UNIVERSIDADE FEDERAL DE MINAS GERAIS

FACULDADE DE LETRAS

Bárbara Guimarães Rocha

THE FOCUS POSITION IN CINYANJA:

A TYPOLOGICAL AND FORMAL APPROACH

BELO HORIZONTE

2025

Bárbara Guimarães Rocha

THE FOCUS POSITION IN CINYANJA:
A TYPOLOGICAL AND FORMAL APPROACH

Dissertation presented to the Post-Graduate Program in Linguistic Studies (POSLIN) at the Federal University of Minas Gerais (UFMG) – College of Letters (FALE), as a partial requisite to obtain the title of Doctor in Linguistics.

Area: Theoretical and Descriptive Linguistics
Research line: Formal Studies in Languages
Advisor: Professor Doutor Fábio Bonfim Duarte

Belo Horizonte, Brazil
College of Letters
August – 2025

Rocha, Bárbara Guimarães.
R672f The focus position in Cinyanja [manuscrito] : a typological and formal approach / Bárbara Guimarães Rocha. – 2025.
1 recurso online (271 f.: il., tabs., maps.,p&b.) : pdf.

Orientador: Fábio Bonfim Duarte.

Área de concentração: Linguística Teórica e Descritiva.

Linha de pesquisa: Estudos Formais de Língua.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 231-240.

Anexos: f. 241-271.

Exigências do sistema: Adobe Acrobat Reader.

1. Línguas banto – Moçambique – Estudo e ensino – Teses. 2. Línguas banto – Sintaxe – Teses. 3. Línguas africanas – Teses. 4. Língua umbundu – Teses. I. Duarte, Fábio Bonfim. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 496.39

Ficha catalográfica elaborada pela Bibliotecária Priscila Oliveira da Mata - CRB/6- 2706
Biblioteca Professor Rubens Costa Romanelli - FALE/UFMG

UNIVERSIDADE FEDERAL DE MINAS GERAIS

THE FOCUS POSITION IN CINYANJA: A TYPOLOGICAL AND FORMAL APPROACH

BÁRBARA GUIMARÃES ROCHA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Estudos Linguísticos, como requisito para obtenção do grau de Doutor em Estudos Linguísticos, área de concentração Linguística Teórica e Descritiva, linha de pesquisa Estudos Formais de Línguas.

Aprovada em 20 de agosto de 2025, pela banca constituída pelos membros:

Prof. Fábio Bonfim Duarte

Universidade Federal de Minas Gerais

Prof. Quesler Fagundes Camargos

Universidade Federal de Rondônia

Profa. Helena Guerra

Universidade de Brasília

Profa. Simone Lúcia Guesser

Universidade Federal do Paraná

Prof. David Alberto Seth Langa

Universidade Eduardo Mondlane

Belo Horizonte, 20 de agosto de 2025.

Documento assinado eletronicamente por **Fabio Bonfim Duarte**, **Professor do Magistério Superior**, em 21/08/2025, às 14:13, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Helena da Silva Guerra Vicente**, **Usuária Externa**, em 23/08/2025, às 18:27, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Quesler Fagundes Camargos**, **Usuário Externo**, em 24/08/2025, às 21:36, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **David Alberto Seth Langa**, **Usuário Externo**, em 22/09/2025, às 05:03, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Simone Lúcia Guesser**, **Usuário Externo**, em 24/09/2025, às 09:56, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4394249** e o código CRC **EF08679E**.

Referência: Processo nº 23072.244290/2025-10 SEI nº 4394249

This dissertation is dedicated to my advisor, my family, and my boyfriend, for their unvaluable support.

ACKNOWLEDGEMENTS

First of all, I need to thank my advisor, Prof. Fábio Bonfim Duarte, whose dedication, advice, and support was essential to the development of this dissertation.

I also need to thank profa. Dra. Jânia Martins Ramos, for introducing me to the study of formal syntax.

I would also like to thank Prof. David Alberto Seth Langa for his feedback, help with data collection, and invaluable advice.

Moreover, I would like to thank Prof. Aroldo Leal de Andrade for introducing me to the world of the study of information structure, and for his feedback on the earlier versions of this dissertation.

In addition, I would like to thank my family for their unconditional support.

I would like as well to thank my colleagues at Laliafro, who helped me develop my hypothesis and data analysis.

Furthermore, I would like to thank my colleague Kondwani Bewala, who provided the data used in this work. This dissertation would not happen without his contribution.

Finally, I would like to thank my dear Paulo, for helping me develop the analysis, for proofreading the text, and for his unconditional love.

Thank you all so much!

"[...] the faculty of language plays a crucial role in every aspect of human life, thought and interaction. It is largely responsible for the fact that, alone in the biological universe, human beings have a history, a diversity and cultural evolution of some complexity and richness, and even biological success, in the technical sense that their numbers are enormous. A Martian scientist observing the strange occurrences on Earth could hardly fail to be impressed by the emergence and importance of this apparently unique form of intellectual organization. It is all the more natural that the topic, with its many mysteries, has stimulated the curiosity of those seeking to understand their own nature and their place in the wider universe."

CHOMSY, Noam. **Linguagem e mente**. Universidade de Brasília, 1998.

Abstract

This dissertation investigates the syntactic encoding of focus in Cinyanja, a Bantu language spoken in Mozambique and other countries, in the context of generative syntax and information structure. The study addresses a central debate in Bantu linguistics: whether focus is realized in a dedicated syntactic position. The theoretical framework is based on Rizzi's (1997) cartographic approach, which integrates pragmatic features such as focus and topic into syntactic structure, and Aboh's (2010) proposal that information structure originates in the numeration. The study employs mixed methods: corpus analysis of elicited data (question-answer pairs, felicity and grammaticality tests), and theoretical analysis within the principles of the Minimalist Program and Distributed Morphology. Based on the data analyzed, I argue that Cinyanja marks focus through cleft and pseudocleft constructions, in which the focused constituent appears immediately after a copular verb 'ndí'. I propose that the functional head Focus is mapped between InflP and VoiceP, i.e. a low focus projection, in the matrix clause of the cleft / pseudocleft. The argument in focus assimilates the high floating tone of the copula that precedes it and is flanked by a phonological boundary. In this way, my analysis is in line with Zubizarreta's (2010) proposal. The empirical evidence includes: (i) the optionality of the copula 'ndí' and the complementizer 'méne' in focus constructions; (ii) the agreement between the focus constituent and the overtly realized complementizer 'méne'; and (iii) prosodic marking (assignment of high tone) in focused arguments. The results contribute to broader discussions on the syntax-pragmatics interface and Bantu linguistic typology, highlighting the role of prosody and syntax in the realization of focus.

Keywords: Information Structure; Argument Structure; Focus, Formal Syntax, Bantu Languages.

Resumo

Esta dissertação investiga a codificação sintática do foco em Cinyanja, uma língua Bantu falada em Moçambique e outros países, no contexto da sintaxe gerativa e da estrutura da informação. O estudo aborda um debate central na linguística Bantu: se o foco é realizado em uma posição sintática dedicada. A fundamentação teórica se baseia na abordagem cartográfica de Rizzi (1997), que integra características pragmáticas como foco e tópico à estrutura sintática, e na proposta de Aboh (2010) de que a estrutura da informação se origina na numeração. O estudo emprega métodos mistos: análise de corpus de dados elicitados (pares de pergunta-resposta, testes de felicidade e gramaticalidade), e análise teórica dentro dos princípios do Programa Minimalista e da Morfologia Distribuída. Com base nos dados analisados, defendo que o Cinyanja marca o foco por meio de construções de clivada e pseudoclivada, em que o constituinte focalizado aparece imediatamente após um verbo copular 'ndí'. Proponho que o núcleo funcional Foco é mapeado entre InflP e VoiceP, isto é, uma projeção de foco baixa, na oração matriz da clivada/pseudoclivada. O argumento em foco assimila o tom flutuante alto da cópula que o precede e é flanqueado por uma fronteira fonológica. Desse modo, minha análise se alinha com a proposta de Zubizarreta (2010). As evidências empíricas incluem: (i) a opcionalidade da cópula 'ndí' e do complementizador 'méne' em construções de foco; (ii) a concordância entre o constituinte em foco e o complementizador 'méne' realizado segmentalmente; e (iii) marcação prosódica (atribuição de tom alto) em argumentos focalizados. Os resultados contribuem para discussões mais amplas sobre a interface sintaxe-pragmática e a tipologia linguística Bantu, destacando a função da prosódia e da sintaxe na realização do foco.

Palavras-chave: Estrutura Informacional; Estrutura Argumental; Foco; Sintaxe Formal; Línguas Bantu.

ABBREVIATIONS

ACC	Accusative
APPL	Applicative
ASP	Aspect
CAUS	Causative
CL	Clitic
COMP	Complementizer
COP	Copula
CP	Complementizer Phrase
DAT	Dative
DET	Determinant
DP	Determinant Phrase
EMPH	Emphatic
EPP	Extended Projection Principle
F	Feminine
Foc	Focus
FocP	Focus Phrase
FUT	Future
FV	Final Vowel
IAV	Immediately After the Verb
InflP	Inflection Phrase
INTER	interrogative
LAD	Language Acquisition Device
LOC	Locative
NFOC	Non-focus
NOM	Nominative

NPA	Nuclear Pitch Accent
NUM	Number
OCP	Obligatory Contour Principle
OM	Object Marker
p-boundary	Phonological boundary
p-phrase	Phonological phrase
PASS	Passive
PAST	Past tense
PERF	Perfective Aspect
PL	Plural
PRES	Present tense
pro	empty category with Case
PRO	empty category without Case
PRS	Present Tense
PST	Past Tense
SG	Singular
SM	Subject Marker
SPEC	Specifier
SUBJ	Subjunctive
TAM	Tense, Mood and Aspect
Top	Topic
UG	Universal Grammar
UTAH	Uniform Theta-Assignment Hypothesis
VoiceP	Voice Phrase
VP	Verb Phrase
XP	any given phrase/projection

LIST OF TABLES AND FIGURES

SUMMARY

CHAPTER 1

INTRODUCTION

This dissertation aims to investigate the focus position in the Nyanja language (Bantu family). The reason for the choice of this phenomenon has to do with the fact that there is a big debate in Bantu literature whether the focus constituent occupies a dedicated syntactic position in the structure of a given pragmatically structured proposition. The hypothesis I am investigating in this dissertation can be formulated as follows: prosodically, the focused constituent bears a high floating tone and is flanked by a phonological boundary (p-boundary); syntactically, Cinyanja uniformly encodes constituent focus via movement of the focused DP into the specifier of a low Focus projection, rendering a cleft or pseudocleft construction. The evidence for my proposal comes from five facts, namely:

(i) the copula *ndí* bears a high tone which spreads to the focused argument;

(ii) the copula *ndí* may optionally occur immediately before the focused constituent;

(iii) the focused DP bears a high floating tone even when the copula is not segmentally realized;

(iv) the complementizer *méne* 'that' is also optional in clefts and pseudoclefts;

(v) when overt, the complementizer enters an agreement relation with the focused constituent.

The question about how arguments are introduced in a sentence has been investigated since the inception of the generative program. A few seminal works should be mentioned: Marantz's *On the Nature of Grammatical Relations* (1984), Baker's *Incorporation* (1988), Levin and Rappaport Hovav's *Argument Realization* (2005), and Pylkkänen's *Introducing Arguments* (2002). All the works above delineate how syntax and semantics interact to allow for the derivation of sentences and are still influential in generative literature.

There is, however, evidence that argument structure is deeply related to information structure. Rizzi (1997) proposes that pragmatic properties should be represented in syntax, since they constitute pieces of information that ought to be part of the derivation. His cartography allows for the mapping of the properties related to the left periphery, that is, pragmatic factors and elements such as clause typing, Force, relative pronouns, and topic and focus dynamics. Rizzi's proposal

of "exploding" the complementizer layer (CP) into different projections of Force, Topic, and Focus makes a number of phenomena related to the C system more amenable to analysis, such as adjacency and non-adjacency effects involving elements of the C system, and the different kinds of fillers of the subject position (overt DP, PRO, pro, trace/copy). Rizzi proposes that Ā-movement (concerning the CP system) is driven by Criteria, that is, features that have interpretive import (Wh, Negation, Focus, Topic, …) and determine the interpretation of the category bearing them and of its immediate constituents.

Revisiting Rizzi, Aboh (2010) proposes that "information structure starts at the numeration". Based on analysis of a number of languages, especially Gungbe, and a range of phenomena such as question-and-answer pairs and overt particles that encode Topic and Focus functions, Aboh concludes that information structure starts at the numeration in the form of discourse-related lexical items which drive the derivation. These discourse-related particles, which encode features such as Force, Topic, and Focus, project in syntax. These features are comparable to other optional formal features (such as Case features) that are added arbitrarily when the lexical item enters the numeration (2010:12). The author proposes the following postulate (2010:14):

(1) A numeration N pre-determines the Information Structure of a linguistic expression.

Under such a view, “information structure and syntax interact indirectly (via PF and presumably LF)” (2010:16). The author argues that “the choice of a linguistic expression is sensitive to discourse considerations external to syntax. But what matters for the present discussion is that whatever linguistic expression is chosen by the speaker has very specific syntactic properties” (2010:19).

Based on the discussions outlined above, the motivation for the current dissertation is to investigate how focused arguments are introduced in syntax and to unify the syntactic and pragmatic factors involved in the derivation of a sentence. That is, this research has the objective to investigate how focused arguments are introduced and how this is related to information structure factors. In the following sections the aim is to address the problem at hand, the assumed hypothesis, the research questions, and the methodology used in this dissertation.

This chapter is organized as follows: section 1.1 presents the problem at hand. Section 1.2 shows the research questions and hypothesis. Section 1.3 provides the methodology employed in this work. Section 1.4 deals with the structure of this dissertation.

In the next subsection I will present the problem at hand.

1.1: THE PROBLEM AT HAND

In this section I will discuss the relation between argument structure and information structure.

I assume in this dissertation the idea that language is a manner of thought structuring (similarly to what has been advanced by Fodor), and language output will reflect how the speaker organizes the information/idea that will be conveyed. This should happen during the Selection and Numeration steps of derivation, following the idea of Aboh (2010) that information structure features are already present in the lexicon through bundles of features that will enter the derivation along other necessary features for the linguistic representation to be constructed. Therefore, argument structure alternations would consist in different ways of conveying information according to the speaker's intention and conceptualization of the truth-value of the clause.

Consider the following data:

(2) ENGLISH

(2a) John broke the vase with a hammer.
(2b) John broke the vase.
(2c) The hammer broke the vase.

(2d) The vase was broken (by John) (by the hammer).
(it is not possible to add both "by-phrases" at the same time)
(2e) The vase broke.

Almost all the sentences in (2) above convey the same truth-value interpretation: there is an object "vase" that was subjected to a change of state and became "broken", either by the action of John and/or by the hammer, except in (2e), in which there is no cause involved in conveying the information that the object underwent a change of state. There are, however, differences among all the examples: along some changes in grammatical functions, depending on the structure, the informational status of the arguments involved in the event described by the verb changes from sentence to sentence. The phrasings used in (2a) and (2b) give discursive salience to the action of the causer of the event and will be used in specific contexts. (2c), on the other hand, emphasizes the instrument and will be used in other contexts. (2d) and (2e) focus on the change of state and its results, and the patient is more salient.

The literature on informational structure discusses two pragmatic functions which are performed by the arguments in a given pragmatically structured proposition: Topic and Focus. In the present work I will deal with the projection of Focus. According to Lambrecht (1994), this pragmatic function is associated with signaling a portion of the proposition that is not pragmatically recoverable in the common ground. Thus, focus is the semantic component by

which the assertion differs from the presupposition. Focus is also related to the idea of "a set of alternatives that are relevant for the interpretation of a linguistic expression" (Krifka and Musan 2012). The literature recognizes different types of focus constructions and domains. The notion of Focus is presented in detail in section 3.5.2.

