ to distinguish it from the other nine donkeys referred to in the text.

(50)  
\[ Lad \ lad \ \chi\text{-}gib \ \chi \ \text{bur.} \]
side side PROG-seek 3S donkey

He's looking everywhere for the donkey. [TB19]

(50)
\[
\text{[NP } \text{bur} \text{]} \\
Gënd \ s\text{-}tib \ \text{bur.} \\
\text{there.is.not} \text{FUT1-one donkey} \\
The\text{other donkey is not there. } [TB20]\]

The previously-uninvolved context (S4) includes any changes of subject not included in S2 or S3 (Dooley and Levinsohn 2001:130). Noun phrases (including focused and nonfocused) account for more than any other single encoding device used for references in this context. Therefore, I propose that noun phrases are the default value for this context. Example (51) from SB [82–83] shows the use of the noun phrase ni\text{\textquoteleft}b knu ‘that boy’ to encode a participant that was uninvolved in the previous clause (S4 context).

(51)  
\[ Lë\text{'} \ lad \ \lo \ \xo\text{\textquoteleft}b\text{-}ga \ \text{tib tabl} \ \text{redond} \ \text{ga.} \]
FOC side face be.put-the one piece.of.wood round there

In the front there was another round board. [SB82]

(51)
\[
\text{[ NP ]} \\
U\text{-}ya \ \text{ni\text{\textquoteleft}b} \ \text{k-nu} \\
\text{COMP-go boy exactly-that} \\
That\text{boy went…} [SB83]\]

The final subject context I consider is that of first introductions of participants. I propose the default encoding device for these introductions is a noun phrase, specifically a noun phrase made up of a quantifier (generally ‘one’ unless the participant is plural like ‘ten donkeys’) and the noun. The puma is introduced in MM [2] using this default, as shown in example (52).

(52)  
\[ \ldots \text{u\text{-}yu} \ \text{ti} \ \text{me\textquoteleft}dz \]
COMP-be one puma

…there was a puma [MM2]

Some participants are introduced as subjects of embedded clauses, rather than matrix clauses. Example (53) shows how the mule is introduced as the subject of an embedded headless relative clause.
Participants in nonsubject roles are not encoded in the same way that they would be if they were subjects. Table 21 shows the proposed default encoding for each nonsubject context.

Table 21. Default encoding for nonsubject contexts

<table>
<thead>
<tr>
<th>The nonsubject was previously . . .</th>
<th>CONTEXT</th>
<th>DEFAULT ENCODING</th>
</tr>
</thead>
<tbody>
<tr>
<td>in same role</td>
<td>N1</td>
<td>noun</td>
</tr>
<tr>
<td>speaker, but now addressee</td>
<td>N2</td>
<td>—</td>
</tr>
<tr>
<td>in different role</td>
<td>N3</td>
<td>noun</td>
</tr>
<tr>
<td>uninvolved</td>
<td>N4</td>
<td>noun</td>
</tr>
</tbody>
</table>

When the referent fills the same nonsubject role as in the previous clause (N1), a noun is used to encode it. The one clear example of this context is found in TB and is encoded by a noun. Example (54) from TB shows the usage of the noun bur ‘donkey’ in clause [19] to refer to the direct object. The donkey was also the direct object of the previous clause, so this is the same-role context (N1).

(54) Lëf [N ] xi uy-uslo x-gib s-tib bur.
FOC 3S COMP-begin PROG-seek FUT1-one donkey
He began to look for the other donkey. [TB18]

Lad [N ] lad xgib xi bur.
side side PROG-seek 3S donkey
He's looking everywhere for the donkey. [TB19]

The previous-speaker context (N2) is seemingly unimportant in SFOZ. That is, when the addressee of a speech in the current clause was the speaker of a speech in the previous clause, there is no need to encode the participant at all. In this context, the speaker is mentioned, but there is no mention of the addressee. Thus, when the addressee is not specified, it can be assumed that the addressee was the speaker of the previous clause. (See examples in MM clauses (12), (26), and (43) in the appendix.)

In role-change contexts (N3), the referent is involved in the previous clause in a role other than that described in N2. In these contexts a noun is again proposed as the default encoding. Nouns account for the majority of the role-change contexts (N3) in the texts. The noun mgiy ‘man’ is used in MM [72] after the man was one of the subjects of the previous clause. This is shown in the following example.

---

40 Some nonsubjects were charted with a dash (—) to indicate the participant was the addressee of a speech, but these are not considered as references to be analyzed. Unless a nonsubject is called to be obligatorily expressed (e.g., by way of verb subcategorization or other grammatical restraints) I did not mark the reference as a null reference.
When they arrived there [MM71]

\[
\begin{align*}
\text{FOC puma PROG-answer face man} \\
\text{the puma says to the man,} & \ldots [MM72]
\end{align*}
\]

In previously-uninvolved (N4) contexts, participants that were in no way involved in the previous clause are now filling a nonsubject role. (Introductions of participants are excluded from this context and have been considered a separate context.) Given the frequent use of nouns to encode participants in other nonsubject roles, it is no surprise to see that this context is by default also encoded by a noun. Though an equal number of occurrences were encoded by a noun phrase as by a noun, many of the noun phrases were necessary for disambiguation, and thus, are not a good indication of what the default is. Example (56) shows the default encoding of the puma in MM [89] with the noun \textit{me’dz} ‘puma’.

\[
\begin{align*}
\text{U-gitsyë-u.} \\
\text{COMP-explode-3IN} \\
\text{It (the gun) fired. [MM88]}
\end{align*}
\]

\[
\begin{align*}
\text{U-ga’-u} & \quad \text{ldo’} \quad \text{me’dz}. \\
\text{COMP-hit-3IN stomach puma} \\
\text{It hit the puma’s stomach. [MM89]}
\end{align*}
\]

There is only one example of a participant being introduced in a nonsubject context, the old person, referred to in example (57). As with introductions in subject contexts, a noun phrase was used to encode the referent. However, the noun phrase fills a genitive role and does not include a quantifier. I do not attempt to propose a default for nonsubject introductions, since there is only one example in these texts.

\[
\begin{align*}
\text{Kë’ foto chën minn gol.} \\
\text{hang photo belonging.to person old} \\
\text{There were pictures of the old person. [SB36]}
\end{align*}
\]

These proposals for default encodings explain how encoding devices are often chosen on the basis of the contexts in which they are found. However, many of the references do not reflect the default encoding device, but rather, use a different device.

In the following sections, I examine other factors that influence referential choice.
5.6 Other factors

5.6.1 Pronoun restriction

One factor that may strongly affect referential choice is the restricted use of pronouns. This differs from the other studies of Zapotec mentioned previously (Kreutz 1980 and Newberg 1987) where third-person pronouns are used widely throughout texts. Of the three SFOZ texts studied, two do not use pronouns at all, while the third uses many pronominal references. There is no striking difference between the texts that would lead to an obvious reason for the difference in pronoun usage. So, I draw on my knowledge of the language as a whole. From informal conversations and observation, it seems that the pronoun is used when respect is being shown by the speaker. One possibility is that respect is shown for the referent when the pronoun is used. Another possibility is that respect is shown to the addressee of the speech when pronouns are used to refer to other people. This remains an area for further study. However, either of these possibilities implies that pronoun usage is more a pragmatic issue than anything else. This is one of the areas where it becomes very clear that participant reference patterns involve an interplay between grammatical and pragmatic considerations.

Looking at the TB text, it is clear that pronouns can be used in a great number of contexts including at least S1, S3, S4, and N4. Perhaps the only limitation on the use of pronouns when the speaker wants to show respect is that the participant be identifiable. In other words, if the participant has been uninvolved previously, but remains accessible in the hearer’s mind because it is a very important participant (VIP), a pronoun may be used. Therefore, it is likely that the speaker still has to consider such matters as referential distance, potential interference, episode boundaries, and importance of the participant before using a pronoun to insure that the addressee will understand the reference. The use of a pronoun would require that the referent has at least been introduced previously and remains in the mind of the hearer (Gundel et al. 1993:278).

The wide use of pronouns in the same-subject context would fit the expectation of using an encoding device that marks highly continuous topics. This leads me to propose the division of the S1 context into two subcategories: (a) the subject is the same as in the previous clause and the speaker does not necessarily want to show respect and (b) the subject is the same as in the previous clause and the speaker desires to show respect. With that division, the proposed default encoding device for S1a would be null reference, while S1b would be a pronoun.

5.6.2 Introductions

Introductions often call for specific referential forms, regardless of the context in which the reference is made. In SFOZ, a newly introduced participant is encoded by a quantifier + noun. The use of the quantifier is important, since this is the main way of indicating singular versus plural subjects. Without the quantifier in the introduction, the number of participants (other than proper names) would be left ambiguous. Because introductions seem to be a more specific context than merely S4 or N4, I included introductions as a separate context in the list of contexts already examined in section 5.5.

---

I intentionally leave vague who is being shown respect, as that remains to be investigated.
5.6.3 Episode boundaries

In chapter 3, episode boundaries were mentioned as a factor shown to affect referential choice, according to research done by Anderson, Garrod and Sanford (1983). Payne (1988:346) found that when a boundary between discourse units occurs, references following the boundary tend to be more fully specified. This happens, in spite of the fact that the participant may already be highly accessible in the hearer’s mind.

Shifts in location or time can constitute boundaries. In SFOZ, these shifts are sometimes indicated by transition words that have been thought of as paragraph markers. The precise meaning of words such as gadzekna ‘then’ and uye'lolola ‘that being finished’ seem to be less important than the idea that there is a shift of some sort. Following Staley’s practice (1995:118), I took note of stages of journeys. These changes in location due to a participant’s travel are frequently indicators of boundaries. That is, verbs like ucha ‘left’ (SB3), sta ‘went’ (MM57), and udzinn ‘arrived’ (MM9) often indicate episode boundaries due to the change in location implied.

Dooley and Levinsohn (2001:43) claim that more encoding is needed to refer to a participant after the flow of information has been disrupted. In some languages, episode boundaries are a disruption to the information flow and, therefore, may lead to overencoding participant references that immediately follow them. This is the case in SFOZ, where overencoding seems to be another indicator of episode boundaries, as it occurs regularly following boundaries.

A concept related to episode boundaries is the idea of a peak episode within a narrative. Longacre considers the concept of plot as the notional structure of a narrative. There are often surface structures that correlate to various notional structures of a narrative. For instance, sometimes there is an episode of a narrative that is characterized by a variety of unexpected surface-level discourse features and correlates to the climax or denouement of the plot. This is what Longacre calls the “peak” episode (1996:37). Because peak is typically marked by unexpected surface structures, unexpected participant reference patterns may also mark the peak.

Table 22 shows the boundaries that are found in the MM text. (The full text can be found in the appendix.) The factors used in determining the boundaries were: use of transitional words, change in location (including parts of a journey), change in time and overencoding of a reference. Clauses [70–89] are the peak episode of the text. Some of these boundaries will be discussed in chapter 6 as factors influencing referential choice.