In this dissertation I will discuss the realization of focused arguments (identificational and informational constituent focus) in the language Cinyanja, a Bantu language spoken in Mozambique and other countries in southeastern Africa. Typologically, in Bantu languages there are four possible syntactic positions for the projection of the formal Focus category: sentence-initially, pre-verbal, post-verbal, and sentence-finally (Gibson et al 2017, cf. section 3.5.3). I will investigate in which position the focused arguments in Cinyanja are realized syntactically, that is, how this language fits in the typology proposed by the authors. My proposal is that, in Cinyanja, focused arguments occur post-verbally, that is, after a copular verb. This copula is manifested prosodically (high floating tone and p-boundary insertion), and segmentally in some constructions. Hence, this language behaves just as noted by Zubizarreta (2010; cf. section 3.5.4).

The next section presents the research questions and hypothesis assumed in this work.

1.2: RESEARCH QUESTIONS AND HYPOTHESIS

The following questions will guide the work developed in this dissertation:

i. How to map the position of the focus projection in a sentence which encodes a pragmatically structured proposition in Cinyanja?
ii. How to represent the interaction between syntax and information structure expression in Cinyanja?
iii. To what extent do nonsemantic factors such as information structure and heaviness govern argument realization?
iv. What are the prosodic and syntactic properties of focused constituents of Cinyanja in the framework of Zubizarreta (2010)?
v. How to assess the way Cinyanja fits in the typology of focus expression in Bantu languages proposed by Gibson et al (2017)?

The hypothesis advanced in this dissertation is that, in Cinyanja, focused arguments are syntactically realized immediately after a copular verb, rendering the immediately after the verb (IAV) position. Although the focused argument is realized in sentence-initial, sentence-medial, or sentence-final position, there

seems to be a residual high floating tone that comes from a grammaticalized copula that is not segmentally realized. There also occurs the insertion of a p-boundary flanking the focused argument. Thus, my hypothesis is that Cinyanja conforms to Zubizarreta's (2010) proposal regarding the prosodic marking of focused arguments.

In the next section I will discuss the methodology employed in this dissertation.

1.3: METHODOLOGY

The aim of this section is to describe the methodology employed in this dissertation. I will use an empirical approach to data collection and analysis, that is, I will gather, describe and explain a range of data from the Nyanja language as spoken in Mozambique. The reason I am investigating the Mozambique variety is due to an ongoing cooperation project between the Universidade Federal de Minas Gerais in Brazil and the Universidade Eduardo Mondlane in Mozambique. This is justified by the nature of the investigation I am developing in this work. The goal of the methodology employed here consists in a tool to describe the interaction between information structure and argument structure.

There is a discussion in the field work on semantics (and by extension, formal syntax) about the experimental character of the data collection (cf. Matthewson 2004, among others). The non-experimental approach aims to gather natural data. However, as Matthewson points out (2004:371), "negative evidence […] is crucially required" and thus a more experimental approach is considered to gather more accurate data. In the present work I pursued an experimental approach to data collection, by using imaginary discourse contexts, question-and-answer pairs, and questions such as "can this structure be used in this specific context?". The goal of the question-and-answer pair method was to generate a database of sentences that encode a pragmatically structured proposition, and to pair these sentences with their context of realization. To achieve this, I designed a set of questions to be translated and answered by two native speakers of Cinyanja from Mozambique. One of these speakers (who is a language teacher and researcher) also provided felicity conditions and grammaticality tests. These sets of data were constructed using the VerboWeb (Cançado et al 2017) database, which organizes and describes a range of verbs according to their syntactic and semantic properties. The questionnaire was elaborated in Portuguese in order to facilitate the collection of the data. I was thus able to build a database of several different constructions which occur in the language in question. This preliminary instrument is presented in the Annex 1 of this dissertation. Other contexts and

sentences were tested after the use of the questionnaire. Based on the gathered data, I will investigate the validity of my research hypothesis.

Regarding the presentation of aspects of the Nyanja language, I chose authors such as Ngunga (2004), Macalane (2013), and Biriate (2023), among others, who provide a good perspective and description of the language in question.

As for the literature review and the theoretical framework, the present dissertation builds on top of a range of works, including works within Chomsky's Generative Program of research, on informational structure, and on argument structure. The aim of this review was to provide the background for the description and analyses I propose in this thesis. As for the informational aspects of the phenomena I am investigating, I chose the authors which seem to be more relevant in the field, such as Prince (1981) and Lambrecht (1994), among others.

Concerning the morphosyntactic aspects, my choice of theoretical framework, namely Chomsky's Generative Research Program and Hale and Marantz's Distributed Morphology, is due to my wish to present an elegant analysis of a formalist character. Taking the Occam's Razor principle into account, I assume that "the principle gives precedence to simplicity: of two competing theories, the simpler explanation of an entity is to be preferred" (Duignan 2024), and thus the Minimalist Program and Distributed Morphology

approaches seem adequate to fulfill my goal of describing and explaining the phenomena in question, and to provide the best answers to the research questions I am investigating.

The methods of data collection are summarized below.

i. **Literature review**: I read and cited a range of works on the Nyanja language, information structure, argument structure, the Minimalist Program, and Distributed Morphology.

ii. **Question-and-answer pairs**: I presented a few situations to two native speakers and asked them questions about the participants of the event described in those situations.

iii. **Felicity conditions of sentences**: I asked a native speaker about the answers that were more adequate to the questions and context of utterance.

iv. **Grammaticality tests**: I asked a native speaker to judge if a given sentence is grammatical or ungrammatical.

Thus, based on the considerations above, I developed my research by using an experimental and empirical approach. See chapter 2 for the presentation

of aspects of the Nyanja language, and chapter 3 for the presentation of the theoretical framework assumed in this dissertation.

1.4: STRUCTURE OF THE DISSERTATION

This dissertation is organized as follows: **Chapter 1** comprised the introduction, namely the problem at hand, the research questions, the assumed hypothesis, and the methodology used in this work. **Chapter 2** presents a brief overview of the language analyzed in this work, namely the Bantu language Nyanja. **Chapter 3** presents the theoretical framework assumed in this dissertation. **Chapter 4** presents the discussion of data on the realization of focused arguments according to their informational status. **Chapter 5** presents the theoretical proposal, and the syntactic and pragmatic analysis of data related to argument realization and information structuring. **Chapter 6** concludes with the final remarks. In the next chapter I will present the Nyanja language.

CHAPTER 2

THE NYANJA LANGUAGE

Cinyanja[1] is a Bantu language, N.31 in Guthrie's classification (1967-71). The term “Bantu” refers to the notion of “people” and “human beings” and was first proposed by Bleek (1869) to designate this group of languages that share many properties.

The Nyanja language is spoken in three provinces of Mozambique, namely: Niassa (in the city of Lichinga and the districts of Cuamba, Ngaúma, Mecanhelas, Mandimba and Lago); Zambézia (in the districts of Dere, Lugela, Milange, Mocuba, Molumbo and Morrumbala); and Tete (in the city of Tete and the districts of Angónia, Chiúta, Chifunde, Doa, Mágoe, Macanga, Marávia, Moatize, Mutarara, Tsangano and Zumbo). According to Ngunga et al (2023), Cinyanja is also spoken in the republics of Malawi and Zambia, where it enjoys the status of a “national” language in the former case and “one of the seven national languages” in the latter case. The following variants are also part of the sub-family Chewa-Nyanja: Cicewa (Chichewa) and Cimang'anga (Cingoni).

It is worth remarking that, although Nyanja and Cicewa are part of the same sub-family, there are many differences between the two variants. While

[1] The prefix {ci-} means “language”.

Cicewa is widely studied, Cinyanja still needs to be more documented and analyzed. I intend to contribute to the knowledge of this language with the present work.

The classification of this language can be illustrated as follows.

- Niger-Congo
    - Atlantic-Congo
        - Volta-Congo
            - Benue-Congo
                - Bantoid
                    - Southern Bantoid
                        - Bantu
                            - Nyasa
                                - Chewa-Nyanja

The map below illustrates the geographical distribution of the Nyanja language.

Figure 1: Geographical distribution of Cinyanja (Source: Ethnologue)

According to Biriate (2023:14-15), the Nyanja language was introduced in the formal education system of Mozambique in 1993 as an experiment, and definitely in 2004. Over ten-thousand students are studying the language, as seen in Minedh (2017).

Cinyanja presents very interesting phenomena, such as a rich agreement system, 18 noun classes, an intricate verbal complex, the so-called verbal extensions, and a dual tone system.

This chapter is organized as follows: section 2.1 presents a brief account of the phonology of the language, that is, the consonant and vowel inventories, as well as phonological processes and an account of the tone system. Section 2.2 provides some facts on morphology. Section 2.3 shows some aspects of the syntax of the language. Section 2.4 deals with the verbal extensions.

2.1: BRIEF PRESENTATION OF THE CINYANJA PHONOLOGY

In this section I will present the vowel and consonant inventories, as well as the phonological processes which involve these phonemes. The next subsection will describe the vowel system.

2.1.1: THE VOWEL SYSTEM OF CINYANJA

The vowel system of the Nyanja language consists of five contrastive vowel phonemes: /a ɛ i ɔ u/, written as <a e i o u>. Long vowels occur in some words and the context of penultimate syllable in a phonological phrase. The table below illustrates the structure of the vowel system.

	front	central	back
close	i		u
open-mid	ɛ		ɔ
open		a	

Table 1: Cinyanja vowel system (source: Ngunga and Faqir (2011)

Macalane (2013) describes the following phonological processes that involve vowels: semivocalization, elision, crasis, insertion, and vowel harmony. Regarding semivocalization, Macalane (2013:14-15) describes this process as follows: a high vowel /i/ and /u/ becomes a semivowel (also known as glide) when it precedes another vowel, as seen below.

(3) mwana (> mu-ana) 'child'
myala (> mi-ala) 'stones'

This process is illustrated by the phonological rule below. This rule changes high vowels /i/ and /u/ respectively into the semivowels /y/ and /w/ in order to avoid hiatus. The difference between a vowel and a semivowel has to do with phonotactics[2].

[2] In many Bantu languages the syllable structure consists in -CV(N)C- (Meeussen 1967). Any structure that breaks out of this configuration is resolved through phonological processes such as the ones I present in this section. Additionally, in many languages, glides function phonologically as consonants, even though they are pronounced as vowels phonetically. These facts seem to apply to Cinyanja.

(4) V[+high] → G/_V

Macalane (2013:15-16) also presents the elision process, in which a low or high vowel is erased when it occurs immediately after another vowel.

(5)	phoka	(> pa-okha)	‘isolated place’
	cala	(> ci-ala)	‘finger’
	maaso	(> madiso > maiso)	‘frog’

This process may be described by the rule below. This rule erases low and high vowels which occur immediately before another vowel. This process is somewhat similar to the semivocalization, as it applies in the same context.

(6) V[+low/+high; rec] → ø/-V

Another process described by Macalane (2013:17) consists in the crasis, that is, the amalgamation of two vowels which results in a third different vowel. In the examples below, the conjunction of a low vowel with a high vowel results in a mid vowel.

(7)	comwela	(> ca kumwela > ca+umwela)	‘tool for drinking’
	podyela	(> pa kudyela > pa+udyela)	‘cafeteria’

According to Macalane (2013:18), the phenomenon of glide insertion occurs when a new segment is realized in a position which is not formally

occupied inside the word or morpheme. This process may be applied in initial, medial or final position in the word. In Cinyanja, this process consists in the insertion of a glide[3], namely /w/ or /y/, in intervocalic position, as seen in the data below.

(8)	mayiko	(> maiko > ma-dziko)	‘days’
	awo	(>ao)	‘those’

According to Macalane (2013:18-19), vowel harmony is a process in which two vowels assimilate at a distance. Thus, the occurrence of a given vowel is conditioned by the quality of the vowel in the syllable that precedes it. Observe the situation in the examples below, in which a stem vowel determines the quality of the vowel in the applicative extension (cf. section 2.4).

(9)	-yika	‘put’	-yik**il**a	‘to keep’
	-ponya	‘to shoot’	-pony**el**a	‘to shoot with/to’

The data above suggests that the vowel of the applicative extension may occur as [+high] or [-high, -low], given that the vowel harmony rule is obligatory. Macalane (2013:20) illustrates this rule with the triangle below.

[3] Recall that glides function phonologically as consonants and thus, the syllable structure is preserved (cf. footnote 2).

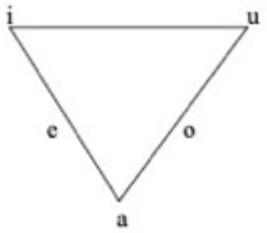
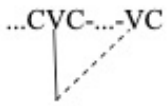


Figure 2: Structure of the vowel system of Cinyanja (source: Macalane (2013:20)

The vowel harmony rule applying to the word *-ponyela* 'to shoot with/to' in (9) may be described as follows. The filled line indicates the features of the vowel in the leftmost syllable, namely /o/, and the dashed line indicates that these features are being copied to the vowel on the right, i.e. /e/. Thus, the quality of the applicative extension vowel is determined by the influence of the stem vowel.

Vowel harmony:



[-high, -low]

The next section presents an account of the consonant inventory of Cinyanja, and the main phonological processes that may involve the combination of these segments.

2.1.2: THE CONSONANT SYSTEM OF CINYANJA

The consonant system of Cinyanja comprises plain, labialized, and palatalized consonants. Labialized consonants are written with <w> and palatalized consonants are written with <y>. Voiced, aspirated and prenasalized consonants also occur. Aspirated consonants are written with <h>, and prenasalized are written with <n>. The consonant inventory is presented below.

manner/point	Bilabial	Labio-dental	Alveolar	Palatal	Velar	Glotal
OCLUSIVE						
Non-voiced	p	-	t	-	k	-
Voiced	bh	-	dh	-	g	-
IMPLOSIVE						
Voiced	b	-	d	-	-	-
NASAL						
Voiced	m	-	n	ny	ng'	-
FRICATIVE						
Non-voiced	-	f	s	sh	-	h
Voiced	-	v	z	-	-	-
AFFRICATE						
Non-voiced	ps	pf	ts	c	-	-
Voiced	bz	bv	dz	j	-	-

LATERAL	-	-	l	-	-	-
TRILL	-	-	r	-	-	-
SEMIVOWEL	w	-	-	y	-	-

Table 2: consonant inventory of Cinyanja/Chichewa (source: Sitoe and Ngunga (2000:86))

Macalane (2013) describes some phonological rules that involve consonants, namely: pre-nasalization, elision, and assimilation.

Regarding pre-nasalization, Macalane (2013:23) defines it as a rule in which the nasal consonant assimilates the articulation point of the following consonant, and both consonants are pronounced at the same time. This process is similar to vowel harmony, in the sense that the leftmost consonant copies the features of the consonant that follows it. This process may be illustrated as indicated below.

(10) N+[b, d, v, g] = mb, nd, mv, ng

The examples below illustrate the occurrence of pre-nasalization in the following words:

(11) **mb**uzi ‘kid’
mvula ‘rain’
ndevu ‘beard’
ngozi ‘accident’

As for elision, Macalane (2013:24) describes this process as being the erasing of the consonants /k/, /d/ and /dz/ when the prefix of the classes 6 or 16, respectively {pa-} and {ma-}, precedes them. This rule can be stated as indicated below: the non-nasal, non-continuous consonant, in the context of being preceded by the noun prefixes of class 6 or 16 and occurring between two vowels, is deleted.

(12) C[-cont, -nas] → ø /V_V

The examples below illustrate this process.

(13)	potalika	'far place'	(cf. pa utalika < pa (**k**)utalika)
	maaso	'eyes'	(cf. ma iso < ma (**d**)isso)
	mayiko	'countries'	(cf. ma (**dz**)iko)

Concerning assimilation, Macalane (2013:25-26) notes that this process occurs when nasal consonants assimilate the point of articulation of the following consonant. Consider the examples below, in which the consonant of the class noun prefix copies the features of the stem consonant and thus change their articulation.

(14)	m'phunzitsi	(>m(u)phunzitsi)	'professor'
	n'tengo	(>mutengo)	'price'
	nkazi	(>mikazi)	'woman'

In this subsection I presented an account of the segmental phonology of Cinyanja. The next subsection provides a brief overview of the tone system of the language.

2.1.3: BRIEF ACCOUNT OF THE TONE SYSTEM OF CINYANJA

As defined by Pike (1948) *apud* Hyman (1975:213), a tonal language presents a significative, contrastive and relative pitch in each syllable. Thus, in these languages one can find words with the same segments (consonants and vowels) whose meanings are distinguished by the relative pitch of some syllables. This distinction can be lexical or grammatical, that is, it can distinguish between words, or between grammatical properties such as tense-mood-aspect. Lexical tones belong to the dictionary, according to Ngunga (2004:90). On the other hand, grammatical tones belong to the grammar of the language and not in the dictionary.

According to Downing and Mtenje (2017:110), typical characteristics of Bantu tone systems are comprised as follows:

i. They are two-tone systems, contrasting high vs low (or ø) tone.

ii. Tone is contrastive for all lexical categories – e.g., noun, verb, adjective – as well as for many functional morphemes.

iii. The application of various tone processes – especially those leading to what Odden (2015) calls tonal mobility and tone reduction, can obscure the source of the underlying tone associations of a particular form.

iv. Verb tone patterns are typically the most complex, as they may be conditioned by a combination of factors like the contribution of a lexical (underlying) high tone on the verb stem and the Object Marker, as well as by high tone(s) contributed by particular tense-aspectual-mood (TAM) markers.

Like many Bantu languages, Cinyanja is tonal, meaning that variations in pitch – specifically high and low tones—affect lexical and grammatical distinctions. In this language, there is a contrast between a high tone and a low, underspecified tone, resulting in a high vs. ø tone system.

According to Downing and Mtenje (2017), the low tone does not actively contribute to tonal contrasts in this language, contrary to high tones which may double, plateau, or retract, and trigger tonal processes motivated by the

Obligatory Contour Principle (OCP) (see Odden 1986; Myers 1987, 1997). Observe the following examples of minimal pairs which are distinguished through tones (taken from Macalane (2013:32)). In the examples below there are minimal pairs which are distinguishable through lexical (15) and grammatical (16) tones. In example (15), for instance, the two words *nkángo* 'lion' and *nkango* 'curl' have the same segmental structure, but the presence of the high tone in the former signals a different meaning from the latter. On the other hand, in (16) there are two verbal words with the same morphophonological structure, but the difference in which vowel bears a high tone is reflected in the grammatical aspect of the verb.

(15a) nkángo 'lion'
(15b) nkango 'curl'

(16a) ndímapíta 'I usually go'
(16b) ndimápíta 'I used to usually go'

As for the marking of tone in writing, Biriate (2023:20) only marks low and high tones in minimal pairs in which meaning can only be distinguished by relative pitch differences (Ngunga and Faquir 2011:100 *apud* Biriate 2023). In the present work, tone will be marked with a <'> diacritic in the vowel that bears the high tone. For instance, the minimal pair /té/ and /tè/ will be marked, respectively, as high tone <té> and low tone <te>.

The aim of the next section is to provide a brief account of the nominal and verbal morphology of the language in question.

2.2: BRIEF PRESENTATION OF THE MORPHOLOGY OF CINYANJA

In nominal morphology, we have an invariable nominal theme that is joined to a variable prefix, the latter being responsible for inflection and derivation. Nouns are organized into nominal classes, which consist in “the set of nouns with the same prefix and/or agreement pattern” (Ngunga and Simbine 2012:92). There are 18 classes and their respective prefixes. The classes are organized into singular and plural pairs, the infinitive verb class, and the locative classes. According to van der Wal and Fuchs (2019), nouns in Bantu languages are formed by joining a root with the appropriate class prefix. The authors analyze deverbal nouns, where *n* is a nominalizer, as well as denominative nouns, such as diminutives and locatives. Thus, van der Wal and Fuchs (2019) propose that nominal/gender classes are located in the functional head n, which categorizes roots as nominal. Following Kramer (2014, 2015), the authors (2019:2) assume that there are several "flavors" of n, the number and nature of which depend on how particular languages organize nouns.

The table below illustrates the noun classes.

Class	Prefix	Gender	Semantics	Examples
1	mu- / Ø	singular of 2	people (all), also animals	mu-nthu ‘person’ galu ‘dog’
2	a-	plural of 1		a-nthu ‘people’ a-galu ‘dogs’
3	mu-	singular of 4	plants, body parts, places	m-tengo ‘tree’
4	mi-	plural of 3		mi-tengo ‘trees’
5	di-/dzi- / Ø/	singular of 6	parts of human body and animals, things, places	di-so ‘eye’
6	ma-	plural of 5 and 11		ma-so ‘eyes’
7	ci-	singular of 8	places, things (products, tubers)	ci-manga ‘corn’
8	zi-	plural of 7		zi-manga ‘corns’
9	N-	singular of 10	animals, things	mbelele ‘sheep’
10	N-	plural of 9		mbelele ‘sheep’
11	lu-	singular of 6	long things, places	lu-zi ‘rope’
12	ka-	singular of 13	diminutives	ka-mwana ‘little child’
13	ti-	plural of 12		ti-mwana ‘little children’
14	wu-	substances	massive, abstracts	u-ci ‘honey’
15	ku-	infinitive	----------	ku-lemba ‘write’
16	pa-	locative	situational	pa-munda ‘in the machamba’

17	ku-	locative	directional	ku-munda ‘towards the machamba’
18	mu-	locative	insideness	m'-munda ‘inside the machamba’

Table 3: noun classes and their semantics (source: Biriate 2023:?)

The following table illustrates the prefixes and suffixes of the noun classes described above. Their distribution depends on whether the root corresponds to a verb, an adjective, a possessive pronoun, a demonstrative or a numeral.

class	**prefix / pronoun**	**verb**	**adjective**	**possessive**	**demonstrative**	**Numeral**
1	mu-	w-	w-	w-	-yu	m-
2	a-	a-	o-	a-	-wa	a-
3	mu-	u-	w-	w-	-wu	u-
4	mi-	yi-	y-	y-	-yi	yi-
5	di-/dzi/ ø	li-	lo-	l-	-li	li-
6	ma-	a-	o-	a-	-wa	a-
7	ci-	ci-	c-	c-	-ci	ci-

8	zi-	zi-	z-	z-	-zi	zi-
9	N-	yi-	y-	y-	-yi	yi-
10	N-	zi-	z-	z-	-zi	zi-
11	lu-	li-	lo-	l-	-li	li-
12	ka-	ka-	ko-	ka-	-ka	ka-
13	ti-	ti-	to-	t-	-ti	ti-
14	wu-	wu-	w-	w-	-wu	wu-
15	ku-	ku-	kw-	kw-	-ku	ku-
16	pa-	pa-	po-	pa-	-pa	pa-
17	ku-	ku-	ko-	kw-	-ku	ku-
18	mu-	mu-	mo-	mw-	-mu	m-

Table 4: noun classes prefixes and suffixes (source: Macalane (2013:51)

The following examples illustrate the processes in which the noun classes are involved.

(17) $\sqrt{pusi}$ pusi ‘cat’
a-pusi ‘cats’

(18) nyani yanga ‘my monkey’
a-nyani anga ‘my monkeys’

[Biriate 2023:24]

(19) *Anyani anga ákúba mikanda.*

a-nyani	*anga*	*a-ku-b-a*	*mikanda*
2-my	2monkey	2SM-PRES-steal-FV	4bead

'My monkeys are stealing beads.'

[adapted from Biriate 2023:80]

As for verbs, Cinyanja has a verbal root which is combined with prefixes, suffixes, and the verbal extensions. Verb prefixes encode subject and object agreement, tense/aspect, negation, mood, etc. The order of prefixes is relatively rigid. Verbal extensions, on the other hand, may change the argument structure of the verb. They can be analyzed as verbal heads that were grammaticalized into bound morphemes (Marantz 1984; Williams 1981 *apud* Mchombo 2004). Additionally, the order of the extensions is more flexible. The table below illustrates the structure of a verb in Bantu (adapted from van der Wal 2015).

Table 5: structure of a Bantu verb

NEG	subject	NEG	TAM	object	$\sqrt{root}$	extensions	final suffix	post-final

The data below illustrate a typical sentence in Cinyanja.

(20) *N'nyamata anadya nthoci imodzi dzulo kuphwando.*

n'nyamata	*a-na-dy-a*	*nthoci*	*imodzi*
1boy	1SM-PAST-eat-FV	9banana	NUM-one
dzulo	*ku-phwando*		
yesterday	17LOC-party		

'The boy ate one banana yesterday at the party.'

[Biriate 2023:39]

In this section I presented a description of noun and verbal morphology of Cinyanja. The goal of the next section is to provide a brief account of the syntax of this language.

2.3: BRIEF PRESENTATION OF THE SYNTAX OF CINYANJA

Regarding its structural sentential organization, Ciyanja follows the general Bantu syntactic patterns, including agreement patterns related to noun classes and the canonical SVOX basic word order. In a simple transitive sentence, the verb follows the subject, and the object follows the verb.

(21) *Abambo anayipha mbudzi.*

Abambo	*a-na-yi-ph-a*	*mbudzi*
1dad	1SM-PAST-9OM-kill-FV	9kid[4]

'Dad killed the kid.'

[Biriate 2023:40]

[4] Kid is the English word for a young goat in this context.

The object must be adjacent to the verb. On the other hand, the subject is not fixed in the position immediately before the verb, as it may occur postposed. However, the subject cannot be realized between the verb and the object. See the examples below.

(22a) *Anayipha mbudzi abambo*[5].

a-na-yi-ph-a	*mbudzi*	*abambo*
1SM-PAST-9OM-kill-FV	9kid	1dad

'Killed the kid dad' (lit.)

[Biriate 2023:40]

(22b) * *Anayipha abambo mbudzi.*

a-na-yi-ph-a	*abambo*	*mbudzi*
1SM-PAST-9OM-kill-FV	1dad	9kid

'Killed the dad the kid.'

Cinyanja speakers employ a subject (SM) and an object (OM) marker. Regarding the subject marker (SM), Mchombo[6] (2004:23) notes that this morpheme must obligatorily appear in the initial position of the verb structure. The author argues that, depending on the context, it may function as an agreement

[5] As I will argue in chapters 4 and 5, this construction may correspond to a "covert" pseudocleft construction, depending on the context of utterance.

[6] Mchombo (2004) considers the Cewa variant in his analysis. Since the facts are generally the same in Cinyanja I will extend his proposals to the Mozambique variant.

marker or as an incorporated pronoun. The pronominal analysis is suggested by the fact that, when the SM is present, the subject DP can be erased. See the example below.

(23) *Ikuzísăka*.

i-ku-zí-săk-a
3PL.SM-PRES-3PL.OM-hunt-FV
'They are hunting them.'
[adapted from Mchombo 2004:20]

More evidence of the pronominal status of the SM comes from anaphoric agreement with DPs in other clauses, that is, the marker may enter a non-local agreement relation. In example (24), the fact that the verb *iphe* 'to kill' is construed as having *mikángo* 'lions' as its subject derives from the anaphorical relation between the SM and this argument, since this noun is in a different clause. The example below illustrates this phenomenon. Note that the SM in the structure of the verb *iphe* 'should kill' in the relative clause is correferent with the subject *mikángo* 'lions' in the matrix clause.

(24) *Mikángo ikudzíwa kuti osaka akufuna kuti iphe mbidzi.*

mikango	*i-ku-dziw-a*	*kuti*	*osaka*
4-lion	4SM-PRES-know-FV	that	2-hunter

a-ku-fun-a	*kuti*	*i-ph-e*	*mbidzi*
2-PRES-want-FV	that	4sm-kill-SUBJ	3-zebra

'The lions know that the hunters want them (the lions) to kill the zebra.' (lit. 'The lions know that the hunters want that they (lions) should kill the zebra'.)

[adapted from Mchombo 2004:23]

Thus, the SM can occur without a pronounced subject DP; it may agree with the argument that functions as the grammatical subject, or it can agree with an antecedent with topic function.

As for the object marker (OM), this marker occurs after the tense affixes and immediately before the verb root. Furthermore, the OM is not obligatory[7]. According to Mchombo (2004:20-23), this marker duplicates the φ-features of the object and works as an incorporated pronoun. This pronominal analysis is suggested by the fact that the full object DP can be erased from the sentence when the OM is present, as well as the possibility of non-local (anaphoric) agreement between the OM and the object DP. Consider the examples below. In (25) the subject and object DPs are not pronounced. On the other hand, in (26) there occurs an anaphoric relation between the OM {*-i-*} and the object DP *mikángó* 'lions' in the relative clause.

(25a) *ti-da-**wa**-dul-a*

[7] However, according to Biriate (2023:49), some ditransitive verbs require OM coreferential with the indirect object.

3PL.SM-PAST-1PL.OM-cut-FV
'we had cut them'

(25b) *ti-da-dul-a*
3PL.SM-PAST-cut-FV
'We had cut'

[Biriate 2023:29]

(26) *Mikángó ikudzĭwa kutí njovu zikufúná kutí anyaní aigúlítsé kwá alenje.*

mikango	*i-ku-dziw-a*	*kuti*	*njovu*
4-lions	4SM-PRES-know-FV	that	10-elephants

zi-ku-fun-a	*kuti*	*anyani*	*a-i-gul-its-e*
10SM-PRES-want-FV	that	2-baboons	2SM-4OM-sell-SUBJ

kwa	*alenje*
to	2-hunters

'The lions know that the elephants want the baboons to sell them (the lions) to the hunters.'

[Mchombo 2004:21]

Concerning the double object constructions, Biriate (2023:43) argues that Cinyanja is a totally asymmetric language. However, Rocha and Bewala (to appear), based on the tests proposed by Bresnan and Moshi (1990), namely passivization, word order, and realization of the OM, propose that the language

is, in fact, partially asymmetric[8]. Examples (27) below illustrate the word order of the objects.

(27a) *Nkángo waphéla osáka mbiidzi.*

nkango	*w-a-ph-el-a*	*osaka*	*mbiidzi*
3-lion	3SM-PAST-kill-APPL-FV	2-hunters	9-zebra

'The lion killed for the hunters the zebra.'

(27b) *Nkángo waphéla mbiidzi osáka.*

nkango	*w-a-ph-el-a*	*mbiidzi*	*osaka*
3-lion	3SM-PAST-kill-APPL-FV	9-zebra	2-hunters

'The lion killed the zebra for the hunters.'

[Rocha and Bewala, to appear]

In addition, examples (28) below illustrate that the direct and applied objects can both become the subject of a passivized verb.

(28a) *Osáka, aphélédwa mbiidzi, ndi nkángo.*

osaka)	*a-ph-el-edw-a*	*mbiidzi)*	*ndi*
2-hunter	2SM-kill-APPL-PASS-FV	9-zebra	by

nkango
3-lion
'The hunters were killed the zebra for by the lion.'

[8] The authors show that the position immediately after the verb and movement to subject position of passive verbs are both available to the two objects. However, the OM can only agree with the indirect/applied argument.

(28b) *Mbiidzi, yaphédwéla osáka, ndi nkángo.*

mbiidzi)	*y-a-ph-edw-el-a*	*osaka)*	*ndi*
9-zebra	9SM-PAST-kill-PASS-APPL-FV	2-hunter	by

nkango
3-lion
'The zebra was killed for the hunters by the lion.'
[Rocha and Bewala, to appear]

On the other hand, Rocha and Bewala show that the OM can only agree with the indirect/applied argument, as seen in the examples below.

(29a) *Nkángo waáphéla mbiidzi (osáka).*

nkango	*w-a-a-ph-el-a*	*mbiidzi*
3-lion	3SM-PAST-2OM-kill-APPL-FV	9-zebra

(osaka)
2-hunter
'The lion killed the zebra for them (the hunters).'

(29b) *??Nkángo wazíphéla (mbiidzi) osáka.*

nkango	*w-a-zi-ph-el-a*	*(mbiidzi)*
3-lion	3SM-PAST-9OM-kill-APPL-FV	9-zebra

osaka
2-hunter
'The lion killed it (the zebra) for the hunters.'

In this section, a brief account of the syntax of Cinyanja is presented. The next section will deal with the verbal extensions.

2.4: BRIEF PRESENTATION OF THE VERBAL EXTENSIONS

Cinyanja, along other Bantu languages, makes use of the so-called verbal extensions, morphemes that may produce changes in the argument structure of the verb. These extensions can be categorically classified as verbs, that is, they are heads that may introduce arguments and compound meanings. Mchombo (2004:70) notes that verbal extensions in Bantu originated as verbs which participated in serial constructions, common in Kwa languages of West Africa. In Bantu, these verbs went through morphologization and became affixes. These verbal suffixes change the semantic interpretation of the root of the verb it is attached to; in addition, the extensions may affect the number of DPs that the verb can bear in the syntactic configuration.