Table 22. Episode boundaries in MM text

<table>
<thead>
<tr>
<th>CLAUSES</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–7</td>
<td>Setting</td>
</tr>
<tr>
<td>8–21</td>
<td>Puma and the mule</td>
</tr>
<tr>
<td>22–34</td>
<td>Puma and the cow</td>
</tr>
<tr>
<td>35–44</td>
<td>Puma meets the man</td>
</tr>
<tr>
<td>45–56</td>
<td>Puma and man plan contest</td>
</tr>
<tr>
<td>57–69</td>
<td>Preparations for contest</td>
</tr>
<tr>
<td>70–89</td>
<td>The competition</td>
</tr>
<tr>
<td>90–98</td>
<td>Puma and man discuss whether they’ll compete again</td>
</tr>
<tr>
<td>99–100</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>
5.6.4 Participant ranking

Participants are not all created equal—some are more important in a text than others. Staley (1995:153) measured the importance of participants by the absolute frequency of references to them, claiming that the more references there are to a participant, the more important the participant is. A more qualitative approach was chosen, categorizing participants according to Longacre’s (1995:700–701) categories, previously outlined in example (26). I will refer only to his classification of participants, not props, since they are beyond the scope of this study. The main ranks of participants that Longacre proposes are major versus minor participants. He then further subdivides major participants as central or noncentral participants. Finally, noncentral participants are divided as antagonists or helpers.

As Dooley and Levinsohn (2001:119) suggest, the participants of each text can be classified into these categories, in order to illuminate patterns of overencoding or underencoding for a specific rank of participants. Table 23 shows the general ranking of each participant in MM.

Table 23. Participant ranking in MM

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>puma</td>
<td>Major participant</td>
</tr>
<tr>
<td>man</td>
<td>Major participant</td>
</tr>
<tr>
<td>mule</td>
<td>Minor participant</td>
</tr>
<tr>
<td>cow</td>
<td>Minor participant</td>
</tr>
</tbody>
</table>

As mentioned earlier, persistence values are related to importance of participants. That is, the encoding devices that have the highest persistence values are likely to be used to encode the more important participants. In MM, if we disregard the noun phrases used to introduce participants (which have the highest persistence value of all encoding devices in MM) null reference has the highest persistence value in MM. The correlation between persistence and participant ranking can be seen in MM. The puma and the man (both major participants) are encoded by null reference thirteen and ten times, respectively; whereas in contrast, the mule and the cow (both minor participants) are never referred to by null reference. This is different from Kreutz’s (1980) findings in Guevea de Humboldt Zapotec, where major characters are more likely encoded with overt pronouns while minor characters are usually encoded with null reference or nouns.

5.6.5 Very Important Participants (VIPs)

VIPs are participants that are introduced in a special way and then referred to with a special set of terms that do not necessarily follow those predicted as default encodings (Dooley and Levinsohn 2001:121). So, if many of the marked encodings happen to refer to one participant, it may be possible to identify that participant as a VIP and create separate encoding rules for him or her. This proves to be helpful in examining the references to the man in TB. He is first introduced with a focused noun phrase. This focusing is unusual for participant introductions. Furthermore, the contexts referred to by Dooley and Levinsohn (2001:130–131) seem to be irrelevant in encoding references to the man; a pronoun is used in all contexts. It is helpful to consider the man a global VIP and assume that the default encoding will not be used for referring to him.
In addition to global VIPs, local VIPs may also play a role in SFOZ texts, though the examples are more ambiguous. It is possible that the driver in SB functions as a local VIP when he interacts with the various store owners. It seems that each time he is restaged, he is referred to with a focused noun phrase that includes a relative clause specifying his role among the group (as leader and driver) (see SB 20, 53, and 93.). This seems to be a reference pattern used specifically for him. However, once he is restaged, references to him follow the general patterns of reference for nonVIP participants.

5.6.6 Grammar constraints

The only grammar constraints on participant reference that have been mentioned so far are the restrictions with imperative verbs. That is, a null reference is required to refer to the subject of a positive imperative. In contrast, an explicit subject is required if the imperative is negative (see section 2.2.6 for examples).

There is another grammatical issue that potentially affects participant reference. SFOZ has inalienably possessed nouns that include body parts. When a participant of a text is the possessor of an inalienably possessed noun (the boy’s foot), the possessor needs to be explicitly expressed. Any instances of participants as possessors are nonsubject contexts. Therefore, null references have already been eliminated as the default encoding for them. However, grammatical restrictions prohibit using null reference to underencode those references.

5.7 Marked encodings (underencoding and overencoding)

Keeping in mind the default encodings proposed in section 5.5, I identified which references were encoded with nondefault devices. These references are considered marked, since they are encoded in a way that is not anticipated. The marked encodings can be either underencoded or overencoded. The continuum presented in example (43) was used as the basis of determining which encodings were overencoded and which were underencoded. For example, if a noun was expected, but an encoding device for a more continuous topic was used, the reference was considered underencoded. However, if a null reference was expected and any of the less continuous devices was used, the reference was considered overencoded. Because I grouped the focused devices together with their unfocused counterparts in (43), I did not consider a difference in focusing to be a matter of underencoding or overencoding. For example, if the default was a noun and a focused noun was used I did not consider it a marked encoding. (Ideally, focusing should be studied in depth in relation to its role in participant reference, as well as other discourse functions.)

Of all the marked references, there are more subjects (thirty-four) than nonsubjects (ten) in the three texts. However, as table 24 indicates, the percentage of all subjects that are marked (30 percent) and the percentage of all nonsubjects that are marked (40 percent) are relatively close. Thus, nonsubject references and subject references are marked with relatively equal frequency. Subjects and nonsubjects have similar distributions of overencoded versus underencoded references. In both categories, overencoding is slightly more prevalent, as seen in table 24.

42 There is one instance in SB (88) where a possessor is not explicitly stated for нима ‘hand’. This is evidence that in this case the body part may actually form part of a compound verb ‘wash hands’ and, therefore, not require an explicit possessor. Further study is needed.
Table 24. Distribution of marked references

<table>
<thead>
<tr>
<th>Subject</th>
<th>N=114 (100%)</th>
<th>N=25 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overencoded</td>
<td>18 (16%)</td>
<td>6 (24%)</td>
</tr>
<tr>
<td>Underencoded</td>
<td>16 (14%)</td>
<td>4 (16%)</td>
</tr>
<tr>
<td>Total</td>
<td>34 (30%)</td>
<td>10 (40%)</td>
</tr>
</tbody>
</table>

By listing each of the marked references and examining possible causes for using nondefault encoding devices, some further generalizations were made. Factors that affect participant reference in SFOZ and result in marked reference forms are episode boundaries, participant rank, changes at peak, and the need to disambiguate.

Overencoded references commonly mark episode boundaries. In MM, sixty-two percent of the references following episode boundaries are overencoded. This overencoding occurs in clauses 8, 35, 57, 70, and 99. In example (58), the overencoded reference to the puma and the indication of a time lapse mark the episode boundary. The puma is referred to with a focused noun, in spite of the fact that it was the subject of the previous clause (S1 context), which would usually call for a null reference.

(58) Konde s-tib rat lë' me'dz uy-ann
     when FUT1-one moment FOC puma COMP-see
     A little later the puma saw…[MM35]

In MM, the majority of the overencoding after boundaries are expressed by focused nouns. However, sometimes nonfocused nouns are also used. Overencoding following boundaries also occurs in the other texts [see SB (20) and TB (18) in the appendix, as examples].

Overencoding is especially prevalent in the peak of MM examples (72–89). Of the ten subjects in the peak, only four are encoded with the default. All six that are overencoded, or differ from the default by focusing, are focused nouns. Two that are encoded with the default devices are also focused nouns. In the peak, this use of focused nouns, which are often used after breaks in continuity, seems to slow down the action and emphasize the turn-taking of the dialogue and action.

Underencoding in the SFOZ texts tends to be more related to participant ranking and VIP status than to boundaries and peak. In MM, five of the nine underencoded subjects are major participants that are encoded with focused nouns, rather than noun phrases. All five of these instances of underencoding occur in contexts where the participant has not been mentioned at all in the previous clause (S4). Presumably, since these participants are major characters in the story, they remain more accessible in the hearer’s mind than minor characters, and thus require less encoding, in spite of larger referential distances (the distinguishing factor between the S4 context and the others). This pattern appears to be predictable enough that it warrants restating the default encodings to include mention of encoding major participants in S4 contexts with focused nouns rather than noun phrases.

Identifying a global VIP may help explain referential choice when reference strategies appear to single out a major participant and encode references to that participant in a different way.

---

43 As mentioned earlier, I did not consider focusing alone as overencoding. However, in this case, it seemed appropriate to call it overencoding, since there was a sudden abundance of focused nouns in the peak.
“After being introduced, the global VIP is often referred to by minimum, but virtually constant, coding” (Dooley and Levinsohn 2001:121). This holds true in TB, where the use of pronouns to underencode references to the man is a clue that the man is a global VIP. In TB, pronouns are used in S1, S3, S4, and N4 contexts (there are no references to the man in any other context in the TB text). The use of such a minimal encoding device is certainly unusual for the S3, S4, and N4 contexts, so the global VIP motivation makes sense.

5.8 Proposed motivations for nondefault encodings

In the previous section, various motivations for marked references were discussed. Some of these motivations allow the default encodings that are proposed in section 5.4 to be revised to account for more of the data. Table 25 shows the revised default encodings.

Table 25. Revised default encodings

<table>
<thead>
<tr>
<th>Referent was previously…</th>
<th>CONTEXT</th>
<th>DEFAULT ENCODING</th>
</tr>
</thead>
<tbody>
<tr>
<td>same subject</td>
<td>S1</td>
<td>Ø</td>
</tr>
<tr>
<td>addresser</td>
<td>S1 (speaker desires to show respect)</td>
<td>pronoun noun</td>
</tr>
<tr>
<td>nonsubject</td>
<td>S2</td>
<td>NP</td>
</tr>
<tr>
<td>uninvolved</td>
<td>S3</td>
<td>NP focused noun</td>
</tr>
<tr>
<td>never mentioned</td>
<td>S4 (for minor participants)</td>
<td>NP (quant + noun)</td>
</tr>
<tr>
<td>in same nonsubject role</td>
<td>Intro</td>
<td>noun</td>
</tr>
<tr>
<td>speaker, but now addresser</td>
<td>N1</td>
<td>noun</td>
</tr>
<tr>
<td>in different role</td>
<td>N2</td>
<td>—</td>
</tr>
<tr>
<td>uninvolved</td>
<td>N3</td>
<td>noun</td>
</tr>
</tbody>
</table>

At times, however, hard and fast rules do not appear to account for the data; but rather, inferences, based on factors that may affect referential choice, account for nondefault encodings. One of these inferences is that a reference to a participant is often overencoded following an episode boundary. Another inference is that a major participant or global VIP that remains accessible in the hearer’s mind, even though not having been recently referred to, is often underencoded.

These results give a clearer picture of how referential choices are made in SFOZ. In chapter 6, I analyze in detail the MM text and discuss the motivations for each participant reference.
6
A Closer Look At One Text

In this chapter, I work through the text *Me’dz ichop mgiy* ‘The puma and the man’, applying the methods of Givón (1983, 1994) and Dooley and Levinsohn (2001), to determine patterns of participant reference.

First, I consider what can be learned by calculating referential distance, potential interference, and persistence of the encoding devices used in this text. The referential distance values for MM are repeated here in table 26, which shows that noun phrases have the highest referential distance (with a mean value of 20) and would, thus, be expected to encode highly discontinuous participants. This is certainly the case in MM, where all four occurrences of noun phrases are introductions of participants that have not been involved in the text previously (see clauses (2), (6), (9), and (23) in the appendix). The participants being introduced are maximally discontinuous and it fits that they are encoded with the device of the highest referential distance value.

Table 26. Referential distance values in MM

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>21</td>
<td>27%</td>
<td>1.1</td>
</tr>
<tr>
<td>focused noun</td>
<td>20</td>
<td>26%</td>
<td>2.1</td>
</tr>
<tr>
<td>noun</td>
<td>33</td>
<td>42%</td>
<td>2.41</td>
</tr>
<tr>
<td>NP</td>
<td>4</td>
<td>5%</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

At the other end of the spectrum are the null references, which would be expected to encode highly continuous participants, due to their low referential distance value of 1.1. Examining the contexts in which null references occur, it becomes clear that these references are used to encode subjects that remain the same from the previous clause or sentence (S1, according to the Dooley and Levinsohn 2001:130 classification). In this context, participants are very continuous and it fits that they are encoded with a device that has a low referential distance value, such as the null reference. The null references in example (59) demonstrate the continuity of the participant, the puma, which was referred to explicitly in clause [8], and then with null references in the next clause [9] and even across the sentence boundary in clause [10].

(59) *Ga-na sia me’dz₁*
    and-now go puma
    So the puma went [MM 8]

*u-dzinn Ø₁ det zu ti gway.*
COMP-arrive (puma) where stand one mule
and arrived at where a mule was. [MM 9]
Nouns and focused nouns fall between the two extremes in terms of referential distance values, so it is not surprising that they are sometimes used to refer to fairly continuous participants and, at other times, are used to refer to fairly discontinuous participants. The referential distance calculations, thus, do correlate with the data of the MM text.

In addition to aiding predictability of when each encoding device will be used, the referential distance calculations also determine the continuity continuum by which markedness is determined. Looking at MM alone, the continuum shown in example (60) is suggested, but it is more accurate to consider the continuum suggested by all three texts, as presented in (61).

(60)

```
MOST CONTINUOUS/ACCESSIBLE
  \[\emptyset\]
  focused noun
  noun
  NP
MOST DISCONTINUOUS/INACCESSIBLE
```

(61)

```
MOST CONTINUOUS/ACCESSIBLE
  \[\emptyset,\text{pronoun, focused pronoun}\]
  noun, focused noun
  NP, focused NP
MOST DISCONTINUOUS/INACCESSIBLE
```

The continuum in example (61), therefore, is here used as the basis for further analysis of this text. When a device lower than the default encoding device on the continuity continuum is used to refer to a participant, that reference is considered overencoded. Alternatively, when a device higher than the default encoding device on the continuity continuum is used, that reference is considered underencoded. Thus, the continuum of SFOZ encoding devices (based on referential distance values) identifies which participant references have more or less encoding than a default device. For example, the focused-noun reference to the puma in clause 24 (shown here in example 62) is considered overencoded because the default encoding for its context, as identified by Dooley and Levinsohn’s method, is a null reference. Although the participant is the subject of the previous clause and sentence, and highly continuous, a device associated with lesser continuity is used.

(62)

```
[Foc N ]
Lë' me'dz \[ch-ëp,\]
FOC puma PROG-answer
The puma says, [MM 24]
```
Besides referential distance, potential interference was also calculated. The potential interference values for the encoding devices in MM are here repeated as table 27.

### Table 27. Potential interference values in MM

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>21</td>
<td>1.48</td>
<td>1</td>
</tr>
<tr>
<td>focused noun</td>
<td>20</td>
<td>1.75</td>
<td>2</td>
</tr>
<tr>
<td>noun</td>
<td>33</td>
<td>1.82</td>
<td>2</td>
</tr>
<tr>
<td>NP</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

As might be expected when considering both referential distance values and potential interference values, noun phrases are used when there is high potential interference. In the case of all four noun phrases in MM, other participants were involved in the previous contexts that could have interfered with proper identification of the referents. However, all of the NPs in this text are introductions of participants that require noun phrase references, regardless of potential interference. Without more NP references in the text, it is hard to confirm that NPs are used frequently when there is potential interference.

Null references in MM have the lowest potential interference value. Though null references are sometimes used when there is potential interference (as indicated by a mean value ‘greater than one’), the median value is ‘one’, indicating no potential interference. So, the null reference is more frequently used when there is no potential interference. Though Dooley and Levinsohn’s (2001) method does not deal with potential interference, the same-subject context where null references are considered default would often be a context where there is no potential interference. That is, when the subject remains the same clause after clause, there is less room for other participants to be involved and less likelihood of their interfering with the proper identification of the referent.

Beyond noun phrases and null reference, the potential interference values for MM do not provide much information that is helpful in determining participant reference patterns in SFOZ. Each of the other encoding devices is used at times when there is potential interference and at times when there is not.

The final quantitative measurement proposed by Givón (1983, 1994) is persistence. Persistence values measure how long a referent continues to be an uninterrupted semantic argument in the clauses that follow a given reference. Givón (1983:15) claims that there will be more references to the more important participants and, therefore, they will have a better chance of persisting longer.

As mentioned in chapter 5, persistence values of participants do not shed much light on participant reference patterns in the MM text. One example of why persistence values in the MM text do not reveal referential patterns is seen in looking at the use of NPs. Even though participants have the highest persistence value when referred to by NPs, the limited use of NPs for introductions in this text means that the persistence values do not correspond to the importance of participants.

In the case of MM, persistence values do not help in determining participant reference patterns for SFOZ. In fact, some of the data of MM seems to contradict the expectations predicted by Givón’s (1983, 1994) theory.
Of the three quantitative measures proposed by Givón (1983, 1994), referential distance proved to be the most helpful for analyzing SFOZ references in this text.\footnote{Givón (1994:10–11) only mentions two of the three quantitative measures: referential distance and persistence.} Potential interference measures did not give an inaccurate picture of the data in this text. Though the data did not go against predictions of potential interference values, the values were not definitive enough to be much use in making generalizations about participant reference patterns. The final measure, persistence, led to inaccurate hypotheses about the SFOZ data, due to the limited use of certain encoding devices. Therefore, in this text, it was of no help in analyzing participant reference. In longer texts that have a greater variety of referential devices, these measures might prove to be more helpful. After employing Givón’s (1983, 1994) quantitative method to this text, it becomes clear that other factors also need to be considered when studying participant reference. In order to do that, look at the second method that was employed in this study.

I turn to Dooley and Levinson’s (2001) method to consider a variety of factors that influence referential choice. Their format can show how each reference was analyzed and makes the strategies of SFOZ participant reference more accessible to the reader. As suggested by Dooley and Levinsohn, I prepared a chart to show each third-person reference in the MM text. After examining each reference to determine whether it uses the default encoding, I marked overencoded references with a plus sign (+) and underencoded references with a minus sign (−) in the Subject- and Nonsubject-value columns. When the default encoding does not call for focusing and the reference uses the default encoding, but is focused in addition, a lowercase ‘f’ precedes the S- or N-value.\footnote{For example, in clause (22) the lion is encoded with a focused noun. The default encoding for the lion who was previously the addressee (S2) is a noun. So the default is used, with the addition of the focus marker. This is marked in the chart as ‘fS2’. Contrast this with the focused noun reference in clause (36), where the man was previously unmentioned (S4). Because the man is a major participant, the default encoding is a focused noun. Since the focusing is included in the default, it is not marked with a lowercase ‘f’ in the S-value column. It is labeled as ‘S4’.} I have divided the charts into the various episodes of the text, as listed in table 22. The charts are given in English. As described earlier, the lines of the chart are divided by sentences and clauses, with the exception of speeches kept with the clauses in which they are embedded. Embedded clauses are shaded so that the given clause can be compared with the preceding matrix clause, in order to determine the context.

Table 28 charts clauses (1–7) of MM, in which the setting is given. The puma and man are both introduced with the default encoding of a quantifier + noun in clauses (2) and (6), respectively. After the puma is introduced, it remains the subject for clauses (3–5) and is encoded by the default for same-subject references (S1): null reference. Thus, all the subject references in these initial clauses are encoded with the default values, except for the subject reference in clause (7). The plus sign indicates that it is overencoded. That is, looking beyond the embedded clauses (4–6), the subject of sentence (2) is the same as the subject in sentence (1). So, given the same-subject context (S1) a null reference is expected, but a noun is used. This overencoding is presumably to disambiguate the reference, as another participant (a man) has been mentioned since the previous matrix clause. The nonsubject references are each encoded with a noun, which is the default expected when the nonsubject has changed roles (N3) as it has in clauses (6) and (7).
Table 28. Setting (Clauses 1–7)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>(one time)***</td>
<td></td>
<td></td>
<td></td>
<td>be</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>Intr</td>
<td></td>
<td></td>
<td>be</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>want</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td>Ø [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>do test</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td>Ø [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>see</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td></td>
<td>if</td>
<td>Intr</td>
<td>puma [1]</td>
<td>N3</td>
<td>if man be stronger than puma.</td>
</tr>
</tbody>
</table>

* The first column shows the sentence number. Column two shows the clause number. Column three gives any connecting material between clauses. Column four shows the encoding device used for the subject of each clause, with the number assigned to each referent in brackets (1=lion, 2=man, 3=mule). The fifth column, ‘S Value’, shows the context of the subject of the clause, according to Dooley and Levinsohn’s (2001) classification. The sixth column shows participants in a nonsubject role, with bracketed numbers to track the references (as in column four). The seventh column, ‘N Value’, shows the nonsubject context. The final column gives a literal translation of the remaining content of the clause.

** Information included in the translation column is a rather literal translation of the context of each clause. In general, the subject is not repeated, except where confusion might occur without making it explicit in the translation column.

*** Parentheses are used to indicate subjects of clauses that are not third-person participants.

Table 29. Episode one (Clauses 8–21)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>8</td>
<td>And now</td>
<td>puma [1]</td>
<td>+S1</td>
<td></td>
<td></td>
<td>go</td>
</tr>
<tr>
<td></td>
<td>9a*</td>
<td>where</td>
<td>Ø [1]</td>
<td>S1</td>
<td>Intr</td>
<td></td>
<td>arrive</td>
</tr>
<tr>
<td></td>
<td>9b</td>
<td></td>
<td>one mule [3]</td>
<td></td>
<td></td>
<td></td>
<td>where be mule.</td>
</tr>
</tbody>
</table>

* For the charting, clauses with embedded relative clauses were divided with each clause listed on a separate line, resulting in clauses (9a) and (9b).