Nine different verbal extensions occur in Cinyanja, according to Ngunga (2004) and Macalane (2013), as follows. Note the allomorphic variation in the expression of the extensions, due to vowel harmony effects.

(i)	passive	-idw/edw-
(ii)	applicative	-il/-el/-ir/-er
(iii)	pseudo-passive/stative	-ik/-ek
(iv)	causative	-its/-ets/-y
(v)	intensive	-its(its)-/-ets(ets)-

(vi)	reversive	-ul/-ol
(vii)	perfective	-ilil/-elel
(viii)	reciprocal	-an/-an(an)-
(ix)	associative	-an

The extensions can be subdivided into three groups (Mchombo 2004:75):

(i) those that increase by one the number of DPs that can appear in the sentence
(ii) those that reduce by a corresponding amount the number of DPs the suffixed or extended verb can support
(iii) those that do not alter the array of DPs

Guthrie (1962) classifies the extensions as O+, O- and neutral (O: object). Typical instances of O+ extensions are the applicative and causative. O- is represented by passive, stative and reciprocal. The reversive constitutes a neutral extension. Thus, when a causative or applicative extension occurs with an intransitive verb it becomes transitive, and when a transitive verb is causativized or applicativized it becomes ditransitive. On the other hand, when a transitive verb receives the passive, stative, or reciprocal extensions it becomes intransitive.

The semantics of the verbal extensions used in Cinyanja are described by Biriate (2023:31-38) as follows.

i. **Passive**: indicates that the subject is the patient of the event.
ii. **Pseudo-passive/stative**: indicates state, impending quality, possibility.

iii. **Causative**: indicates that the subject of the event (causer) makes the object (causee) do something.

iv. **Intensive**: indicates intensity or force with which the event is realized.

v. **Reversive**: indicates the opposite interpretation of the non-extended verb.

vi. **Perfective**: indicates perfection, accuracy, finish, taking the action to its end without stopping.

vii. **Reciprocal**: indicates that the event denoted by the verb involves more than one subject which is the object of the action executed by the other and vice-versa.

Thus, some extensions change the argument structure of the verb it is attached to, while others do not alter verbal valency. Some extensions may also change the meaning of the verb, such as the examples below. In (30a) the non-extended verb *yend* means 'to walk'. However, in (30b) the meaning is changed to 'to visit' when the applicative extension appears in the verbal complex.

(30a) *Jowawu wayenda.*

Jowawu	*w-a-yend-a*
1-John	1SM-PAST-walk-FV

'John walked.'

(30b) *Jowawu wayendela Maria.*

Jowawu *w-a-yend-el-a* *Maria*
1-John 1SM-PAST-walk-APPL-FV 1-Mary
'John visited Mary'. (lit: 'John walked for Mary'.)
[adapted from Rocha (214)]

The data above show that the verbal extensions are deeply connected to the meaning of the verb and, consequently, of the sentence in which they occur.

CHAPTER SUMMARY

In this chapter I presented a brief account of the Cinyanja grammar. It addressed facts about its phonology, that is, the phonemic inventories and phonological processes, the tone system, and morphosyntax, especially the noun classes, verb structure, and verbal extensions.

Cinyanja's vowel inventory consists of five vowels that may be subject to lengthening in the penultimate syllable of a phonological phrase. The consonant inventory consists of 25 consonants, including aspirated and prenasalized consonants, and two semivowels.

Next, I noted that Cinyanja has a tonal system based on the opposition between a high tone and a low underspecified tone. Tone in this language can

distinguish between words (lexical tone) and grammatical features such as TAM (grammatical tone).

Concerning the nouns in this language, they are organized into 18 grammatical classes which are partially determined by semantic considerations. On the other hand, verbs are structured in a complex of prefixes, extensions, and suffixes that are related to agreement, negation, tense-mood-aspect (TAM), and argument realization.

Finally, I presented a brief account of the verbal extensions, morphemes that are related to possible changes in argument structure. There are three kinds of extensions in Cinyanja, namely morphemes that increase verbal valency, morphemes that reduce verbal valency, and neutral extensions which do not affect the array of DPs in the argument structure of the verb. Some extensions may also change the meaning of the verb.

In the next chapter I will provide the theoretical framework that will support the theoretical analysis that will be developed in chapters Chapter 4 and Chapter 5.

CHAPTER 3

THEORETICAL FRAMEWORK

The present work is developed in the lines of current generative approaches to natural language, that is, the Minimalist Program and Derivation by Phases. I assume the hypothesis that language is a manner of thought structuring, along the lines of Fodor's ideas, and that language output will reflect how the speaker organizes the information that will be conveyed, according to both universal and language-particular principles. This organization should be determined by the Selection and Numeration steps of derivation, following Aboh's (2010) proposal that information structure features (such as Topic and Focus) are present in the lexicon along all other necessary features, and that they may establish probe-goal relations and project in syntax. Therefore, argument structure alternations are assumed to consist in different ways of conveying information according to the speaker's intention and conceptualization of the truth-value of the sentence.

This chapter is organized as follows: section 3.1 presents an overview of the minimalist program. Section 3.2 shows a brief account of the notions of argument structure and argument realization. Section 3.3 deals with Carstens' proposal regarding the operation AGREE and the EPP-feature in Bantu

languages. Section 3.4 provides a brief account of the definition and properties of cleft constructions. Section 3.5 presents considerations on information structure. Section 3.5.1 presents the notion of information. Section 3.5.2 presents the notion of focus. Section 3.5.3 presents a typology of focus marking in Bantu languages. Section 3.5.4 presents an account of the syntactic and prosodic focus expression in Bantu languages. Section 3.6 presents the relation between information structure and morphosyntax. Section 3.7 presents considerations on the methodology of question-and-answer pairs.

The aim of the next section is to present a brief account of the Minimalist Program.

3.1: THE MINIMALIST PROGRAM AND DISTRIBUTED MORPHOLOGY

This dissertation is based on the theoretical framework of the Minimalist Program (Chomsky 1995), in the Derivation by Phase model (Chomsky 2001, 2004, 2008, 2012). I will present here a brief overview of the assumptions of this theoretical approach.

One of the central aspects of Chomsky's Generative Research Program is the innate character of language, that is, all human beings are endowed with the

capacity for learning a natural language. This genetic endowment is called Universal Grammar (UG), also known as Language Acquisition Device (LAD).

The Principles and Parameters (Chomsky 1981) approach is concerned with the nature of the UG, the initial state of the grammar of language that is present in the mind-brain of all human beings. This language acquisition device allows for the interpretation of input (pieces of language to which the child is presented) and processing into output (the grammar of a particular language and the sentences this grammar generates). This approach understands the UG as organized into universal principles and parameters that will be set as the child acquires the grammar of his/her language. An example of one such principle would be the Principle A of Government and Binding, which states that, in all languages, anaphors must have an antecedent within their domain. An example of a parameter would be the Pro-Drop Parameter, according to which some languages are able to "drop" the subject of a sentence in some contexts, while others cannot and may rely on expletives to fill this syntactic position.

The Minimalist Program simplifies the structure of the UG, by taking a derivational approach – contrary to previous representational approaches. According to this framework, the machinery that pairs form and meaning into utterances consists of only two syntactic levels – the computation (narrow syntax) and the spell-out – the point at which the constructed objects are

transferred to be interpreted at the interfaces: the articulatory-perceptual system – Phonological Form (phonology); and the conceptual-intentional system – Logical Form (compositional semantics). Lexical items – roots and bundles of features (including functional categories) – are drawn from the narrow lexicon and enter the derivation in the Select and Numeration steps. The features can be interpretable or uninterpretable in the interfaces, and those which are uninterpretable must be valued/resolved in probe-goal relationships, via c-command[9] or spec-head relations. Narrow computation – or strict syntax – constructs syntactic objects, relationships, and structure via the operation Merge – a binary operation that takes lexical items (roots and bundles of features) and phrases already constructed, and thus creates new syntactic objects (pairs of items), in such a way that the structure and relations are interpretable at the interfaces.

In order to reduce computational load, the derivation occurs in phases, cyclical stages that are closed and impenetrable (Chomsky 2008), according to the Strict Cycle Principle (Chomsky, 1973). Phases are "propositional/ independent units" that have phonological and semantic correlates, and are defined by Chomsky (2001, 2004) as domains of uninterpretable φ-feature

[9] Constituent command (aka c-command) is defined by Reinhart (1976) as "A commands node B iff the branching node α_1 most immediately dominating A either dominates B or is immediately dominated by a node α_2 which dominates B, and α_2 is of the same category type as α_1" or, in other words, "α c-commands β iff every branching node dominating α dominates β".

valuation. Phases are, then, stages of the derivation in which the features are valued in such a way that the structure can be transferred to the interfaces to be interpreted. Chomsky argues that valuation and deletion of φ-features occur during Transfer (spell-out). According to this view, transfer only targets the complement of the phase, leaving the edge intact and visible for further computation. Chomsky (2008) proposes that v* and C are phases. Other authors, such as Svenonius (2003) and Hiraiwa (2005), argue that DP is also a phase. McGinnis (2001, 2004, 2008) proposes that High Applicative (cf. Pylkkänen 2008) is also a phase. Arad (2003) argues that the first categorization of a root is a (micro) phase.

All operations apply only at the phase level, following the Principle of Earliness (Pesetsky 1989). The information is erased when transferred to the subsequent stages of derivation. Figure 3 illustrates the minimalist model of the computational system.

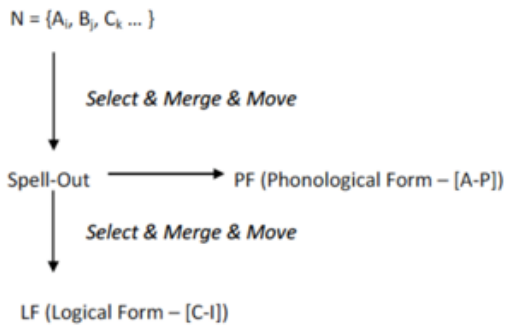


Figure 3: Minimalist model of the computational system

I will also assume the Distributed Morphology theory, as proposed by Hale and Marantz (1993, 1994), among others. This theory proposes that lexical items do not come from the narrow lexicon with phonological features, and that roots are acategorial and become verbs/nouns/ adjectives/etc. when merged into the appropriate syntactic environment. As mentioned above, Arad (2003) argues that the first categorization is a phase, and the resulting syntactic object is sent to spell-out. Phonological features are associated with terminal nodes via correspondence rules in the Vocabulary list, at the Morphological Structure component (associated with Phonological Form), and semantic interpretation occurs at the Logical Form interface taking as input all of the derivation – compositional meaning is interpreted in Logical Form, and special (noncompositional) meaning is interpreted via the Encyclopedia list.

Thus, the argument structure of a sentence like "John broke the vase" would be derived as follows. The steps in (31) show that the Voice phase has unvalued φ-features that need to be valued by a goal. The DP *the vase* is then merged to VP and can function as a goal. The φ-features are inherited by VP to be valued by the goal. When all the features are valued they are deleted, and the complement of the Voice phase is transferred to the interfaces. This model assumes that transitive verbs will require a goal to value the phase φ-features, or

else the derivation will crash. After this phase is closed, the edge is still accessible to the next steps of the derivation: the CP/TP system.

(31a) [$_{VoiceP}$ John *v*φ [$_{VP}$ break [$_{DP}$ the vase]]]

(31b) *v*φ percolates to VP (inheritance)
[$_{VoiceP}$ John *v* [$_{VP}$ breakφ [$_{DP}$ the vase]]]

(31c) deletion – transfer (the φ-features are valued by the goal)
[$_{VoiceP}$ John *v* [$_{VP}$ ~~breakφ~~ [$_{DP}$ ~~the vase~~]]]

The steps in (32) show that the CP/TP system also has φ-features to be valued, and the external argument introduced by Voice, which sits at the edge of the phase, can function as a goal, valuing the φ-features of C/T and, at the same time, it can get its Case features are valued by the probe T. When all the features and criteria (cf. Rizzi 1997) are satisfied/valued, the phase closes and is sent to spell-out. There is no need for C/T to introduce an argument since there is a DP in the edge of the phase available to act as a goal for the C/T probe.

(32a) [$_{CP}$ [+affirmative, + finite] [$_{TP}$ Tφ [+past] [$_{VoiceP}$ [$_{DP}$ John] [$_{VP}$ breakφ [$_{DP}$ ~~the vase~~]]]]]

(32b) [$_{CP}$ [+affirmative, + finite] [$_{TP}$ John broke [+past] [$_{VoiceP}$ ~~John~~ ~~broke~~φ ~~the vase~~]]]

In sum, this dissertation will be developed following the most recent proposals of the phase-based theories, that is, the most recent minimalist approaches to natural language processing. These theoretical assumptions will be particularly useful in Chapter 5, when I propose a derivational approach for dealing with the different word order facts regarding cleft and pseudocleft constructions.

In the next section I will provide a brief account of the notion of argument structure assumed in generative literature.

3.2: ON THE NOTION OF ARGUMENT STRUCTURE AND ARGUMENT REALIZATION

"Argument" is a notion from predicate logic and refers to the item that saturates the meaning of a "predicate" – an unsaturated element. For instance, a verb such as *break* requires two arguments to saturate its meaning: the element that causes the change of state, and the element that undergoes the change of state. In the neutral (presentational or event-reporting) case, the element that triggers the event is encoded as the subject, whereas the element that undergoes the change of state is encoded as the object, as illustrated by the example below.

(33) John broke the vase yesterday.

Notice that, in (33) the predicate *break* is saturated by two arguments, the agent/subject *John* and the patient/object *vase*. The element *yesterday* is circumstantial and is not required by the verb to saturate its meaning, since it is a temporal adverb adjoined in a high domain of the sentence, usually at the C/TP level. Thus, the verb *break* can be analyzed as having two "slots" to be saturated by the core arguments: the agent *John* and the patient *vase*.

Argument structure can be understood, then, as the c-selection and s-selection information contained in the bundle of features that the predicate exhibits. The predicate (generally a verb, but some prepositions, deverbal names, and adjectives) contains lexical-semantic requirements that must be satisfied at the moment of the syntactic derivation. Thus, argument structure is related to the number and semantic properties of the required arguments (Cançado, Godoy and Amaral 2017; Cançado and Amaral 2016). Consider again the verb *break*.

(34a) John broke the vase.
(34b) The vase broke.

The sentences in (34) indicate that there was a change-of-state event in which the referent *vase* changed its state to *broken*. (34a) shows that the event was triggered by the referent *John*, whereas in (34b) there is no triggering event,

only the change-of-state event. The role played by the argument *vase* is the same in both sentences, that is, it performs the same thematic (or semantic) role in both sentences, although its grammatical function is changed from *object* to *subject*. Thematic roles (theta-roles or θ-roles) are semantic functions that a predicate assigns to its arguments (cf. Fillmore 1967, Dowty 1991). The verb *break* can assign two thematic roles: agent (the argument that triggers an event) and patient (the argument that undergoes a change of state). Thus, one way to represent the argument structure of the verb *break* could be stated as follows.

(35) *Break*: {(agent), (patient)}

Notice that the agent can be optional, depending on what kind of information the speaker wishes to convey: which of the subevents of *break* are more discourse salient or in need of activation. In this sense, either the causation event *John causes the vase to break* or the change-of-state event *the vase breaks* can be salient information in a given pragmatic situation.

Hale and Keyser (1993, 2002) argue that semantic roles are assigned structurally, in specific syntactic positions. This proposal conforms to Baker's (1988) proposal of the Universal Theta-Assignment Hypothesis (UTAH), in which identical thematic relationships between items are represented by identical

structural relationships between these items at the deep structure level[10]. The UTAH posits that thematic relations are identical at deep and surface structures.