The first episode of MM (table 29) tells about the puma meeting the mule. This is considered a new episode because of the change in location. The puma is journeying to the place where the mule is. The episode opens in clause (8) with an overencoded reference to the puma. A noun is used instead of a null reference as expected, since the puma was also the subject of the previous clause. This overencoding occurs because the reference follows an episode boundary, as discussed in chapter 5, section 5.6.3. The other instances of same-subject contexts (S1) within episode one use the default null reference, since they do not follow a boundary.
The mule is introduced in this episode using the default encoding device for introductions (quantifier + noun). The nonsubject referent (the mule) in clauses (10–11) is not considered underencoded, even though the default for a role-changing nonsubject is a noun; and here, there is no explicit reference to the mule. This is because the mule is the addressee of a speech, so not obligatorily encoded (see footnote 42). The nonsubject reference to the addressee in (12–21) is not encoded explicitly, which is the default for a previous-speaker context (N2).

Table 30. Episode two (Clauses 22–34)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>22</td>
<td>Then</td>
<td>F puma [1]</td>
<td>fS2</td>
<td>Ø [1]</td>
<td>S1</td>
<td>go</td>
</tr>
<tr>
<td></td>
<td>23a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>arrive</td>
</tr>
<tr>
<td>23b</td>
<td></td>
<td></td>
<td>one cow [4]</td>
<td>Intr</td>
<td></td>
<td></td>
<td>where be cow.</td>
</tr>
<tr>
<td>9</td>
<td>34</td>
<td>Then</td>
<td>F puma [1]</td>
<td>fS2</td>
<td></td>
<td></td>
<td>hide behind plant.</td>
</tr>
</tbody>
</table>

Episode two (table 30) is identified by another change in location. The puma now goes and meets the cow. The connecting word gadzekna ‘then’ is another good indication that this is an episode change. Though this episode is not characterized by an overencoded reference at the beginning, the subject reference in the first clause (22) of the episode is focused. In this case, the focusing seems to replace the overencoding normally found after boundaries.

In clause (23b), the final participant, the cow, is introduced with default encoding. The overencoded subject reference in sentence (7) is a little more mysterious. One could consider sentence (7) as the beginning of a subsection within the episode (i.e., a lower-level boundary), since it begins the action of the episode after the puma arrives (a stage of his journey) in sentence (6). (A dotted line in table 30 indicates the location of this lower-level boundary, if it, indeed, exists.) This assumption would claim that overencoding is used, due to the boundary. However, the subject reference in sentence (4) is very much parallel to this reference and yet the reference in sentence (4) is not overencoded. The difference may be that this reference occurs further along in the story, not as near to the setting and introduction as the parallel reference in sentence (4). If this is significant, it may mean that even after lower-level boundaries, overencoding is desired in SFOZ, except when near the beginning of the text. But perhaps it would be most accurate to say simply that overencoding is an option when a reference follows a boundary.

Though clause (34) is introduced with the connector gadzekna ‘then’, there are not enough other indicators to consider it a new episode. There is no shift of time or change of location (other than a very local move from a visible location to behind some plants) to suggest that this is a new episode. The subject of clause (34) is default encoded with the addition of focusing. Again, I leave the use of focusing for future study. The remainder of the references in episode two use default encoding devices.

---

46 The subject references in sentences (4) and (7) are parallel in the sense that both references to the puma occur in quotation formulas immediately following statements of the puma leaving one place and arriving to a new location.
Table 31.  Episode three (Clauses 35–44)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>35</td>
<td>When later</td>
<td>F puma [1]</td>
<td>+S1</td>
<td></td>
<td></td>
<td>when later see</td>
</tr>
<tr>
<td></td>
<td>37b</td>
<td>cow [4]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>carry man plant</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>S1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>put down plants</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>in front of cow.</td>
</tr>
<tr>
<td>41–</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>say to man, “Are</td>
</tr>
<tr>
<td>43–</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>you man?”</td>
</tr>
</tbody>
</table>

Episode three (table 31) begins the encounter between the puma and the man. This episode is identified by a change in time explicitly stated in the connecting material between episodes: *konde stib rat* ‘when it was later’. The first reference following the episode boundary (in clause 35) is overencoded with a focused noun instead of the default null reference. The reference to the man in the following clause is then encoded with a focused noun. Because the man is a major participant, the focused noun is the default for the given context (previously-uninvolved subject/S4). If he was a nonmajor participant, this would be an underencoded reference. This is also the case for the reference to the puma in clause (40).

The next marked encoding in this episode is the nominal reference to the cow in clause (37b). This reference is encoded with a noun, rather than the default noun phrase, even though the cow was uninvolved in the previous clause and sentence. The underencoding may be allowed, since the cow is the subject of an embedded clause and, therefore, does not need to be so prominent. Being the only cow in the story, there is no ambiguity when a noun is used instead of the default noun phrase.

To identify the context for the subject of clause (38), the man, I look past the preceding embedded clauses (36–37). It then becomes clear that there is a change in subject from sentence (10), where the puma is the subject, to sentence (11), where the man is the subject. Thus, a focused noun would be expected as the default for this context (S4 for a major participant). In fact, a nonfocused noun is used to encode the referent. Though this is a slight difference from the default, it is not considered a marked reference since the difference is only a matter of focusing.

As is expected, given the defaults, most of the nonsubject references in episode three are encoded with nouns. When the man is the addressee in (41–42), he is encoded with a noun rather than being left implicit. This is necessary, since the cow was mentioned in (39) and, therefore, the puma and man are not the only participants on stage. It is therefore necessary to disambiguate who is being addressed in (41–42). The addressee in (43–44) is not encoded, which is the default when the addressee was previously the speaker, as in this case.

Episode four (table 32) is the dialogue surrounding the proposal to have a screaming contest between the puma and the man. The main signal that this is a new episode is the connector *gadzekna* ‘then’ that seems to be used mainly to mark episode breaks.
Table 32. Episode four (Clauses 45–56)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>say, “Okay, go look for where we’ll scream, I’ll leave my bucket.”</td>
</tr>
<tr>
<td></td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The only participants on stage in episode four are the puma and the man. The puma is referred to by the default encoding device for a subject that was previously an addressee: a noun. The man is also referred to by a noun; however, the noun is focused. This reference may be focused for the sake of contrasting the two participants as the action rises towards a climax. As the dialogue unfolds between the two participants, it is not necessary to encode the addressees, and therefore, they are not stated explicitly.

Episode five (table 33) describes side trips each of the major participants takes to prepare for the contest. The episode could be divided into three subsections (57–59, 60–65, and 66–69), as indicated in the table by horizontal lines. The initial references of each subsection [see (57), (60), and (66)] are encoded with focused nouns. The reference in (57) is overencoded, which can be attributed to the fact that it follows an episode boundary. Following this reference, null reference is used, since the puma is the only participant on stage and he remains the subject throughout the subsection.

Table 33. Episode five (Clauses 57–69)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>57</td>
<td>F puma [1]</td>
<td>+S1</td>
<td>S1</td>
<td>go</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>60</td>
<td>F man [2]</td>
<td>S4</td>
<td>S1</td>
<td>go home</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>63</td>
<td>Ø [2]</td>
<td>S1</td>
<td>S1</td>
<td>carry gun.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>Ø [2]</td>
<td>S1</td>
<td>S1</td>
<td>put.in much</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>Ø [2]</td>
<td>S1</td>
<td>S1</td>
<td>ammunition.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65a</td>
<td>Ø [2]</td>
<td>S1</td>
<td>S1</td>
<td>carry it</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65b</td>
<td>where (place)</td>
<td></td>
<td></td>
<td>where high is</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>66</td>
<td>Ø [2]</td>
<td>S1</td>
<td>S1</td>
<td>go</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ø [1]</td>
<td>S1</td>
<td></td>
<td></td>
<td>say, “Let’s go, the place to scream was found.”</td>
<td></td>
</tr>
</tbody>
</table>
The next subsection, in which the man goes to his home, begins with a focused noun in clause (60) since the new subject, the man, was not involved in the preceding clause or sentence (S4). The focused nouns used to refer to the man (60) and the puma (66) are the defaults because both the man and the puma are major participants. Even though in each case, they have been uninvolved in the previous clause, the fact that they are major participants means they remain more accessible to the reader and can be encoded with focused nouns rather than full noun phrases (the default for minor participants in this context). Following the reference to the man in clause (60), all the references to him in the second subsection are encoded by null reference since he remains the uninterrupted subject of that subsection. The reference to the cow (65) is underencoded by a noun, when a noun phrase is expected, since the cow is a minor participant. This underencoding is permissible since there is no ambiguity caused and the reference occurs in an embedded clause where less prominence of the subject is warranted.

In the third subsection the two major participants reunite before going off to have the screaming contest. In this subsection, default values are used for all the encodings.

As the text progresses to episode six (table 34), the plot reaches its climax. I consider this episode the peak of the text. That is, it corresponds to the climax of the action and is marked by unexpected surface features. Therefore, I expect to find some unusual characteristics of participant reference in this episode. A brief look at the chart shows that there are more marked and focused references in this episode than have been seen in the previous episodes.

---

47 The default for the previously-uninvolved (S4) context was determined by looking at all three SFOZ texts. Noun phrases were used with slightly more frequency than focused nouns in the S4 context, but their usage seemed to be determined by whether the participant was major or minor. Major participants are encoded with focused nouns, while minor participants are generally encoded with noun phrases. Though this contrast exists, it cannot be seen in this text alone where both S4 contexts that involve minor participants are encoded with nouns. Since both of the references to minor participants in S4 contexts in this text occur within embedded clauses, the default was determined by considering the instances from other texts where a minor participant was the subject of a matrix clause [see SB clauses (78, 83, and 87) in the appendix, as examples].
Table 34. Episode six (Peak) (Clauses 70–89)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>Conn</th>
<th>Subject</th>
<th>S Value</th>
<th>Nonsubject</th>
<th>N Value</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>71</td>
<td>When</td>
<td>Ø [1,2]</td>
<td>S1 S1</td>
<td>man [2] N3</td>
<td>arrive there</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>72–73</td>
<td></td>
<td>F puma [1]</td>
<td>+S1</td>
<td>man [2] N3</td>
<td>say to man,</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>78</td>
<td>That being finished</td>
<td>F puma [1] (echo)</td>
<td>+S1</td>
<td></td>
<td>scream.</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>86</td>
<td></td>
<td>F puma [1] (it)</td>
<td>fS2</td>
<td></td>
<td>say, “Don’t look at me… I’ll scream.”</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>87</td>
<td>F man [2] (it)</td>
<td></td>
<td></td>
<td></td>
<td>look one side.</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sit gun on chest.</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>hit stomach of puma.</td>
<td></td>
</tr>
</tbody>
</table>

The episode begins with the connecting material *gadzekna* ‘then’. Note that there is more frequent use of this connector within the peak than elsewhere in the text, where it seems reserved for episode breaks.