Argument realization has been heavily investigated in modern syntactic theory. Levin and Rappaport Hovav (2005), for instance, present a comprehensive account of argument realization theories, the thematic hierarchy, and how to map the interaction between lexical semantics and morphosyntax. The authors note that a theory of argument selection must address five major questions (Levin and Rappaport Hovav 2005:3):

(a) Which facets of the meanings of verbs are relevant for the mapping from lexical semantics to syntax?
(b) What is the nature of a lexical semantic representation that encompasses these components of meaning? That is, what are the primitives of this representation and the principles for combining these primitives into representations of specific verb meanings?
(c) What is the nature of the algorithm which derives the syntactic expression of arguments?
(d) To what extent do nonsemantic factors such as information structure and heaviness govern argument realization?
(e) To what extent are the semantic determinants of argument realization lexical and to what extent can some of them be shown to be nonlexical?

[10] In Baker's (1988) framework, syntax comprised three levels of representation: deep structure (in which thematic relations hold), surface structure (related to Case assignment and movement), and the spell-out to the interfaces. The Minimalist Program dispenses with the representational model and the two levels of representation are scrapped in favor of a derivational approach based on the operation *merge*, as sketched in the beginning of the present chapter.

The authors also discuss the challenge of identifying grammatically relevant facets of meaning. In this dissertation I intend to propose an answer to question (d), that is, how the interaction between information structure and morphosyntax determine the realization of focused arguments in Cinyanja.

The present work will assimilate recent approaches to argument structure, such as Kratzer's (1996) proposal of a Voice head, which is the functional head that introduces the external argument. I will also assume Pylkkänen's (2008) proposal, according to which the Voice head introduces the external argument, the Cause head introduces a causation event, and the Applicative head introduces an extra argument that is normally not a part of the verb argument structure. The main current idea on argument structure deals with the fact that arguments are introduced by syntactic heads which project, and theta-assignment occurs structurally, in SPEC-HEAD or C-COMMAND relations. Following Carstens (2005), I will assume that, in Bantu languages, agreement relations occur in spec-head probe-goal operations.

Semantically, predicates such as verbs and prepositions can be analyzed according to three different levels of semantic properties. The first level corresponds to predicate decomposition. According to Levin e Rappaport Hovav (2005), predicate decomposition consists of "a representation of meaning formulated in terms of one or more primitive predicates chosen to represent

components of meaning that reoccur across significant sets of verbs". Abstract semantic predicates, which have become primitive predicates in more current theories, such as CAUSE and BECOME, were first proposed by McCawley (1968b) and Lakoff (1970).

Another type of semantic analysis consists of the thematic roles of the arguments (Fillmore 1967, Dowty 1989, 1991). Thematic roles are semantic functions that the verb establishes with its arguments. This level of analysis takes the sematic relationship between predicates and arguments based on the semantic role the argument plays in the proposition. Thematic roles remain constant in argument structure alternations, that is, there is not a 1:1 mapping between grammatical function and semantic function. However, there is a general tendency to associate roles such as the agent with the subject of the sentence and the patient with the object. This tendency led to the proposals of the so-called thematic hierarchy. Some authors, such as Bresnan and Kanerva (1989:23-24), propose that the thematic hierarchy is actually a topicality scale, since arguments with discourse saliency tend to occupy the subject position. Levin and Rappaport Hovav (2005:170) argue that the thematic hierarchy can be interpreted as a natural prominence scale. The authors remark (2005:172) that "languages vary as to which scales, if any, the surface expression of arguments is sensitive to". For instance, some languages have differential object/subject marking, in which

some arguments will present a morphosyntactic expression according to discourse saliency, animacy hierarchy, grammatical person hierarchy. The proposed hierarchies have empirical consequences, such as the analysis of alternations.

Finally, there is the formal analysis based on neo-davidsonian logic (Kratzer 1996; Heim and Kratzer 1998; Pylkkänen 2008, among others). According to Pylkkänen (2008:5),

> verbs in general will be taken to have neo-davidsonian meanings, where the verb itself names a property of an eventuality (which I take to be a cover term for events and states, following Bach 1981) and the syntactic arguments of the verb name event participants, that is, individuals who stand in thematic relations to the eventuality (Parsons 1990, building on proposals in Castañeda 1967 and Davidson 1967). In this type of framework, the meanings of sentences involve underlying quantification over events.

See the example below from English which presents the formal semantic representation of a transitive verb and the semantic relations of the verb and its arguments.

(36a) Brutus stabbed Caesar.
(36b) (∃e) stabbing(e) & agent(e, Brutus) & theme(e, Caesar)
[Parsons 1990, 97]

The interpretation of this sentence (truth-value) means that there is an event of *stabbing* and this event assign thematic roles to the participants *Brutus* and *Caesar*.

When it comes to argument realization, Kratzer (1996) argues that the external argument is introduced in syntax and semantics via the operation *Event Identification*:

(37) <e,<s,t>> <s,t> → <e <s,t>>

The lambda calculus of this sentence could be represented as follows:

(38a) λe. stabbing(e) & theme(e, Caesar)
(38b) λe. stabbing(e) & agent(e, Brutus) & theme(e, Caesar)
[by Event Identification]

This analysis means that the agent, which is introduced by the syntactic head called Voice, is not a core argument of the verb. Likewise, other non-core arguments such as dative, beneficiary, comitative, instrument, locative, etc., are introduced by *Functional Application* and *Predicate Modification*. The formal definition of these operations can be represented as follows:

(39) Functional Application

> If a is a branching node, {b, g} is the set of a's daughters, and [[b]] is a function whose domain contains [[g]], then [[a]] = [[b]]([[g]])
> [Heim and Kratzer 1998:44]

(40) Predicate Modification

> If a is branching node, {b,g} is the set of a's daughters, and [[b]] and [[g]] are both in $D_{<e,t>}$, then [[a]] = $\lambda x\, e \in D_e$. [[b]](x) = [[g]](x) = 1.
> [Heim and Kratzer 1998, 65]

The operations described above clarify the argument structure of the verb and how the arguments are introduced in a sentence. As seen previously (cf. section 3.5), the truth-value of sentences involved in argument alternations can remain the same, although grammatical functions and informational status will vary. Consider the semantic and informational analysis of the verb *mbil* 'shape' below:

(41) Semantic and informational analysis of $\sqrt{wumb}$ <SHAPE>

(42) *N'zimáyi wawúmbíla mbíya ya mádzi mwána.*

n'zimayi	*wa-wumb-il-a*	*mbiya ya madzi*	*mwana*
1-woman	1SM-shape-APPL-FV	3-pot of water	1-child

'The woman shaped the waterpot for the child.'
[presentational sentence]

(43) *Cacitika ndi ciyani ndi mwana?*

cacitika	*ndi*	*ciyani*	*ndi*	*mwana?*
happen	with	what	with	1-child

‘[$_{foc}$ what] happened to [$_{top}$ the child]?’

[question about a topic]

(44) *Mwana wawumbilidwa mbiya ya madzi.*

mwana	*w-a-wumb-il-idw-a*	*mbiya ya madzi*
1-child	1sm-past-shape-appl-pass-fv	3-pot of water

‘[$_{top}$ the child] [$_{foc}$ was shaped the waterpot for]’

[beneficiary topicalization by passivization]

(45) *Cacitika ndi ciyani ndi mbiya ya madzi?*

cacitika	*ndi*	*ciyani*	*mbiya ya madzi*
happen	with	what	3-waterpot

‘[$_{foc}$ what] happened to [$_{top}$ the waterpot]?’

[question about a topic]

(46) *Mbiya ya madzi yawumbilidwa ndi m’nzimayi.*

mbiya ya madzi	*w-a-wumb-il-idw-a*	*ndi*	*m’nzimayi*
3-pot of water	3sm-past-shape-appl-pass-fv	by	1-woman

‘[$_{top}$ the waterpot] [$_{foc}$ was shaped by the woman]’

[theme topicalization by passivization]

SHAPE = [[X DO <MANNER-SHAPE$_i$>] CAUSE [RESULT-THING$_i$ Y]]

SHAPE = {DP$_{agent}$, DP$_{result}$, DP$_{beneficiary}$}

SHAPE = (∃e) shape(e) & agent(e, m’nzimayi) & theme(e, mbiya ya madzi) & beneficiary(e, mwana)

We can see in the representations above that the verb root $\sqrt{wumb}$ ‘shape’ is a bitransitive verb via applicativization. The agent is introduced by Event

Identification, in the specifier position of VoiceP, the theme is introduced as the complement of v*, and the beneficiary is introduced by Functional Application and the occurrence of the applicative extension on the verb stem. The truth-value of the sentences and the thematic roles assigned to the arguments remain constant despite the changes which occur in grammatical function. The agent argument is introduced by active Voice in active sentences and passive Voice in passive sentences.

However, as noted by Gallego (2015), it is important to point out that the notions of "predicate" and "argument" are intuitive notions from logic and not strictly "linguistic". Therefore, the definition of "argument" and "predicate" will depend on the theory assumed by each author. In the present work, I will follow Gallego in his proposal that (syntactic) "argument" is the noun phrase (DP) that values the uninterpretable features of a phasic head.

Applying the facts to the previous example of the verb *break*, I will assume that the root $\sqrt{break}$ is merged with a light verb v*, which has uninterpretable unvalued φ-features that need to be resolved before spell-out. Thus, light v* acts as a probe that seeks a goal to value its features. This goal consists in the DP *the vase*. This probe-goal relation, established via c-command, values both the features of the probe and the Case features of the goal (assuming Case assignment via Agree, as proposed by Chomsky 2000, 2001, 2004; and

Legate 2008). Then the external argument *John* can be introduced by active Voice, also via probe-goal relationship. Voice also bears uninterpretable features that must be valued by a goal: EPP, φ-features, Case features, and possibly Topic/Focus features. Voice assigns an agentive thematic role to the argument in its specifier. Crucially, Gallego assumes that light *n** cannot introduce arguments because it does not have uninterpretable features to be valued – all the features of *n** are interpretable and do not need to be resolved.

Thus, in this framework, "predicates" that are phasic functional categories have unvalued uninterpretable features which need to be satisfied in a probe-goal relationship with a DP. As Gallego (2015:199) puts it,

> the lexicon is not formed by units that can take arguments and units that cannot ('relational' and 'non-relational', to use Mateu's 2002 terminology). All the elements present in the lexicon are relational (mergeable) by assumption – because syntax is. I will thus claim that the selection of internal arguments is forced due to feature valuation.

In sum, I will assume that predicates are probes and arguments are goals, and that they enter into c-command or spec-head relationships to derive the argument structure of the sentence.

Argument structure alternations (passivization, applicativization, causativization, etc.) can be analyzed as different ways in which the thematic roles are assigned to arguments and how they are realized in sentences. In the present work I will deal with the realization of focused arguments in Cinyanja, that is, the interaction between syntax, information structure, and prosody, which are related to the expression of focus in this language.

In this section I presented an account of the notion of argument structure and argument realization assumed in this dissertation. In the next section I will refer to Carstens' proposal regarding EPP and Agreement operations in Bantu languages. Her proposal will help us explain the AGREE operation that occurs in cleft and pseudocleft constructions in Cinyanja, regarding the hyper-raising of the thematic arguments out of the embedded CP.

3.3: EPP AND AGREEMENT IN BANTU

In this section I will explore Carstens' (2005) proposal regarding the mechanism of the operation AGREE in Bantu languages, as well as considerations on φ-features and EPP-features valuation. This proposal is relevant for the derivation I will propose in Chapter 5, for accounting for the apparent different positions of focused constituents in Cinyanja.

Carstens (2005:249) proposes that, in Bantu languages, functional heads bear uninterpretable φ-features which force the raising of XPs to their specifier positions. This proposal may be summarized as follows:

(47) Bantu φ-EPP: ***u***φ-features have EPP features in Bantu

A consequence of this generalization (47) is that "subject agreement is present if and only if spec-TP is filled, and C-agreement is present if and only if spec-CP is filled." (2005:249). Hence, the spec-head relation is the result of the operation AGREE.

Consider the following example from the Kilega language. In this interrogative sentence the WH-word *bikí* 'that' is realized *in situ*, immediately after the benefactive argument *mwámí* 'chief'. The locative argument *mumwílo* 'of the village' occurs in sentence-final position. Notice that the verb agrees with the agent argument *bábo bíkulu* 'those women'.

(48) *Bábo bíkulu bákásílé mwámí bikí mumwílo?*

babo	bikulu	b-a-kas-il-e	mwami	biki
2that	2woman	2SA-A-give-PERF-FV	1chief	8what

mu-mwilo
18-3village
'What did those women give the chief of the village?'

[Carstens (2005:220)]

This interrogative phrase can also be raised into a position in the left periphery. In this case it is the interrogative word that agrees with the verb. Note the example below. In this sentence the WH-word *bikí* 'what' triggers subject agreement with the verb – through the class 8 prefix {*bi-*}, and the agent argument *bábo bíkulu* 'those women' occurs immediately after the verb, followed by the benefactive argument *mwámí* 'chief' and the locative argument *mumwílo* 'of the village' in sentence-final position.

(49) *Bikí biákásílé bábo bíkulu mwámí mumwílo?*

biki	bi-a-kas-il-e	babo	bikulu	mwami
8what	8SA-A-give-PERF-FV	2that	2woman	1chief

mu-mwilo
18-3village
'What did those women give the chief in the village?'

[Carstens (2005:220)]

Thus, agreement with WH-words in Kilega may occur *in situ* or raised to spec-CP. When this argument is raised there emerges an agreement pattern in the verbal complex, namely the subject marker that agrees with the appropriate noun class prefix. This evidence clearly demonstrates that the argument was successively raised to a higher position, and at some point, it was probed by the

C/TP system in order to establish the AGREE operation. Carstens notices that the same pattern of word order and agreement that occurs in Kilega also appears in 70 other languages she examines. Thus, Carstens treats the data in the paper as representative of the Bantu family generally.

In this section I provided an account of the operation AGREE in Bantu languages. In the next section I will briefly discuss the properties of cleft constructions.

3.4: ON CLEFT AND PSEUDOCLEFT CONSTRUCTIONS

Cleft constructions are widely used to syntactically mark focus, especially contrastive focus, even though it also marks informational and identificational focus. Lambrecht (2001:467) proposes the following definition of the grammatical category "cleft construction":

> a cleft construction (CC) is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions.

Thus, a cleft construction corresponds to a complex sentence with a matrix and a dependent/relative clauses in which the focused argument is licensed by a copular verb and the complementizer may be used to connect the two clauses.

Guesser (2008) proposes a definition of cleft constructions as a syntactic strategy employed in order to mark constituent focus. Discussing data from Brazilian Portuguese, this author distinguishes two types of clefts, namely (i) canonical, and (ii) inverted. The canonical cleft consists in a construction in which the focused argument is raised from its base position in the dependent clause into the position immediately after the copular verb and followed by the complementizer. The following structure illustrates this kind of cleft:

(50) (copular verb) + (XP) + (complementizer) + (IP)

(51) *Foi a MARIA que comeu o bolo.*

foi	a	Maria	que	comeu	o	bolo
COP	the	Mary	COMP	eat.PAST.3SG	the	cake

'It was MARY that ate the cake.'

On the other hand, the inverted cleft consists in a construction in which the copular verb is realized after the focused argument. The structure of this type of cleft is illustrated below.

(52) (XP) + (copular verb) + (complementizer) + (IP)

(53) *A MARIA foi que comeu o bolo.*

a	Maria	foi	que	comeu	o	bolo
the	Mary	COP	COMP	eat.PAST.3SG	the	cake

'MARY was who ate the cake.'

Kato (2017), also investigating Brazilian Portuguese data, provides a diachronic discussion about the syntactic change in the structure of clefts. "Old" Portuguese employed the inverted cleft as in (53), but "modern" Portuguese makes use of the canonical cleft as in (51). The author also shows that the copula may be erased, as in (54a-b). Additionally, the complementizer is also optional, as in (54a).

(54a) *Onde ø os meninos dormem?*

onde	ø	os	meninos	dormem
where	COMP	the	boys	sleep.PRES.3PL

'Where do the boys sleep?'

(54b) *Onde que os meninos dormem?*

onde	que	os	meninos	dormem
where	COMP	the	boys	sleep.PRES.3PL

'Where that the boys sleep?'

Furthermore, according to Almeida (2011), a pseudocleft construction is found in Portuguese, in which the complementizer occurs in sentence-initial position, followed by the dependent/relative clause, and then the copula is realized immediately before the focused constituent. This structure is illustrated below.