Even though the puma was the most recent subject preceding this episode, the reference to it in (70) is overencoded, due to the episode boundary.

The null reference in clause (71) refers to both the puma and the man, even though according to the chart, the puma alone was the previous subject. The man could be considered a part of the subject in clause (70) which would then lead naturally to the null reference in (71), since the subject would remain the same (the puma and the man). However, in clause (70) the reference to the puma is focused and positioned preverbally, while the reference to the man follows the verb. This dividing of the two references leads me to analyze the accompaniment of the man as an oblique rather than as part of the subject. Nonetheless, the use of the null reference is unambiguous, since both participants were clearly mentioned as leaving together in the previous clause (70).

In (72–73), a focused noun is used to single out the puma from the dual subject of the previous clause. Therefore, the overencoding of this reference may be a matter of disambiguation. It is used to identify to which of the two previous subjects is being referred. The

---

48 In (70), the word which expresses accompaniment *i-chop* ‘FUT2-two’ indicates that two participants were together. Whether or not both the puma and the man are subjects, this expression is used.

49 Dooley and Levinsohn (2001:131) consider the S1 context to include “situations in which the subject and nonsubject of the previous sentence combine to form a single, plural subject,” such as the case in (71).
reference to the puma may also be overencoded, since the reference occurs in the peak, emphasizing the confrontation between the participants. The addressee of (72–73), the man, is explicitly stated, rather than being left implicit, as it often is. This may be because the puma and man have arrived to a new location and it would be possible for another participant to be there. So, the explicit reference to the addressee makes it clear that the puma is, indeed, talking to the man and not to anyone else.

In this episode, there are three subject references to either the puma or the man that are focused nouns, rather than the default nouns. All three of these references (76–77, 82–85, and 86) occur in contexts where the subject was previously the addressee and is now the speaker. This use of focusing seems to slow down the action and emphasize the turn-taking of the dialogue and action. This unusual pattern of reference can be attributed to the fact that it occurs in the peak of the narrative.

The marked reference to the puma, in (78), is likely overencoded because of the shift from the dialogue to the continuing action. That is, the puma has just spoken, so the reader may expect the man to reply; but instead, the puma remains the subject because the dialogue has ended. Instead of the usual null reference used to refer to a subject that remains the same, an explicit reference is used. Even though the referential form used in (78), a focused noun, is the same that was used to refer to the puma in the dialogue, the fact that the reference is overt is enough to keep the reader on track. The overt reference when a null is the default is, therefore, used to catch the attention of the readers and make sure they are successfully tracking who is doing what. This is an example of the interplay between pragmatics and grammar, where the speaker may actually choose a specific reference, based on his/her interpretation of how effectively the listener is tracking the participants, rather than strictly basing the choice on default rules.

This episode also includes two more examples of focused nouns being used to reference major characters when they have been uninvolved in the previous clause (S4). These references are found in (80–81) and (87).

In spite of the different factors that have influenced referential choice in the peak, the resulting choices are very often focused nouns (80 percent of all the subject references in the peak are encoded by focused nouns). This repeated use of a fairly heavily encoded reference heightens the attention given by the reader to the peak. The peak ends with the bullet hitting the puma in the stomach. Following that event, episode seven (table 35) opens with a connector uye’loga ‘that being finished’, indicating that the previous events are finished and the plot is moving forward again.

Table 35. Episode seven (Clauses 90–98)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
</table>
The first reference in episode seven (referring to the man) is not the default focused noun, nor is it overencoded, as might be expected, due to the episode boundary. Rather, it is a noun that is not focused. The lack of focusing may be allowed, since the most recent previous subject, besides those that are inanimate, is the man. That is, between this reference to the man and the previous reference to him, there are no intervening participants as subjects of clauses. The only intervening reference to a participant is to the puma, but this is a nonsubject reference found in clause (89). So, although clauses (90–91) involve a change of subject, the change is not as much of a change as it would be if another participant had been the previous subject. The shift of using a noun after so many instances of focused nouns in the peak also indicates to the reader that the peak is over.

The reference to the puma in clauses (92–98) is a noun, the default encoding for a speaker that was previously an addressee. The addressees are not explicitly referred to in this episode, since the participants in the dialogue have already been established and there is no ambiguity.

The narrative draws to a close with clauses (99–100) (see Table 36). The episode boundary is marked by a time expression, axta or knu ‘from that time on’. There are no real surprises in terms of participant reference here in the conclusion of the story. The first reference (referring to the puma) is overencoded, due to the boundary, so a noun is used in place of the default null reference. The final clause does not use any marked reference forms, indicating that SFOZ allows default forms rather than requiring specific constructions, as does Guevea de Humboldt Zapotec (Kreutz 1980), for exiting participants.

### Table 36. Conclusion (Clauses 99–100)

<table>
<thead>
<tr>
<th>S</th>
<th>C</th>
<th>CONN</th>
<th>SUBJECT</th>
<th>S VALUE</th>
<th>NONSUBJECT</th>
<th>N VALUE</th>
<th>TRANSLATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>99</td>
<td>From that time</td>
<td>puma [1]</td>
<td>+S1</td>
<td></td>
<td></td>
<td>return not</td>
</tr>
</tbody>
</table>

Having looked at participant reference strategies in three SFOZ texts (chapter 5) and the MM text in detail (chapter 6), I now draw conclusions from this study as a whole in chapter 7.
Conclusions

7.1 Default encodings

Applying Dooley and Levinsohn’s (2001) and Givón’s (1983, 1994) methods to the SFOZ texts resulted in identifying certain patterns of participant reference and tracking. The fact that many referential devices were used consistently in a given context confirms the idea that “language is language only in context” (Longacre 1996:1). The specific contexts actually refer to a variety of features which influence referential choice. These features include the participant’s grammatical and semantic role in a clause, participant ranking, referential distance, and changes that occur within these features (e.g., change of subject). Factors that often lead to nondefault encoding of references in SFOZ are episode boundaries, participant rank, changes at peak and the need to disambiguate.

The default encodings as outlined in the results of this investigation are listed here again in table 37.

<table>
<thead>
<tr>
<th>Referent was previously . . .</th>
<th>CONTEXT</th>
<th>DEFAULT ENCODING</th>
</tr>
</thead>
<tbody>
<tr>
<td>same subject</td>
<td>S1</td>
<td>Ø</td>
</tr>
<tr>
<td>addressee</td>
<td>S1 (speaker desires to show respect)</td>
<td>pronoun</td>
</tr>
<tr>
<td>nonsubject</td>
<td>S2</td>
<td>noun</td>
</tr>
<tr>
<td>uninvolved</td>
<td>S3</td>
<td>NP</td>
</tr>
<tr>
<td>never mentioned</td>
<td>S4 (for minor participants)</td>
<td>NP (focused noun)</td>
</tr>
<tr>
<td>in same nonsubject role</td>
<td>Intro</td>
<td>NP (quant + noun)</td>
</tr>
<tr>
<td>speaker, but now addressee</td>
<td>N1</td>
<td>noun</td>
</tr>
<tr>
<td>in different role</td>
<td>N2</td>
<td>—</td>
</tr>
<tr>
<td>uninvolved</td>
<td>N3</td>
<td>noun</td>
</tr>
<tr>
<td></td>
<td>N4</td>
<td>noun</td>
</tr>
</tbody>
</table>

Though not specifically mentioned in the chart of default encodings above, identifying global VIPs, as suggested by Dooley and Levinsohn (2001), reveals other patterns of participant reference. For instance, if the speaker is desiring to show respect by using the pronoun, the global VIP can be encoded almost exclusively with the pronoun, no matter what the context is. This is the case in *Tsi’ll bur* ‘Ten donkeys’, where the owner of the donkeys is the global VIP. Besides his introduction and one case of a focused pronoun, he is always referred to with a pronoun. Since this was the only global VIP in this study, the question remains: How do speakers encode a global VIP when they do not necessarily want to show respect? I hypothesize that they would use a null reference, which is the counterpart to the pronoun in the default encodings. However, it could be that null reference does not give enough information in all contexts to disambiguate the participants. Consider that in TB, the global VIP is a person, whereas the other
participants are animals. This means that the pronoun (used only for humans) is enough to identify the human referent without ambiguity. However, if a null reference were used, there would be some references that could be ambiguous as to whether they referred to a man or an animal.

Another finding about participant reference in SFOZ is that major participants seem to be introduced as subjects of matrix clauses, whereas minor participants tend to be introduced as subjects of embedded clauses.

Given the above results, I now specify how null reference, nouns, and other, less common, encoding devices are used in SFOZ.

7.2 Null reference

Null reference occurs in a very limited context. Its primary use is to reference a subject that is the same in the preceding clause. This is common in many languages (e.g., Chinese, Li and Thompson 1979:313). Not only does null reference occur across clause boundaries, it also occurs across sentence boundaries when the subject remains the same. This is a freer usage of null reference than English and can cause English speakers to lose track of participants in SFOZ, while native SFOZ speakers seem to have no difficulty. Null reference is used in the same-subject context (S1) regardless of participant rank as major or minor participant. This differs from other languages (e.g., Mambila) where main participants can be encoded by null reference, but other participants are usually referred to by nouns or pronouns (Staley 1995:99). SFOZ also differs from Guevea de Humboldt Zapotec, which takes participant rank into account in its use of null reference. Kreutz (1980:13) claims that when two participants are referred to pronominally in a paragraph, the higher ranking one is referred to by an overt pronoun, and the non-topic one is referred to by Ø. In SFOZ, null reference can be used for a participant of any rank.

In spite of the fact that null references have no encoding material, they can still be used in potentially ambiguous situations without ambiguity. For example, there may be more than one participant on stage (a potentially ambiguous situation), and yet a null may clearly reference the subject of the preceding clause.

The analysis also shows, however, that a null reference is not always used in every same-subject context. Specifically, when the speaker desires to show respect, as in TB, a pronoun may be used in place of a null.

Not every case of missing encoding is to be considered a null reference. The most common example of this is that of an addressee in a dialogue. As long as a dialogue is taking its normal course between two speakers who alternate turns, the addressee is not explicitly encoded. This is not considered null reference, since the addressee, by definition, always fills a nonsubject role and is not required to be explicit by grammatical rule.

7.3 Nouns

Unlike null reference, nouns occur in a variety of contexts. The widest usage of nouns is in nonsubject contexts. By default, any participant in a nonsubject role is encoded by a noun (with the exception of addressees, as mentioned above). Therefore, depending on the type of text, and how many nonsubject roles occur in a text, a discourse can include a good many nouns.