(55) (complementizer) + (IP) + (copular verb) + (XP)

(56) *Quem quebrou o vaso foi JOÃO.*

quem	quebrou	o	vaso	foi	João
who	break.PAST.3SG	the	vase	COP	John

'Who broke the vase was JOHN.'

Regarding the Nyanja language, the following two constructions are found: (i) the canonical cleft, and (ii) the pseudocleft. The canonical cleft is realized just like the Brazilian Portuguese version, as seen in (51). Notice that the focused argument *mwána* 'child' agrees with the complementizer *méne* 'that', which suggests cyclical movement upwards from the dependent clause into the matrix clause.

(57) (copular verb) + (XP) + (complementizer) + IP

(58) *Ndí MWÁNA, áméne wang'amba kabudula pawaya obaya.*

ndi	*mwana*	*a-mene*	*w-a-ng'amb-a*	*kabudula*
COP	1-child	1-COMP	1SM-PAST-tear-FV	5-shorts

pa-waya	*obaya*
LOC-wire	5-barbed

'It was the CHILD that tore his shorts in the barbed wire.'

On the other hand, in example (59) below the focused argument *n'zimáyi* 'woman' appears in sentence-final position, immediately after the copular verb *ndí*. The dependent clause occurs higher than the main clause, and the complementizer does not appear in this construction. This construction corresponds to a pseudocleft which resembles the Brazilian Portuguese structure shown in (56). However, in Cinyanja pseudocleft constructions the complementizer is optional. The structure below illustrates this pseudocleft in Cinyanja.

(59) (complementizer) + (IP) + (copular verb) + (XP)

(60) *(Amene) wawúmbíla mbíya ya mádzi mwáná, ndí N'ZÍMÁYI.*

a-mene	*w-a-wumb-il-a*	*mbiya ya madzi*
1-COMP	1SM-PAST-shape-APPL-FV	3-pot of water

mwana	*ndi*	*n'zimayi*
1-child	COP	1-woman

'Shaped the waterpot for the child was the WOMAN.'

In the analysis of cleft constructions developed in Chapter 5, I will propose that in the canonical cleft the focused argument is successively moved upwards, triggering agreement operations with T and C in the relative/dependent clause, before it lands into the specifier of a low Focus projection sitting in the main clause headed by the copula *ndí*. Regarding the pseudocleft, I will propose that the whole dependent clause (=CP) is moved upwards to the left periphery of the main clause and is thus linearized to the left of the copula.

In this section I briefly presented the definition and properties of cleft constructions. In the next section I will discuss the notion of information structure.

3.5: ON INFORMATION STRUCTURE

Prince (1981:224) argues that information structure – also known as information packaging - deals with the "the tailoring of an utterance by a sender to meet the particular assumed needs of the intended receiver. That is, information-packaging in natural language reflects the sender's hypotheses about the receiver's assumptions and beliefs and strategies." As Lambrecht (1994:1) puts it, "grammatical analysis at this level is concerned with the relationship

between linguistic form and the mental states of speakers and hearers and that the linguist dealing with information structure must deal simultaneously with formal and communicative aspects of language." It is crucial to note that "the study of information structure is not concerned with lexical and propositional content in the abstract but with the way such content is transmitted." (Lambrecht 1994:3).

Rodrigues and Menuzzi (2009:1) argue that "the crucial element of 'informational structure' is its *interface* character in the organization of linguistic knowledge. In a first approach, information structure can be characterized as the 'meaning' component that encode distinctions such as new/old information, topic/commentary, focus/presupposition, etc.". The authors note that these distinctions can alter truth-value conditions of the sentence, but the most elementary observation regarding information structure deals with different ways in which the propositional content of a sentence can be organized without changes in truth-value conditions. See the following data from English.

English

(61a) John broke the vase with a hammer.
(61b) The vase, John broke it with a hammer.
(61c) The hammer, John broke the vase with it.
(61d) John, he broke the vase with a hammer.
(61e) The vase was broken (by John) (with a hammer).

All the sentences in (61) have the same propositional content, that is, they all have the same truth-value, as seen in (62).

(62) ∃e [**Break**(e) ∧ Agent(e,John) ∧ Patient(e,vase) ∧ **With**(e,hammer)]

Thus, all the sentences in (61) describe an event in which a person called John causes a change of state of an object called vase using an instrument called hammer. However, the same content can be conveyed in many different ways, which change the informational status of the participants of the event: (61a) and (61d) seem to be "about John", whereas (61b) and (61e) seem to be "about the vase", and (61c) seems to be "about the hammer". That is, the different ways of organizing the sentence correspond to different ways of presenting the information, depending on the context/discourse.

According to Rodrigues and Menuzzi (2009:3), the contribution of meaning from information structure has to do with encoding or signaling whatever the speaker judges as shared or not between the speaker and the hearer in a given context/discourse and deals with updating the mental representations of the hearer. As proposed by Engdahl and Vallduví (1996:460)

> information packaging is a structuring of sentences by syntactic, prosodic, or morphological means that arises from the need to meet the communicative demands of a particular context or

> discourse. In particular, information packaging indicates how information conveyed by linguistic means fits into the (hearer's mental model of) context or discourse. When communicating a proposition ϕ, a speaker may realize it by means of different sentential structures according to his/her beliefs about the hearer's knowledge and attentional state with respect to ϕ.

Lambrecht (1994:5) proposes the following definition of information structure:

> that component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts. (…) The information structure of a sentence is the formal expression of the pragmatic structuring of a proposition in a discourse.

Thus, information structure deals with the different ways of representing referents and organizing the arguments of a sentence in such a way as to convey specific bits of meaning and information. The next section comprises some facts about the notion of information.

3.5.1: ON THE NOTION OF INFORMATION

The notion of information corresponds roughly to the notion of knowledge. According to Lambrecht (1994), the universe of discourse comprises the world external to the text (constituted by the participants of the discourse and the "context" setting of the discourse - place, time, circumstances), and the world internal to the text, an abstract world of linguistic representations and their meanings, created in the minds of the interlocutors in the process of communication. It is precisely the manipulation of such representations that allows information to be expressed. The linguistic component of information structure necessarily involves these two worlds (external and internal) since it matches form-meaning pairs with the mental states of the interlocutors. Lambrecht argues that "information can strictly speaking only be conveyed relationally, via propositions" (1994:46). That is, "to inform a person of something is to induce a change in that person's knowledge by adding one or more propositions" (1994:44).

The first distinction that can be made regarding information is the *old/new* information. According to Lambrecht (1994:45), "it is a fundamental property of information in natural language that whatever is assumed by a speaker to be NEW to a hearer is information which is ADDED to an already existing stock of knowledge in the hearer's mind". Thus, there is no "new" information without "old" information. Prince (1981) proposed three levels of givenness to better

distinguish between “old” and “new information”: (i) predictability/ recoverability; (ii) saliency; and (iii) shared knowledge.

i. Givenness: the speaker assumes that the hearer CAN PREDICT OR COULD HAVE PREDICTED that A PARTICULAR LINGUISTIC ITEM will or would occur in a particular position WITHIN A SENTENCE.

ii. Givenness: the speaker assumes that the hearer has or could appropriately have some particular thing/entity/… in his/her CONSCIOUSNESS at the time of hearing the utterance.

iii. Givenness: the speaker assumes that the hearer “knows”, assumes, or can infer a particular thing (but is not necessarily thinking about it).

Lambrecht (1994:48) proposes to treat the difference between "new" and "old" information in terms of the concepts of "presupposition" and "assertion", and, also, with the concept of "focus". For the author (1994:49), the relevant distinction for the study of informational structure has to do with PRAGMATIC STATES of the denotata of the constituents of an individual sentence and the PRAGMATIC RELATIONS between these denotata and the propositions in which they play predicate or argument roles. "It is the establishment of such pragmatic relations that makes information possible" (1994:49).

The notion of information structure is dependent on the notion of “common ground” – the set of information that is mutually known and shared in communication, according to Krifka and Musan (2012:2). The authors argue that

"communication can be seen as continuous change of the *common ground*, i.e., of the information that is mutually known to be shared in communication; speakers plan their contributions with respect to the common ground." In order for communication to hold between the speaker and the hearer, the process of accommodation allows for "repairing" the common ground. It was defined by Lewis (1979:340) as

i. If at time t something is said that requires presupposition P to be acceptable, and if P is not presupposed just before r, then - *ceteris paribus* and within certain limits - presupposition P comes into existence at t.

The following examples, taken from Krifka and Musan (2012:1-2), show how accommodation works:

(63a) I have a cat, and I had to bring my cat to the vet.
(63b) #I had to bring my cat to the vet, and I have a cat.

In the examples in (63), it is observed that the second sentence in (63b) presents information that was already presupposed in the first sentence, that is, the information is already present in the common ground after the utterance of the first sentence, which renders (63b) as not felicitous in context (symbolized by #).

The analysis of information structure is dependent on the notions of proposition, pragmatically structured proposition, presupposition, and assertion.

Proposition is the part of the meaning of a sentence that remains constant despite changes in such things as voice or illocutionary force of the sentence. It can be related to other units of its type through interpropositional relations, such as temporal or logical relations. For instance, in the following sentences the proposition is the same, although there are changes in the argument structure.

(64a) John broke the vase.
(64b) The vase was broken by John.

Despite differences in word order, voice, and grammatical functions, the two sentences present the same propositional content and thematic relations:

i. There is a change of state event "to break"
ii. John is the trigger of the event
iii. The vase is affected by the event

A pragmatically structured proposition is one that is structured in terms of presupposition (if any) and assertion. Such a proposition will have a prosody characteristic of its own and its felicity of enunciation depends on the context.

PROPOSITION

(65) Mariana made the marmalade.

PRAGMATICALLY STRUCTURED PROPOSITION

(66a) [MARIANA[11]]$_{FOC}$ made the marmalade. (who made the marmalade?)
(66b) Mariana made [THE MARMALADE]$_{FOC}$. (what did Mariana make?)

The notion of presupposition is defined by Lambrecht (1994:52) as "the set of propositions lexicogrammatically evoked in a sentence which the speaker assumes the hearer already knows or is ready to take for granted at the time the sentence is uttered."

On the other hand, assertion is defined by Lambrecht (1994:52) as "the proposition expressed by a sentence which the hearer is expected to know or take for granted as a result of hearing the sentence uttered."

According to Lambrecht (1994:77), another important notion in information structure is identifiability, a cognitive category whose name was suggested by Chafe (1976), who noted that

> to designate referents for which a representation exists in the addressee's mind the term 'identifiable' is preferable to the sometimes suggested terms 'known' or 'familiar'. […] what counts for the linguistic expression of the cognitive distinction in question is not that the addressee know or be familiar with the referent in question

[11] Small caps indicates prosodic prominence.

> […] but that he be able to pick it out from among all those which can be designated with a particular linguistic expression and identify it as the one which the speaker has in mind.

Just as an asseverated proposition is one of which only the speaker has the mental representation at the moment of utterance, the identifiable referent is one for which a shared representation already exists in the minds of the speaker and the interlocutor at the moment of utterance (inference: the unidentifiable referent is represented only in the mind of the speaker).

The concept of identifiability is related to the notion of SPECIFICITY of referents of indefinite nominal phrases. Natural languages have several strategies for marking DEFINITENESS and SPECIFICITY, and these strategies do not establish a binary relationship. Definiteness is one of the grammatical mechanisms for marking identifiability, i.e., it is a discrete grammatical category. Identifiability, according to Lambrecht, is a continuous cognitive category. Other mechanisms can be used to mark identifiability: word order (topicalization); presence/absence of numerals; presence/absence of case, etc.

Identifiability can be expressed in languages in different ways, as in (i) double distinction (definite article x indefinite article) or (ii) triple distinction (bare noun x definite article x indefinite article). The examples below illustrate the triple distinction in Portuguese.

PORTUGUESE

(67a) Livro é sempre um bom presente. [*generic*]
Book is always a good gift
'Book is always a good gift.'

(67b) Comprei um livro ontem. [*indefinite specific/non-specific*]
buy.PAST.1SG a book yesterday
'I bought a book yesterday.'

(67c) Comprei o livro ontem. [*definite = specific*]
buy.PAST.1SG the book yesterday
'I bought the book yesterday'

Generic indefinites are always identifiable, since the referent is either a whole class or some representative member of the set, which requires the hearer to identify the semantic class designated by the lexical head.

Another important notion is that of activation, or accessibility, which concerns the level of awareness of certain referents. Chafe (1987) proposes the following classification:

i. **Active** – "currently lit up, a concept in a person's focus of consciousness at a particular moment."

ii. **Semi-active** (accessible) – "it is in a person's peripheral consciousness, a concept of which a person has a background awareness, but one that is not being directly focused on."

iii. **Inactive** – "it is currently in a person's long-term memory, neither focally nor peripherally active."

These states are related to short-term (+ active) and long-term (- active) memory and to the speaker's center of attention (centering).

Lambrecht (1994:95) states that the category ACTIVATION has grammatical correlates in prosody (phonological attenuation), morphology (pronominal, inflectional or zero coding), and syntax (more difficult to demonstrate). Phonological attenuation is not a necessary condition for activation.

Activation states are correlated to formal aspects of sentence structure:

i. Active state
   a. Lack of tonal prominence [systematic]
   b. Pronominalization [non-systematic]

ii. Inactive state
   a. Accentuation [iconic relation]
   b. Nominal expression

According to Lambrecht (1994:98), while "the ABSENCE of prosodic prominence on a constituent necessarily indicates active status of the coded referent or denotatum, the PRESENCE of prominence has no analogous distinguishing function."

The semi-active state can also be called accessible: there is a greater difficulty in mentally retrieving the referent, compared to the active state. There are three sources of accessibility:

i. Textual – a distant referent
ii. Inferential – from an inference
iii. Situational – salient presence in the external world

Lambrecht (1994:100) points out that the distinction between the three states of activation of referents is based on the idea that there are different kinds of mental "cost" or "effort" involved in processing mental representations. However, from a strictly grammatical (phonological and/or morphosyntactic) point of view, only one binary distinction is justifiable: the distinction between referents MARKED AS BEING ACTIVE (attenuated pronunciation and/or pronominal encoding) and referents NOT MARKED AS ACTIVE.

Lambrecht (1994:104-105) suggests that the cognitive category ACCESSIBILITY should be treated as a POTENTIAL FOR ACTIVATION rather than THE STATE OF A REFERENT in the person's mind.

> Given accessibility of a referent, a hearer will exploit this potential – by drawing inferences or by searching the text-external or text-internal world – if she is invited to do so ON THE BASIS OF THE PRESUPPOSITONAL STRUCTURE OF THE SENTENCE.

She will NOT look for a referent if such an invitation is not grammatically expressed.

The connection between identifiability and activation can be summarized as follows:

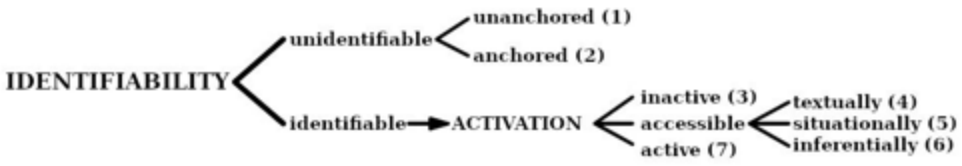


Figure 4: The correlation between identifiability and activation (Source: Lambrecht (1994:109))

The numbers (1) through (7) refer to the following terminological conventions:

1. Unidentifiable/brand-new
2. Unidentifiable anchored/brand-new anchored
3. Inactive/unused
4. Textually accessible
5. Situationally accessible
6. Inferentially accessible
7. Active/given

Hajičová (1993) proposes the following activation scale:

1. items in the focus: maximum activation level
2. Items in the topic: level just below the maximum

3. Items mentioned in a previous utterance or with some associative link: immediately following level of activation

However, the notions of identifiability and activation are not to be confused with the notions of topic/focus - topic and focus are relations between referents and a proposition, that is, they are informational functions. Identifiability and activation are cognitive notions involved in the encoding and processing of these referents.