In addition to nonsubject encodings, there is one subject context encoded by nouns; namely, when a subject is the addressee in the preceding clause. In a text which has a lot of dialogue, such as ‘The puma and the man’, this can occur frequently.
When a major participant is a subject, after not being involved in the preceding clause (S4), it is encoded by a focused noun rather than by a noun phrase. Though such nouns are focused, they are, nevertheless, classified as ‘noun’ in this study. What I first considered to be underencoding of a major participant by a focused noun, I now find to be the default encoding for a major participant.

It can also be assumed that a noun may be used in other cases that call for underencoding. For instance, if a noun phrase is expected as the default encoding where underencoding is allowable or even desired, due to, say, participant rank, a noun may be used in place of the noun phrase.

Nouns serve an anaphoric function of referring to participants that were previously introduced. In many languages, such as English, pronouns function much the same way as SFOZ nouns do. Since nouns have more encoding material than pronouns, this explains why the frequent use of SFOZ nouns sounds redundant and overencoded to English speakers, like myself, who are used to minimally encoded anaphoric references.

7.4 Other encoding devices

While nouns and null reference can be found quite easily in almost any SFOZ text, third-person pronouns are rare. Of the three texts studied, pronouns were only found in one (TB). In that text, pronouns were exclusively used to reference the owner of donkeys. In fact, the donkey owner was almost exclusively referred to by pronouns. There are two conditions that are met in that text which account for the use of the pronouns. First of all, the speaker is apparently desiring to show respect either to the audience or to the referent. From my knowledge of SFOZ beyond the texts in this study, this seems to be the most important condition that must be met for pronoun usage. Another condition that explains some of the marked uses of pronouns is that the donkey owner is the global VIP of the TB text. This condition allows pronouns to be used to refer to the donkey owner even when the default encoding would call for more encoding material than a pronoun.

When a participant is not a global VIP, but respect is still being shown by the speaker, a pronoun can at least be used in subsequent clauses when the participant remains the subject (S1). This is expected because pronouns are classified together with null reference on the topic-continuity continuum in (43) and are expected to encode highly continuous topics such as those in a same-subject context. Further study is needed to determine whether pronouns are used more widely when respect is not being shown.

Noun phrases are used in SFOZ to introduce participants for the first time. Introductions actually use a specific type of NP: a quantifier followed by a noun. Other than these special introductory NPs, I have considered any type of NP as a general NP, without looking at its specific composition. For example, I have not distinguished between those that include demonstrative adjectives, like ni'b knu ‘that boy’, and those that include qualitative adjectives, like nay gol ‘old woman’. A future study of SFOZ participant reference should consider patterns of usage for each different type of NP.

SFOZ noun phrases are also used to mark less continuous participants, such as when the subject of a clause was previously a nonsubject (S3) or for minor participants who were not mentioned at all in the previous clause (S4).

Nouns, pronouns, and noun phrases may all be focused by the focus marker: lë' or by fronting the constituent. The specific use of focusing in SFOZ was not studied in this research,
but some observations were made. Focused nouns are the default encoding device for major participants who were not involved in the preceding clause. Because a major participant is more accessible in the mind of a hearer, a noun can be used rather than the NP used for a minor participant. However, the focusing of a noun helps a reader to recognize that a participant is returning to subject role after not having been in that role for at least one clause. In addition to adding the focus marker, focusing usually includes a change in word order, which also may be significant in tracking participants. At times, focusing is also used at episode boundaries. Though focused forms are classified with their corresponding nonfocused forms in the continuity continuum in (43), it seems that focusing might be a mild form of overencoding available for use in place of a device normally used for less continuous topics, such as a noun phrase.

Proper names were not used in any of the studied texts. They are encoding devices that could potentially have their own set of rules for participant reference. They will require further study.

7.5 Implications for theory

In analyzing SFOZ participant reference, it quickly became clear that a variety of factors influence referential choice. Sequential strategies that consider referential distance, while helpful, do not account for all of the references throughout SFOZ discourse. Indeed, other factors such as VIPs (Dooley and Levinsohn 2001), episode boundaries (Anderson et al. 1983), and participant ranking (Longacre 1995) come into play when describing participant reference and tracking in SFOZ.

Of the two methods used, Dooley and Levinsohn’s method (2001) proved more helpful than Givón’s (1983, 1994) in analyzing patterns of participant reference and tracking, though both methods have their strengths. While Givón’s measurement of referential distance was useful in determining which devices were used for more continuous participant references, Dooley and Levinsohn’s method handled the same idea of referential distance by considering the specific context of each reference (e.g., some contexts had inherently low referential distance, as when the subject remained the same in consecutive clauses). I did not find ‘potential interference’ and ‘persistence’ calculations proposed by Givón (1983, 1994) to shed much light on referential choice in the SFOZ texts in this study, although they might be useful when looking at other texts that are more similar and have a broader usage of the encoding devices.

Since Givón’s method focused on linear features of the texts, it enabled quantification of the data which proved helpful in determining some default reference patterns in SFOZ. In the cases where it was not helpful, it may have been due more to the differences in the texts I examined than to any fault of the method itself. Because of the method’s emphasis on quantification, it does not attempt to address nondefault encodings of participants that are caused by nonlinear factors, such as participant rank and episode boundaries. Because of this, using his method without additional investigative strategies cannot be relied upon to illuminate the wide variety of referential patterns in SFOZ.

Dooley and Levinsohn’s method of analyzing participant reference leaves room for considering a variety of other factors, including discourse factors, that might affect referential choice or cause marked forms to be used. However, one weakness of this approach is the inattention to both clause and sentence division. In their chart, Dooley and Levinsohn (2001:128) have given each sentence a separate number, but listed each clause on a separate line. Though this does allow the reader to know where both clause divisions and sentence divisions are, it is difficult to know whether to consider a preceding clause or entire sentence when labeling contexts. This means that certain factors that influence referential choice can remain hidden from
Another weakness is that the chart does not show the hierarchical structure of a discourse. That is, if the text is divided by clauses and the dependent and embedded clauses are listed separately, just as the independent clauses are, there is no way to easily see the difference in those clauses which might play a role in participant reference. For instance, a reference in an embedded clause might have been chosen, based on the context of a reference in a preceding independent clause, without considering references in the intervening embedded clauses.

One positive aspect of Dooley and Levinsohn’s model is its consideration of subject versus nonsubject references. In SFOZ there is a clear contrast between how subject and nonsubject references are handled. Givón’s method does not account for this difference.

The consideration of various factors influencing participant reference and tracking in the theory of Dooley and Levinsohn lends itself well to analyzing participant reference data. Other theories that speak of the variety of factors are also helpful to consider, although they often do not present a practical way of applying the theory to actual analysis.

### 7.6 Implications for translation

The results of this research in participant reference and tracking have many implications for translating literature into Zapotec. If a text is translated bit-by-bit, without reference to hierarchical structure, wrong referential choices will be made that lead to ambiguity and unnatural discourse. For example, if the MM text were translated clause by clause, without considering episode boundaries and participant introductions, inappropriate encoding devices would be used. It is important, therefore, when translating, to note features that influence referential choice, such as episode boundaries, changes of subject, and introductions (Dooley and Levinsohn 2001:111–135).

Uses of encoding devices in one language may not correlate with those of corresponding encoding devices in another language. A translator cannot simply replace all nominal references in a source language with nominal references in a target language. If all pronouns of a source language, for example, are simply replaced by SFOZ pronouns, listeners may be confused, since SFOZ third-person pronouns are used to show respect, whereas respect may have no relevance to pronouns in the source language.

When literature is translated into SFOZ, the translator must consider the patterns of participant reference and tracking in SFOZ and analyze the text in regard to the relevant features that affect those patterns. Only then will an effective translation be possible.

### 7.7 Strengths and weaknesses of the study

This study provides a firm foundation on which to continue studies of participant reference and tracking in SFOZ. Though some areas of participant reference have not been touched upon by this study, basic default patterns and hypotheses that have been formed will aid in producing more natural and satisfying written literature in SFOZ. In addition, the results and conclusions presented will provide additional analysis that can further comparative Zapotec studies and may shed new light on processes that exist in Zapotec languages that have not yet been studied.

One weakness of this study was the limited scope of the text corpus. By limiting the study to three texts, certain encoding devices, such as proper names, were never exhibited and, therefore, are not accounted for in the study. The discussion (in section 6) of the reference in clause (38) illustrates the need to consider sentence divisions in addition to clause divisions.
could not be studied. Another weakness having to do with the texts was that they were each slightly different genres. Though all three were narratives, MM was an animal folktale, TB was a very short humorous vignette, and SB was a personal experience narrative. TB was narrated using the pronoun to show respect, whereas neither of the other texts used the respect pronoun at all; so there was no other text with which to compare the usage. Since each text was narrated by a different person, it is possible that individual styles of reference were coming across in the texts more than overall language-specific patterns. It would have been good to have a variety of speakers each narrate a number of very similar texts. This would have enabled me to identify referential patterns for each speaker and compare them across the group to find out how similarly or differently various speakers encode the participants of the same text.

Another weakness of this study is the lack of specificity. For instance, I grouped several of the encoding devices together at each level of the continuity continuum, rather than determining the relative levels for each encoding device. Though this may be an accurate picture of the way SFOZ uses these devices, it is possible that with more data, each encoding device would separate into its own level on the continuum. Again, this study provides a good starting point, but further analysis is needed. Another example of classifying devices together rather than separating them is with regard to noun phrases. NPs formed with demonstrative adjectives might be permitted in very different contexts from NPs with qualitative or quantitative adjectives, yet in this study they were classified together and treated as a single encoding device.

7.8 Recommendations for further research

As I researched participant reference and tracking in SFOZ, additional questions arose. Some questions could be answered with further study of the grammar and syntax of SFOZ. Most of the questions are very much a part of the topic of participant reference, but were not covered in this study. Further research on this topic could answer the following questions:

(63)

- How do speakers of SFOZ choose the level of specificity a noun should have? For example, in Tsi\' bur ‘Ten donkeys’ why does the narrator sometimes say ‘animal’ and other times say ‘donkey’ for the same referent?

- How do speakers of SFOZ handle multiple participants on stage that are very similar and, thus, cannot be distinguished from each other with only a noun (e.g., three girls of the same age)?

- How are individuals tracked if their proper names are given in the text?

- How similarly would various narrators encode the participants of the same story? How similarly would one narrator encode the participants in various retellings of the same story? How strictly do native speakers adhere to the patterns of reference outlined in this study?

- What role does focusing play in participant reference and in the language in general?
• How do reference strategies in SFOZ change for a different genre of literature, such as hortatory or expository?

The subject of participant reference and tracking is quite broad as can be seen in the preceding questions. I hope to take the knowledge gained from this investigation and further my understanding by continuing to research this topic and doing further discourse analysis of SFOZ texts.
Appendix:
Corpus of Texts

The three texts analyzed in this study have been included here for the reader to further investigate. Each of the three texts is divided into clauses, with the number of the clause listed at the left. Relative clauses have been left embedded in their matrix clauses. Sentence divisions, though not always clear in SFOZ, have been indicated with a period at the end of one sentence and a capital letter beginning the next sentence.

**Me'dz ichop mgiy**

1. \( Uy-ak-u \) \( tib \) \( bwelt \)
   COMP-happen-3IN one time
   Once upon a time

2. \( u-yu \) \( ti \) \( me'dz \)
   COMP-be one puma
   there was a puma

3. \( xla'n \)
   PROG-want
   wanting

4. \( gu'n \) \( preb \)
   FUT2-do test
   to do a test

5. \( gann \)
   see
   to see

6. \( bi \) \( yu \) \( tib \) \( mgiy \) \( mas-che \) \( ncho'y \)
   INT be one man more-more strong
   \( que \) \( lo \) \( me'dz. \)
   than face puma
   if there was a man who was stronger than the puma.

---

51 This animal folktale was narrated by Fernando R.
52 The labels on the various aspects of SFOZ verbs have not been fully studied yet. That is, some of the aspects seem to be used in various functions and finding a label to describe all those functions is challenging. The label for aspect on each verb describes one of the uses of that aspect marker, but may not fit particularly well with the use in that particular instance.
And-now PROG-know-NEG puma man
But, the puma didn't know man.

So the puma went

So the puma went

and-now go puma

and-now go puma

and arrived at where a mule was.

He says,

He says,

The mule says,

The mule says,

but man exists.

but man exists.

Man has me tied up here.

Man has me tied up here.

Man is causing me to suffer.

Man is causing me to suffer.

The initial 's' in sia could potentially be analyzed as a future aspect marker. Further study is needed before a decision can be made about how it should be analyzed.

Though this verb carries a future marker, it is translated with the English past tense. This is because in SFOZ this verb is marked as future when someone has already gone, but has not yet returned.
(17) Ch-ey-n yag dze'y mgiy.
PROG-carry-1S wood cause.to.burn man
I am carrying wood that the man will burn.

(18) Bi'tsi xla'n'l
If PROG-want-2S
If you want

(19) g-wiy-l mgiy
FUT2-see-2S man
to see man

(20) gwa
go!
go

(21) kib mgiy."
seek! man
look for man."

(22) Gadzekna lë' me'dz sia
then FOC puma go
Then the puma went

(23) u-dzinn det zu ti ngon.
COMP-arrive where stand one cow
and arrived where a cow was.

(24) Lë' me'dz ch-ëp,
FOC puma PROG-answer
The puma says,

(25) “¿Be lu' n-ak mgiy?”
INT 2S STAT-be man
“Are you man?”

(26) Ch-ëp ngon,
PROG-answer cow
The cow says,

(27) “N-ak-d-n mgiy.
STAT-be-NEG-1S man
“I am not man.
(28) *Mgiy ne'g n-këdu'-n*
    man here STAT-be.roped.up-1S
    Man has me tied here

(29) *per x-kë ldo' mgiy na'.*
    but PROG-hit stomach man 1S
    but man likes me.

(30) *X-ne'dz gix ch-ow na' nis na ch-a na'.*
    PROG-give plant PROG-eat 1S water and PROG-drink 1S
    He gives me plants that I eat and water that I drink.

(31) *Bi'tsi x-la'n-l*
    If PROG-want-2S
    If you want

(32) *g-wiy-l mgiy,*
    FUT2-see-2S man
to see man

(33) *goliaz.”*
    wait!
    wait.”

(34) *Gadzekna lë' me'dz uyu'n gëts xits gix.*
    then FOC puma COMP-do hide back plant
    Then the puma hid behind the grass.

(35) *Konde s-tib rat lë' me'dz uy-ann*
    when FUT1-one moment FOC puma COMP-see
    A little later the puma saw

(36) *lë' mgiy z-id-le*
    FOC man FUT1-come-already
    man already coming,

(37) *n-ey gix, g-ow ngon.*
    STAT-carry plant FUT2-eat cow
carrying plants for the cow to eat.

(38) *Konde u-dzinn mgiy,*
    when COMP-arrive man
    When the man arrived,
(39) \textit{u-gix gix lo ngon.}  
COMP-put.down plant face cow  
he put down the plants for the cow.

(40) \textit{Lë' me'dz u-dzinn}  
FOC puma COMP-arrive  
The puma arrived

(41) \textit{ch-ëp lo mgiy,}  
PROG-answer face man  
and says to man,

(42) \textit{“¿Be lu' n-ak mgiy?”}  
INT 2S STAT-be man  
“Are you man?”

(43) \textit{Ch-ëp mgiy,}  
PROG-answer man  
The man says,

(44) \textit{“Na'-u.”}  
1S-3IN  
“I am.”

(45) \textit{Gadzekna ch-ëp me'dz,}  
then PROG-answer puma  
Then the puma says,

(46) \textit{“Bi'lltsi xka-u,}  
if that’s.how-3IN  
“If that's how it is,

(47) \textit{g-ak-n tib preb.}  
FUT2-happen-1I one test  
let’s do a test.

(48) \textit{I-kë-n tib chedz.”}  
FUT2-make-1I one scream  
We'll scream.”

(49) \textit{Lë' mgiy ch-ëp,}  
FOC man PROG-answer  
The man says,
(50) “Z-ak-u.
FUT1-be.able-3IN
“Ok.

(51) Di’ltsi xka-u,
If that’s how-3IN
If that’s how it is

(52) gwa,
go!
go,

(53) kib det i-köï-n chëdz.
seek! where FUT2-make-1I scream
look for where we will scream.

(54) Lë’ na’ tsaña du-kub-na.”
FOC 1S release 1S.POSS-bucket-1S
I'll leave my bucket.”

(55) “Z-ak-u”
FUT1-be.able-3IN
“OK,”

(56) chëp me’dz.
PROG-answer puma
says the puma.

(57) Lë’ me’dz sia
FOC puma go
The puma went

(58) uy-a
COMP-go
and went

(59) kib det gya n-ak.
FUT2.seek where high STAT-be
to look for where it is high.

(60) Lë’ mgiy uy-a lidz
FOC man COMP-go home
The man went home
(61) uy-ey tib gi'b, gu'n n-ak skopet. 
COMP-carry one gun thing STAT-be rifle 
and carried(got) a gun, the kind that's a rifle.

(62) U-gu n-cho'b plom.  
COMP-put.in STAT-grow ammunition  
He put in a lot of ammunition.

(63) Uy-ey-l-u,  
COMP-carry-already-3IN  
He carried it,

(64) sia  
go  
he went

(65) u-dzinn det zu ngon.  
COMP-arrive where stand cow  
and arrived to where the cow was.

(66) Lë' me'dz u-dzinn  
FOC puma COMP-arrive  
The puma arrived

(67) ch-ëp,  
PROG-answer  
and says,

(68) “Sia-na  
go-1I  
“Let's go

(69) lë' det i-kë-n chedz u-dzal-la.”  
FOC where FUT2-do-1I scream COMP-be.found-already to the place we will scream that was found.”

(70) Gadzekna lë' me'dz sia i-chop mgiy.  
them then FOC puma go FUT2-two man  
Then the puma went with the man.

(71) Konde u-dzinn ga  
when COMP-arrive there  
When they arrived there
(72) lë' me'dz ch-ëp lo mgiy,
FOC puma PROG-answer face man
the puma says to the man,

(73) “Kë' chedz na.”
make! scream now
“Scream.”

(74) Ch-ëp mgiy,
PROG-answer man
The man says

(75) “Lu' g-ne'ch.”
2S FUT2-go.ahead
“You will go first.”

(76) Gadzekna lë' me'dz ch-ëp,
then FOC puma PROG-answer
Then the puma says,

(77) “Z-ak-u.”
FUT1-be.able-3IN
“OK.”

(78) Lë' me'dz u-kë' chedz.
FOC puma COMP-make scream
The puma roared.

(79) U-yu mbil giy dën.
COMP-be echo mountain field
There was an echo (in the mountain).

(80) Uy-e'loga lë' me'dz ch-ëp,
COMP-be.finished-afmnt FOC puma PROG-answer
After that, the puma says,

(81) “Na-na lu' kë' chedz na.”
and-now 2S make! scream now
“Now you scream.”

(82) Gadzekna lë' mgiy ch-ëp,
then FOC man PROG-answer
Then the man says,
“Per gwiy-d-la lo na’.
but look.at-NEG-2S face 1S
“But don't look at me.

Yi lad gwiy.
over.there side look.at!
Look over there.

Lë’ na’ i-kë’ chedz.”
FOC 1S FUT2.make scream
I'll scream.”

Lë’ me’dz u-wiy ti kwa.
FOC puma COMP-look.at one side
The puma looked to one side.

Lë’ mgiy u-zo’b skopet laz.
FOC man COMP-cause.to.sit rifle chest
The man positioned the gun at his chest.

U-gitsyë-u.
COMP-explode-3IN
It fired.

U-ga’-u ldo’ me’dz.
COMP-hit-3IN stomach puma
It hit the puma’s stomach.

Uy-e’lo-ga ch-ëp mgiy,
COMP-be.finished-AFM PROG-answer man
Then the man says,

“¿Be s-kë-n s-ti chedz?”
INT FUT1-make-1I FUT1-one scream
“Should we scream again?”

Ch-ëp me’dz,
PROG-answer puma
The puma says,

“Lë’ya-na porke axta ldo’-n në chow.
o-no-AFM-now because until stomach-1S hurt much
“Not now because all the way to my stomach it hurts a lot.
(94) *U-zēn tib yo'b.*
COMP-catch a pain
A pain has grabbed me.

(95) *Lē' bey-le-ga.*
FOC know-already-AFM
It was enough with that.

(96) *Masde, gann*
well.then see
Well then, we'll see

(97) *bok z-id-na*
when FUT1-come-I1
when we come back

(98) *i-kē'-na stib chedz."
FUT2-make-I1 FUT1-one scream
and we scream again."

(99) *Axta or k-nu ň-obiche-chi me'dz*
from hour exactly-that UNR-return-more puma
From that time on the puma didn't go back

(100) *n-dza'l mgiy.*
UNR-meet man
and didn't meet the man again.
This humorous short story was narrated by Camerino R.
(11) *Konde uy-ab gik xi*
when COMP-fall head 3S
When it occurred to him

(12) *chëp xi,*
PROG-answer 3S
he says,

(13) “*Gënd.*”
there.is.not
“It's not here.”

(14) *Uw-iy xi*
COMP-see 3S
He saw

(15) *gënd s-tib bur.*
there.is.not FUT1-one donkey
the other donkey wasn't there.

(16) *Ge'-ga bur nech lo xi.*
nine-AFM donkey lead face 3S
Nine donkeys were in front of him.

(17) *Sia-l uy-uslo uy-ak xi gil-xgab.*
go-already COMP-begin COMP-happen 3S NOM-think
Then he began to worry.