Another concept of information structure is the notion of contrastiveness. According to Lambrecht (1994), contrastiveness is a graded concept involving a set of alternatives. Both topics and foci can be contrastive. The notion of contrast cannot be used to explain the assignment of accents – considering the rather ungrammatical notion of contrastiveness, it is preferable to maintain the proposal of the existence of a focus/activation accent. The notion of contrast is at the pragmatic level: it is a CONVERSATIONAL IMPLICATURE. In the next section I will present the notion of Focus.

3.5.2: ON THE NOTION OF FOCUS

According to Lambrecht (1994), focus is an informational function that pervades the idea of signaling a portion of the proposition that cannot be taken

for granted at the moment of speech – it is an unpredictable or pragmatically non-recoverable element in an utterance. In other words, it is the semantic component of a pragmatically structured proposition by which assertion is differentiated from presupposition. This difference manifests itself at the grammatical and prosodic levels.

There are no structures without focus. For a structure to have independent value, it has to express an assertion, that is, it has to add new content to the common ground.

According to Krifka and Musan (2012:7), building on top of Rooth's (1985) work, "focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expression".

According to Lambrecht (1994:214), "the syntactic domain in a sentence which expresses the focus component of the pragmatically structured proposition will be called the FOCUS DOMAIN". He also notes that "a denotatum which is not in focus is necessarily in the presupposition". Additionally, the author states that "the word or minimal constituent carrying the focus accent will be referred to as the ACCENTED CONSTITUENT."

Consider the question-and-answer pair below.

(68) Q: What did you do last night?
A: $[\text{I}]_{\text{TOP}}$ [went to the MOVIES]$_{\text{FOC}}$

In the answer above, a topic-commentary structure, the constituent *I* is the topic, the focus domain is the predicate *went to the movies*, the accented constituent is *movies*.

The domain of focus can include non-focal elements.

(69) Q: What did John do to the vase?
A: $[\text{John}]_{\text{TOP}}$ [BROKE $[\text{the vase}]_{\text{TOP}}]_{\text{FOC}}$

Focus is usually related to accent (prosodic prominence), but there is also activation accent, which indicates referents present in long-term memory, and which will be used as arguments in new assertions.

Beaver and Clark (2009) argue for the Focus Principle, that is, “some part of a declarative utterance must evoke a set of alternatives containing the propositions of the Current Issue." This assumes that at each stage of common ground development, a current issue can be assumed. The authors assume three degrees of association of the focus with semantic operators: quasi-associations, free associations, and conventional associations.

The different types of focus structures correspond to the different types of pragmatically structured propositions. These structures allow capturing correspondences between functionally identical but formally divergent

sentences, as well as evoking differences in syntactic focus domains and portions of the pragmatically structured proposition. In addition, it allows capturing correspondences between presuppositional structures and focal structures.

There are three basic types of focal structures:

Proposition: *My car broke down.*

i. Predicate focus

Q: what happened to your car?
A: my car [BROKE DOWN]$_{FOC}$.

This is the most common structure, also known as topic-commentary.

It is minimally expressed by prosodic prominence of an element from the predicate.

- Presupposition: 'the car of the speaker is a topic for commentary x'
- Assertion: 'x = broke down'
- Focus: 'broke down'
- Focus domain: VP (verb phrase)

ii. Argument focus

Q: I heard your bike broke down?
A: no, [my CAR]$_{FOC}$ broke down.

Marked structure where the presupposition includes an open proposition; this is also marked as active (by lack of prosodic prominence). Prosodically it is the inverse of the predicate focus structure.

- Presupposition: 'the x of the speaker broke down'
- Assertion: 'x = car'
- Focus: 'car'
- Focus domain: NP (noun phrase)

iii. Sentence focus

Q: What happened?
A: [My CAR broke down]$_{FOC}$

Absence of relevant presuppositions; the assertion extends through all of the proposition; event reporting.

Expressed in certain languages via the same mechanisms of argument focus; or by subject-object inversion; or by clefts.

Common feature: absence of topic/commentary.

- Presupposition: -----
- Assertion = focus: 'the car of the speaker broke down'
- Focus domain: CP (the whole sentence)

These structures can manifest themselves in various ways in languages: some mark by prosody alone (English in some contexts); some mark by prosody and morphology (Japanese); some mark by prosody and syntax/constructions (Portuguese, French, Italian, English in other contexts[12]).

Finally, there is one last type of focal structure: truth focus, also known as counter-assertive or counter-presuppositional. See the following example.

(70) Q: let's go to the kitchen to fix something to eat.
A: [THERE'S] nothing to eat.

The different focus structures can combine themselves.

(71) $[_{PRES}$ THE KITCHEN,$]$ $[_{ASS}$ you$]_{PRES}$ have to clean it$]]_{ASS}$ (not me).

Rooth (1985, 2002) defines focus as a set of alternatives that are relevant to the interpretation of linguistic expressions. However, the notion of alternatives is not only present in the concept of focus - it is a notion that plays a role in language all the time. This definition means that focus especially points to and emphasizes the existence of particular alternatives (Krifka and Musan 2012:7).

[12] Like other languages, English makes use of cleft and pseudocleft constructions which express focus, especially contrastive focus.

Krifka and Musan (2012:7-8) differentiate between EXPRESSION FOCUS and DENOTATION FOCUS. Alternatives can be of form (= expression) or of denotation. The expression of "alternatives" can affect various aspects, such as word choice and pronunciation, and need not necessarily involve meaningful constituents or units.

EXPRESSION FOCUS is generally used for corrections, and sometimes (but not always) occurs together with an explicit negation, as in the example below.

(72) Grandpa did not [kick the BUCKET]$_{FOC}$, he [DIED]$_{FOC}$.

The two relevant alternatives in the sentence in (89) are the focal expressions *kick the bucket* and *dying*. These alternatives cannot be the denotation since they are synonyms. Therefore, they are focal expressions, to distinguish different possibilities of referring to the same event.

Whereas in DENOTATION FOCUS the relevant alternatives are interpreted at the level of denotation, leading to alternative denotations of complex expressions.

According to Krifka and Musan (2012:8):

> Denotation focus on an expression Y with meaning ||y|| leads to the presumption of a set of alternative meanings that play a role in interpreting the constituent in which Y occurs. The alternative denotations need to be comparable to the

> denotation of the expression in focus, i.e., they need to be of the same type, and often also need to be of the same ontological type (persons or tenses), and may be further restricted by the context of utterance. [...] is certainly more important in communication.

In short: focus is the property of a string [Y1...Yi,F...YN] to point to the existence of alternatives to both form (expression) and denotation. Usually this appears as accent, but it can also be achieved by other means such as certain syntactic constructions.

Some semantic operators are associated with focus, such as *only, also, even*. These operators are also involved with the notion of alternatives: specifying an individual from a set (only); expressing the assumption that the assertion extends to other alternatives (also, too); assuming that the denotation of the focused constituent is, in some respect, extreme when compared to other alternatives (even, scalar particles).

According to Krifka and Musan (2012:19), the notions of broad and narrow focus are "imprecise terms that can only be applied when different readings of focus are under discussion." The authors propose (2012:20) the terminology closed focus vs. open focus. This distinction concerns the size of the set of alternatives.

CLOSED (or STRICT) FOCUS concerns sets of alternatives limited to few elements, sometimes only two - the item in focus and an alternative. "It is tempting to call focus with a limited set of alternatives 'contrastive,' as Chafe (1976) did, but the example below seems no more contrastive than the response to non-restrictive questions like WHAT DO YOU WANT TO DRINK?" (2012:20). Closed focus is the case in corrections, contrasts, polar/yes-or-no questions, and in answers to alternative or restricted constituent questions, such as (73) below.

(73) Q: what do you want to drink, coffee or tea?
A: I want [TEA]$_{FOC}$

OPEN (or WIDE) FOCUS involves non-restricted sets, satisfying only the general condition that all alternatives must be compatible with the focus in their semantic type, as shown in (74) below.

(74) Q: what do you want to drink?
A: I want [TEA]$_{FOC}$.

Krifka and Musan (2012:21) suggest using the term CONTRASTIVE FOCUS for cases where there is an assumption that the current common ground contains a proposition with which the current utterance can be contrasted, or that such a

proposition can be accommodated." The typical use of contrastive focus is corrective (75a), but it can also be additive (75b).

(75a) Q: John got married.
A: No, it was [PETER]$_{FOC}$ that got married.

(75b) Q: John wants coffee.
A: [MARY]$_{FOC}$ wants coffee [TOO]$_{FOC}$.

Regarding EXAUSTIVE FOCUS, it

> "is related to the specific way in which the contribution of the alternatives is interpreted. It serves to indicate that the denotation of the focus is the only one that leads to a true proposition, or more generally: “the denotation of the focus is logically the strongest denotation that does so" (2012:21).

This type of focus is incompatible with additive particles as *too/also*, as shown in the examples below.

(76a) It was [JOHN and BILL]$_{FOC}$ who stole a cookie.
(76b) #It was [JOHN and BILL]$_{FOC}$ who also stole a cookie.

In this example the interpretation is that no one else stole a cookie, only John and Bill.

Krifka and Musan (2012:21) also distinguish the IDENTIFICATIONAL FOCUS, which expresses an assertion of identity, as shown in the example below.

(77) The people who stole the cookie were [JOHN and BILL]$_{FOC}$.

And finally, the authors mention SCALAR FOCUS, also called EMPHATIC FOCUS. In this case the alternatives are ordered, and the denotation of the focus is often the smallest or largest element. Scalar particles like *even* or *at least* require scalar focus, as do strong polarity items. Consider the example below, which illustrate this type of focus construction.

(78) [WILD HORSES]$_{FOC}$ would not drag me there.

Kiss (1998) distinguishes between INFORMATION FOCUS and IDENTIFICATIONAL FOCUS. According to the author (1998:245), identificational focus "represents a subset of the set of contextually or situationally given elements for which the predicate phrase can potentially hold; it is identified as the exhaustive subset of this set for which the predicate phrase actually holds". That is, "the constituent which functions as the identificational focus represents the value of the variable bound by an abstract operator expressing exhaustive identification" (1998:245). This constituent syntactically acts as an operator

which moves into a scope position in the specifier of a functional projection and binds a variable.

On the other hand, information focus occurs "if a sentence part conveys new, nonpresupposed information marked by one or more pitch accents – without expressing exhaustive identification performed on a set of contextually or situationally given entities" (1998:246). The author argues that information focus is not associated with movement[13]. Information focus is present in every sentence, but not every sentence contains identificational focus. Kiss (1998:247) argues that this distinction is essential in the analysis of Hungarian, since the two types of focus are never optional interpretational variants but are associated with distinct structural positions, as seen in the examples below (1998:247). Small capitals indicate information focus.

(79a) *Tegnap este Marinak mutattam be Pétert.*

tegnap	este	**Marinak**	mutattam	be	Pétert
last	night	Mary.DAT	introduced.I	PERF	Peter.ACC

'It was **to Mary** that I introduced Peter last night.'

(79b) *Tegnap este be mutattam Pétert* MARINAK.

[13] Erteschik-Shir (1986) argues that most wh-words are not the focus of an open question, but I will assume that, in the cases I am investigating, the interrogative words do express information focus. Thus, there is evidence which suggests that informational focus is indeed related to WH-movement in Cinyanja, in the contexts of open (WH) questions. I will explore this in chapters 4 and 5.

tegnap	este	be	mutattam	Pétert	Marinak
last	night	PERF	introduced.I	Peter.ACC	Mary.DAT

'Last night I introduced Peter TO MARY.'

In the examples above, sentence (79a) contains an exhaustive focus constituent that occurs immediately before the verb. Kiss (1998:247) argues that it means that of a set of individuals present in the domain of the discourse it was Mary and no one else that I introduced to Peter last night. On the other hand, the postverbal focus in (79b) "merely presents Mary as nonpresupposed information, without suggesting that Mary was the only one of a set of relevant persons that I introduced Peter to last night." (1998:247). Thus, according to the author, the structured meaning theory of focus would assign to both sentences the structure below.

(80) ASSERT (<λx.introduced (I, Peter, x), Mary>)

The representation above is divided into a background part that contains a version of the proposition in which the focus is replaced by a variable bound by lambda, and into a focus part which is bound by an illocutionary operator represented by ASSERTION. The interpretation of the structure above consists in the following: "the set of persons for whom it holds that I introduced Peter to them yesterday is under discussion, and it is stated that, among these persons, it holds for Mary that I introduced Peter to her yesterday" (1998:247-248). This

formula above, according to the author, “assimilates information focus to identificational focus in that it represents also information focus as a member of a set of alternatives” (1998:248).

Kiss (1998:248) argues that the difference between identificational focus and information focus resides in the following aspects (1998:248):

i. Identificational focus expresses exhaustive identification; information focus merely marks the nonpresupposed nature of the information it carries;
ii. Certain types of constituents, universal quantifiers, *also*-phrases and *even*-phrases, for instance, cannot function as identificational foci; the type of constituents that can function as information foci is not restricted;
iii. Identificational focus takes scope, whereas information focus does not;
iv. Identificational focus is moved to the specifier of a functional projection; however, information focus does not involve movement;
v. Identificational focus is always coextensive with an XP available for operator movement, whereas information focus can be either smaller or larger;

vi. Identificational focus can be iterated while information focus can project.

Thus, based on the discussion above, the present research will deal mostly with identificational focus, that is, constituents that are a subset of contextually given elements which can bind the focus variable. This type of focus is identified as closed/strict focus by Krifka and Musan (2012), but I will adopt Kiss' (1998) terminology. This choice is due to the nature of the sets of question-and-answer pairs I am employing to investigate the position of focus in the language Cinyanja. The pragmatically structured propositions I am investigating are derived from WH-questions in which all the arguments are given in the context.

The hypothesis that this dissertation will advance is that informational and identificational focus are realized in a dedicated syntactic position, namely the position Immediately After the Verb (IAV). I will base the cleft analysis on Guesser's (2008), Kato's (2015, 2017), and Lambrecht's (2001) proposals regarding cleft constructions (cf. section 3.4). Kato (2015) notices that the cleft construction can be used for both informational and contrastive Focus. As shown earlier, in Cinyanja there occurs the canonical cleft and a pseudocleft.

In the next section I will discuss the typology of focus in Bantu languages.

3.5.3: A TYPOLOGY OF MORPHOSYNTACTIC FOCUS MARKING IN BANTU LANGUAGES

Many languages employ a dedicated position for the expression of focus. Gibson et al (2017) argue that there are basically four possibilities for the syntactic expression of focus: sentence-initial, sentence-final, immediately before the verb (IBV), and immediately after the verb (IAV). Thus, focus is syntactically identified based on verb-adjacency or sentential edges.

To illustrate the typology, consider the following examples from the languages Hausa (sentence-initial), Ostyak (IBV), Aghem (IAV), and Tangale (sentence-final). These examples were taken from Gibson et al (2017:64-66). Small capitals indicate prosody prominence.

(81a) HAUSA (SVO)

Bintà zaa tà biyaa teelà.

Binta	*zaa*	*tà*	*biyaa*	*teela*
Binta	FUT	3SG.F	pay	tailor

'Binta will pay the tailor.'

(81b) *Teelà (nee) Bintà zaa tà biyaa.*

teela	*nee*	*Binta*	*zaa*	*ta*	*biyaa*
tailor	FOC	Binta	FUT	3SG.F	pay

'Binta will pay the TAILOR.'

(82a) OSTYAK (SOV)

*(*Xojna) tam a:n xojna tu:-sa?*

xoj-na	*tam*	*a:n*	*xoj-na*	*tu:-s-a*
who-LOC	this	cup	who-LOC	take.away-PAST-3SG.PASS

'Who took away this cup?'

(82b) *(*Xoj e:lti) tăm a:n xoj e:lti mase:n?*

*(*xoj*	*e:lti)*	*tăm*	*a:n*	*xoj*	*e:lti*	*ma-s-e:n?*
who	to	this	cup	who	to	give-PAST-2SG.OM2

'To whom did you give this cup?'

(82c) *(*Juwana) tam a:n Juwana mase:m.*

*(*Juwan-a)*	*tam*	*a:n*	*Juwan-a*	*ma-s-e:m.*
John-LAT	this	cup	John-LAT	give-PAST-1SG.OM2

'I gave this cup to John.'