(18) *Lë' xi uy-uslo x-gib s-tib bur.*
FOC 3S COMP-begin PROG-seek FUT1-one donkey
He began to look for the other donkey.

(19) *Lad lad xgib xi bur.*
side side PROG-seek 3S donkey
He's looking everywhere for the donkey.

(20) *Gënd stib bur.*
there.is.not FUT1-one donkey
The other donkey is not there.

(21) *Stse'-la uy-ak xgib xi bur*
a.long.time-already COMP-happen PROG-seek 3S donkey
Much time passed as he was looking for the donkey
(22) *konde uw-iy xi*
when COMP-see 3S
when he saw

(23) *det xits s-tib bur sob xi.*
where back FUT1-one donkey sit 3S
that he was sitting on the other donkey.

(24) *Xkasa n-ak kwent.*
that.is.how STAT-be story
That's the story.

(25) *Or k-nu cha ch-ęp xi,*
hour exactly-that more PROG-answer 3S
Then, he said,

(26) *‘xits bur sob na!’*
back donkey sit 1S
‘I'm sitting on the donkey

(27) *y læ’ na’ x-gib s-tib bur.’’*
and FOC 1S PROG-seek FUT1-one donkey
and I am looking for the other donkey.’’

(28) *Gu'nnuta' n-ak kwent u-gu dad nu lo-n.*
that.one? STAT-be story COMP-tell man that face-1S
That is the story that the man told me.
San Bartolo

(1) Lë’ na’ ku tib kwent gu’n u-wiy-nu,
FOC 1S tell a story thing COMP-see-1E
I’m going to tell a story, the thing that we saw

(2) uy-a-nu tib gedz lë San Bartolo.
COMP-go-1E one town name Saint Bartolo
when we went to a town called San Bartolo.

(3) U-cha’-nu Mitl or xun ichol
COMP-leave-1E Mitla hour eight half
We left Mitla at eight thirty

(4) y u-dzi’n-nu San Bartolo or tsi’ ichol.
and COMP-arrive-1E Saint Bartolo hour ten half
and we arrived in San Bartolo at ten thirty.

(5) Udzi’n-nu tib chow tiend.
COMP-arrive-1E one mouth store
We arrived at the entrance of a store.

(6) Ts-u chop mgiy.
HAB-stand two man
Two men were standing.

(7) Ga x-to-zë gis.
there PROG-be.sold-CLTV pot
There many pots are sold.

(8) Lë’ mgiy, minn i-ne’ch lo-nu,
FOC man person FUT2-go.ahead face-1E
uy-et lën kamion
COMP-got.down in truck
The man, the person who would guide us got down from the truck

(9) u-nab-di’dz lo mgiy k-nu,
COMP-request-word face man exactly-that
and asked that man

(10) ch-ëp,
PROG-answer
saying,

A personal account told by Angelina Cruz Juárez.
“Will you give permission for us to enter there?”

Pass-1E over.there for us to see the pots?"

The man replied,

"It can't be."

The man replied,

“(All of you) go down below.

There is another store that is bigger.

There more pots are sold.”

Then the man that drove the truck got up into the truck.
(22) u-dzi'inn-nu chow tiend k-nu.
COMP-arrive-1E mouth store exactly-that and we arrived at the front of that store.

(23) Lët' kamion u-su.
FOC truck COMP-stand
The truck parked.

(24) Lët'-nu uy-et-së,
FOC-1E COMP-get.down-CLTV
We got down

(25) lët'-nu u-ded lën tiend k-nu.
FOC-1E COMP-pass in store exactly-that and we entered that store.

(26) U-ded-nu ga
COMP-pass-1E there
We entered there

(27) u-wiy-nu dil x-to-zë gis.
COMP-see-1E where PROG-be.sold-CLTV pot we saw where all the pots are sold.

(28) S-tsa' minn i-chë ngol,
FUT1-make person FUT2-all doll
The people make all kinds of dolls,

(29) s-tsa' minn i-chë-sa man,
FUT1-make person FUT2-all-CLTV animal
the people make all kinds of animals

(30) s-tsa' minn pchog ya's.
FUT1-make person shell black
the people make masks.

(31) I-chë-sa gu'n-u u-wiy-nu.
FUT2-all-CLTV thing-3IN COMP-see-1E
We saw all those things.

(32) Uloë gadzekna u-wiy-nu dil kë-zë letrer
COMP-finish then COMP-see-1E where hang-CLTV sign
Then we saw where the signs were hung

(33) gann ba klas yu
FUT2-see which class clay
to see from what kinds of clay
(34) **x-ia'** **gis.**
PROG-be.made pot
the pots were made.

(35) **Y u-lox dzekn u-wiy-nu dil kë-zë foto.**
and COMP-finish then COMP-see-1E where hang-CLTV photo
And then we saw where the pictures were hanging.

(36) **Kë' foto chën minn gol.**
hang photo belonging.to person old
There were pictures of the old person.

(37) **Minn u-yab gu'n-u gik tiemp u-za' gu'n-u.**
person COMP-fall thing-3IN head long.ago COMP-form thing-3IN
The person who thought of doing this a long time ago made those things.

(38) **Na-na u-ya'n xin minn.**
and-now COMP-stay.behind child person
Now her children have remained.

(39) **Xin-le minn x-to gu'n-u.**
child-? person PROG-sell thing-3IN
Her children sell them.

(40) **Xin-le minn kë xto-wa.**
child-? person do PROG-sell-3IN
Her children do the selling.

(41) **U-lox ga dzekn, gadzekn lë'-nu u-ka**
COMP-finish-AFM then then FOC-1E COMP-receive
*tib-ka gis.*
one-DSTR pot
Then we each bought our own pot.

(42) **U-lox dzekn u-lox u-zi-nu,**
COMP-finish then COMP-finish COMP-buy-1E
Then when we finished buying,

(43) **(x-ne') lë' minn s-tsë' kamion x-ne',**
PROG-say FOC person PROG-drive truck PROG-say
(says) the man who was driving the truck says,

(44) **"Na-n sia-na.**
and-now go-1I
"Now, let's go.
We'll go to another place where we'll see the actual place where they make the pots.

Here is just where they sell.

Now, let's go where they are made."

Then we returned.

We arrived beside that place.

There was a house where (pots) are made

The man, the one driving the truck, arrived there.

An old woman was standing at the entrance.

He says,
“¿Be s-ne'dz-l  di'dz
INT  FUT1-give-2s  word
"Will you give permission

(57)  ted-nu  yi
pass-1E over.there
for us to enter

(58)  g-wiy-nu
FUT2-see-1E
for us to watch

(59)  gann
FUT2-see
for us to see

(60)  xku'n  x-ia'  gis?"
how  PROG-be.made  pot
how the pots are made?"

(61)  Ch-ëp  nay  gol,
PROG-answer  lady  old
The old lady replies,

(62)  “¿Be kwa'n  xla'nn-l?”
INT  what  want-2S
"What do you want?"

(63)  Ch-ëp  mgiy,  ch-ëp,
PROG-answer  man  PROG-answer
The man replies, saying,

(64)  “Es  que  lë'-nu  xla'nn
is  that  FOC-1E  want
"It’s that we want

(65)  gann  xku'n
FUT2-see  how
to see how

(66)  g-wiy-nu
FUT2-see-1E
to see
(67) *xku*n x-*ia'* gis.'
   how PROG-be.made pot
   how the pots are made."

(68) “A *weno,* *ch-*ëp *minn,*
   oh okay PROG-answer person
   "Oh, okay," says the person,

(69) “*ku-de'd.*”
   IMP-pass
   "enter."

(70) *Uy-e'l*o-ga *dzekn u-ded-nu.*
   COMP-be. finished-AFM then COMP-pass-1E
   Then we entered.

(71) *Uy-e'l*o-ga *dzekn u-ri-y-nu*
   COMP-be.finished-AFM then COMP-see-1E
   Then we saw

(72) *lë' tib ndza'p sob xki*
   FOC one girl sit like.this
   a girl sitting there like this

(73) *x-*gu,
   PROG-put in
   putting

(74) *ch-*u'n *pintar lëd i-*chë* ngol, i-*chë*, i-*chë-sa* gu'n.*
   PROG-do paint against FUT2-all doll FUT2-all FUT2-all-? thing
   painting the sides of all the dolls, all the things.

(75) *U-lox* *dzekn u-dzi'n'nu gax-nu
   COMP-finish then COMP-arrive near-1E
   *dil sob ni'lb s-tsa' gis.*
   where sits boy FUT1-makes pot
   Then we arrived near where a boy who was making pots was sitting.

(76) *lë'* ni'lb k-*nu sob lo tib siy.*
   FOC boy exactly-that sit face one chair
   That boy was sitting on a chair.

(77) *Ga zu tib yag k-*nu.
   there stand one wood exactly-that
   There was a wooden board there.
Then the boy went to the board, a round board.

There was another one on the bottom side.

A round board was on the bottom.

The boy's foot was there.

In the front there was another round board.

He took this much clay

and it was put on top of that.

Then there was another glass of water sitting there.

Then with water that boy mixed the clay.
Then with the same water that he used to wash his hands, that boy made the clay wet.

The boy's other foot was turning the wooden board.

The pot was made.

Then the man that was our guide says,

"How is it made?

Tell it to us."
(98) “Xnu xnu x-ia'.
like this like this PROG-be.made
"Like this, they are made.

(99) Lë'¬u ne'g g-wiy
FOC-3IN here FUT2-see
Here it is, look,

(100) lë'-n i-za'-u.”
FOC-1S FUT2-form-3IN
I will make it.”

(101) Uy-e'lo-ga dzekna, uy-e'lo-ga x-wiy-nu
COMP-be.finished-AFM then COMP-be.finished-AFM PROG-see-1E
Then we watched

(102) lë' gis uy-e'lo-la.
FOC pot COMP-be.finished-already
and the pot was finished.

(103) Ni'b k-nu, non-ga tib du' u-zë't ni'b k-nu
boy exactly-that only-exactly one rope COMP-stick boy exactly-that
With just a string, that boy stuck

(104) u-zë't xan k-nu.
COMP-stick bottom exactly-that
that boy stuck the bottom of it together.

(105) Uy-e'lo-ga lë' uy-e'lo-l gis ga.
COMP-be.finished-AFM FOC COMP-be.finished-already pot there
And then the pot was finished.

(106) U-lox gadzekna u-wiy-nu gu'n-u.
COMP-finish then COMP-see-1E thing-3IN
We finished watching that.

(107) Uy-e'lo-ga dzekn lë' dad x-ne',
COMP-be.finished-AFM then FOC gentleman PROG-say
Then the man says,

(108) “Na-na sia-na.”
and-now go-1I
"Now let's go."
(109) *U-de'd* tib gal pēs lo ni'b.
COMP-give one twenty peso face boy
He gave twenty pesos to the boy.

(110) *Uy-e'lo-ga* lē' u-biy lē'-nu ga.
COMP-be.finished-AFM FOC COMP-return FOC-1E there
Then we came back from there.
References