(83a) AGHEM (S(AUX)VOX)

Fīl á mɔ̀ zī́ án sóm bɛ́kɔ́.

fīl	*á*	*mɔ̀*	*zī́*	*án-↓sóm*	*bɛ́-↓kɔ́.*
friends SM	es	PAST2	ate	LOC-farm	fufu-CL

'The friends ate fufu in the FARM.'

(83b) *À mɔ̀ zī́ áfīn bɛ́↓kɔ́ án ↓sóm.*

à	*mɔ̀*	*zī́*	*á-fīn*	*bɛ́-↓kɔ́*	*án↓sóm.*
es	PAST2	ate	friends	fufu-CL	LOC-farm

'The FRIENDS ate fufu in the farm.'

(84a) TANGALE (SVO)

Pàdgò tàabéè nóŋ?

pàd-gò	*tàabéè*	*nóŋ?*

buy-PERF tobacco who
'Who bought tobacco?'

(84b) *Pàdgò tàabéè kài.*

pàd-gò *tàabéè* *kài.*
buy-PERF tobacco Kai
'KAI bought tobacco.'

Focus can also be marked morphologically in some languages. The conjoint/disjoint alternation consists in an instance of this phenomenon. There are two possibilities of morphological marking of constituent focus: (i) marking on the verb; (ii) marking on the constituent. Colloquial Sinhala, for example, employs an obligatory change in the final vowel of the verb; additional marking on the focused constituent is optional. Compare the examples below.

(85a) COLLOQUIAL SINHALA

Mamə gaməṭə yanna.

mamə *gaməṭə* *yann-a.*
1SG.NOM village-DAT go.PRES-NFOC
'I go to the village.'

(85b) *Mamə gaməṭə(y) yanne.*

mamə *gaməṭə(-y)* *yann-e.*
1SG.NOM village-DAT(-EMPH) go.PRES-FOC
'It is to the village I go.'

(85c) *Mamə yanne gamə̣tə(y).*

mamə *yann-e* *gamə̣tə(-y)*
1SG.NOM go.PRES-FOC village-DAT(-EMPH)
'It is to the village I go.'

Marking on the focused constituent is employed by the language Sri Lankan Malay. See the examples below.

(86a) SRI LANKAN MALAY

(So, [while] the king was planning to make a fool out of Andare)

Raja=jo su-jaadi enco.

Raja=jo *su-jaadi* *enco*
king=FOC PST-become fool
'The king (himself) became a fool.'

(86b) (Now that we've done all this)

...Siinijo araduuduk.

...Siini=jo *ara-duuduk.*
here=FOC PRS-stay
'… here we stay.'

Thus, based on the discussion above, Gibson et al (2017:68) propose the following typology of focus marking.

morphology/typology	initial	IBV	IAV	final
on term				
on verb				

One of the objectives of the present dissertation is to assess the strategies used to mark focus in the Nyanja language as spoken in Mozambique, that is, where in this typology this language fits. The proposal I will develop in this research is that Cinyanja marks constituent focus via prosodical and syntactic means, that is, the focused constituent is marked prosodically and occurs immediately after a copular verb. This copula may be realized segmentally through the word *ndí*, or via a residual high tone which spreads to the penultimate syllable of the focused argument. Thus, my hypothesis is that, in some instances, the copula *ndí* lost segmental phonological material and is realized via prosody effects such as high floating tone spreading.

In this section I presented a typology of focus marking in Bantu languages. The next section will present a brief account of Zubizarreta's (2010) account of the marking of focus in three Bantu languages, including Cichewa, which is a variant of Cinyanja (although, as previously mentioned, there are many differences between them).

3.5.4: SYNTACTIC AND PROSODIC FOCUS EXPRESSION IN BANTU LANGUAGES

Zubizarreta (2010) presents an overview of the literature on focus marking in three Bantu languages, namely Kimatuumbi, Chimwiini, and Cichewa, comparing the data with examples from Italian and Spanish. In this brief account I will focus on the Bantu languages, especially Cichewa, which is a variant of Nyanja. Zubizarreta proposes that these languages present two characteristics in common: (i) an active low Focus position, and (ii) the Focus position triggers the insertion of a strong prosodic boundary, which gives rise to a "ripple effect" in that phrases to the right of Focus are similarly flanked by a comparable prosodic boundary. The author argues for a stronger syntax-prosody connection in the expression of focus.

Zubizarreta (2010:133) assumes that functional categories are visible to the algorithm that inserts p-boundaries and argues that the Focus head (a functional head) can trigger the insertion of a p-boundary. The author notes that studies on Bantu languages suggest a low Focus position, immediately above vP (or VoiceP). Thus, one can assume that there are two possible Focus positions: (i) a high Focus above IP, and (ii) a low Focus between IP and vP/VoiceP. A language with syntactically active Focus projection moves the focused argument

to spec-Foc overtly, whereas languages with non-active Focus move the focused argument covertly. As noted by Aboh (2010), many languages have an overt morphological marker for Focus, such as *wɛ̀* in Gungbe. The examples below illustrate the occurrence of dedicated morphological markers of Focus and Topic. Notice that in this language there are two markers, namely *yà* for Topic and *wɛ̀* for Focus. The morphemes occur to the right of the DP.

GUNGBE

(87a)	Ùn	sè	ɖɔ̀	dàn	lɔ́	yà	Kòfí	hù	ì
	1SG	hear	that	snake	DET	TOP	Kofi	kill	3SG

'I heard that, as for the snake, Kofi killed it.'

(87b)	Ùn	sè	ɖɔ̀	dàn	lɔ́	wɛ̀	Kòfí	hù
	1SG	hear	that	snake	DET	FOC	Kofi	kill

'I heard that Kofi killed THE SNAKE.'

[Aboh (2010:26)]

Following Aboh (2010) and Zubizarreta (2010), the structure below represents the syntactic configuration of a typical pragmatically structured proposition. Note that there are two Focus positions in this proposal so that languages will make use of one or another. Zubizarreta is concerned primarily with the low Focus position.

(88) [spec [**Foc** [spec [Infl [spec [**Foc** [VP

According to Zubizarreta (2010:135), Kimatuumbi presents evidence in favor of the view that the low Focus position is selected by the inflectional node Tense, that is, focal distinctions are marked in the tense system. The type of tense may impose particular focusing requirements on the clause, and most importantly, the focused constituent must appear immediately after the verb. This entails that the verb has crossed the Focus head on its way to the tense head T, as can be observed in the syntactic representation below. In this structure the verb root is merged with the low Focus head, and then it is merged with the Tense/Inflection head.

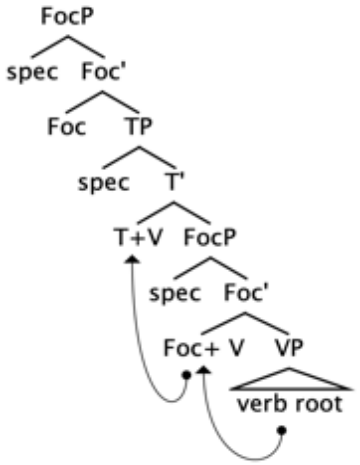


Figure 5: Syntactic structure of the focus realization (Source: Zubizarreta (2010:135))

Regarding prosody, there are two ways of realizing prominence at the post-lexical level, namely (i) culminatively, via main sentence stress (involves associating the rhythmically strongest element of a metrical structure with the Nuclear Pitch Accent – NPA); and (ii) demarcatively, via a prosodic boundary (lengthening, pauses, and/or dedicated phrasal tones). Some languages, such as German make use of the first strategy, whereas Japanese and Korean employ the second process. Languages such as Bengali use both means.

Many Bantu languages clearly belong to the demarcative type of category, according to Zubizarreta (2010:136). In these languages Focus triggers the insertion of a prosodic boundary, namely a right edge prosodic boundary. The author remarks that "in many of the Bantu languages under discussion, the final or penultimate syllable of a phonological phrase (p-phrase) is long. And the one in the last p-phrase is the longest and can therefore be identified with main sentence (or nuclear) stress." (2010:136). That is, the penultimate or last syllable in the last p-phrase is long and may be identified with nuclear stress in a given sentence. However, the author notes that "it is the p-boundary and not nuclear stress which serves to prosodically identify the focused constituent." (2010:136).

Zubizarreta (2010:136) argues that it is useful to distinguish between the Functional domain and the Lexical domain: "a syntactic constituent belongs to the L-domain of lexical head H if it is L-related to H, and it belongs to Fu-domain

if it is Fu-related to functional head F." Thus, XP is L-related to head Y iff XP is a specifier or complement of or adjunct to Y, where Y is a lexical projection YP, whereas XP is Fu-related to function head Fu iff it is a specifier or adjoined to FuP.

Additionally, for languages that make a distinction between weak and strong p-boundaries, weak p-boundaries get inserted at the edge of L-related categories, and strong p-boundaries get inserted at the edge of Fu-related categories. The notation below illustrates this distinction.

(89a) Weak p-boundaries: …XP) YP) ZP) or (XP (YP (ZP…

(89b) Strong p-boundaries: …XP) (YP) (ZP…

Thus, Zubizarreta (2010:137) aims to show that the (low) Focus head triggers the insertion of a strong p-boundary, since it is a functional category. This p-boundary gives rise to a ripple effect such that a strong p-boundary gets copied at the left edge of each phrase that follows the focused constituent. The author discusses data from three Bantu languages. However, in this section I will present only the discussion of Cichewa, which is a language very strongly correlated with Cinyanja.

Zubizarreta (2010:144) argues that in Cichewa p-boundary insertion only happens in the functional domain. In focus-neutral cases no p-boundary appears

within the lexical domain. According to the author, Kanerva (1990) discusses three sandhi-rules that are relevant in identifying the presence or absence of p-boundaries in this language. The rules consist in (i) Penultimate lengthening, (ii) Final H-tone retraction, and (iii) H-tone doubling. The conjunction of (i) and (ii) give rise to a falling (HL) tone contour at the right edge of a p-phrase. On the other hand, (iii) gives rise to a sustained High tone phrase-internally.

The definition of the rules can be stated as follows (2010:144):

(i) *Penultimate lengthening*: a vowel in the penultimate syllable of a p-phrase is lengthened;

(ii) *Final H-tone retraction*: an underlying H-tone is retracted from the final mora of a p-phrase to the penultimate mora;

(iii) *H-tone doubling*: a H-tone is doubled (spread one mora to the right) except when the target of spreading is within the p-phrase-final foot.

The application of the rules above is illustrated by the examples below. Note that Example (98) shows that the verb plus the direct object and the indirect object are all part of the same p-phrase. Example (99) illustrates the H-tone doubling rule within the p-phrases. Example (100) shows that the preverbal subject is located within the functional domain, given the penultimate syllable is

lengthened and the H-tone is retracted. Observe that the verb is located inside another p-phrase, which is indicated by the (V) pattern found in the examples below (Zubizarreta 2010:145).

(90) [IP I- V[VP (V) NP PP]

(anaménya	nyumbá	ndí	mwáála)	Cf. /anaménya/
he-hit	house	with	rock	

'He hit the house with a rock.'

(91) [IP I- V [VP (V) (NPj) PP] NPj]

(Ana-í-ménya	ndí	mwáála)	(nyuúmba)
he-it-hit	with	a rock	the house

'He hit with a rock, the house.'

(92) [IP DP [I-V [VP (V)]]]

(kagaálu)	(kanáafa)	Cf. /kagalú/
(small) dog	died	

'The (small) dog died.'

As in other Bantu languages, in Cichewa Focus also triggers the insertion of a p-boundary. Compare the examples below (2010:145). Notice that in (93) the verb is focused and therefore is contained in a functional domain and followed by a strong p-boundary; the direct object and the indirect object are inside the same p-phrase. On the other hand, in (94) the direct object is focused and thus

located within a p-phrase which also contains the verb. The p-boundary occurs immediately after the focused term.

(93) [What did they do in Mavuto's house?]

(anagóona) (mnyu^{m}bá yá mávúuto)
they-slept in-house of Mavuto
'They SLEPT in the house of Mavuto.'

(94) [What did he hit with the rock?]

(anaménya nyuú^{m}ba) (ndí mwáála)
he-hit house with rock
'He hit THE HOUSE with a rock'

As noted by Kanerva (1990 *apud* Zubizarreta 2010), "the prosodic effects of focus in Cichewa are even more dramatic. Not only is a p-boundary inserted at the right edge of Focus, this p-boundary also triggers the insertion of p-boundaries to the right of the focused constituent." The rule can be stated as follows:

(95) Each L-related constituent that follows the focused constituent is also flanked by p-boundaries.

Zubizarreta calls this phenomenon “ripple effect”. The examples below illustrate this effect (2010:146).

(96) [What did he do to the house with the rock?]

(anaméenya)P (nyuúmba)P (ndí mwáála)P

he-hit house with rock

‘He hit the house with the rock.’

(97) [Just what did he do to the house with the rock yesterday?]

(anaméenya) (nyuúmba) (ndí mwáála) (dzuulo)

Zubizarreta presents the following derivation of example (97) in order to illustrate the phenomenon.

(98) [IP V-Infl [(V) Foc [$_{VP}$ [$_{VP}$ NP (V) PP] Adv]

(Anaméenya) ((nyuúmba ndí mwáála dzuulo)

Foc rule: ↑ nyuúmba) (ndí mwáála) (dzuulo)

Copy: ↑ ↑

Zubizarreta (2010:146) proposes that this p-boundary copying is a product of the Uniformity Condition, which can be stated as follows.

(99) If a PB is inserted at the edge of a constituent C in domain D (e.g. L-domain), then all constituents within that domain must be equally flanked by the same type of PB.

Zubizarreta (2010:147) mentions that, in Cichewa, "overt movement of a focused constituent to the low Focus position is more parsimonious than in other Bantu languages. […] While VP medial XPs, when focused, must move to spec-FocusP, a focused XP in VP-final position in Cichewa stays *in situ*". The example below illustrates this phenomenon.

(100) What did he hit the house with?

(anaménya	nyumbá	ndí	MWÁÁLA)
He-hit	the house	with	*a rock*

'It is with a ROCK that he hit the house.'

Thus, it appears that Cichewa is a language in which the syntactic licensing of a focused constituent may be covert or overt. If the argument is VP medial, the licensing must be overt; i.e. it must move to spec-FocusP. The reason is that this is the only way that it can be prosodically licensed by a SPB (strong p-boundary) to its right. In that position, the Focus rule ensures just that. On the other hand, if the argument is VP-final, the syntactic licensing is (or may be)

covert, because in that position it is already right-flanked by a "strong" boundary, introduced either by the functional domain rule – if the VP is followed by IP-level material, as in the cases with right-dislocated XPs –, or otherwise by the Closure Convention. In Cichewa an in-situ VP medial phrase cannot be prosodically identified as focused, but a VP-final phrase can. Therefore, the former but not the latter must move overtly to Focus position (Zubizarreta 2010:147).

One of the goals of the analysis to be developed in Chapter 5 of the present dissertation is to assess the properties of focused constituents in Cinyanja, that is, if the same rules that occur in Cichewa, according to Zubizarreta, also apply to the Cinyanja variant.

Based on Zubizarreta's proposal outlined above, my aim in this dissertation is to show that, in Cinyanja, the focused DP bears a high floating tone and is flanked by a p-boundary triggered by the focus category. The sandhi-rules noticed by Kanerva (1990) regarding Cichewa may also apply to Cinyanja, that is, there occurs penultimate lengthening, high tone retraction, and high tone doubling or spread.

Based on this discussion I will propose that there is a dedicated syntactic position for focused arguments, and that these arguments are marked prosodically via processes such as p-boundary insertion and H-tone doubling and retraction.

The example below illustrates the phenomena occurring in Cinyanja. Notice that the focused argument *osáka* 'hunter' is flanked by a p-juncture to the right and bears a high tone in the penultimate syllable. My proposal is that the prosodic effects are due to the presence of a "silent" copula which is manifested through a high tone.

(101) (Who caused the lion to kill the zebra?)

Osááka, waphélétsa mbídzi ndi nkángo.

osaka	*w-a-ph-el-ets-a*	*mbidzi*	*ndi*
1-hunter	1SM-PAST-kill-APPL-CAUS-FV	9-zebra	by

nkango
3-lion
'(It was) the hunter caused the lion to kill the zebra.'

In this section I presented an account of prosodic marking of focused arguments in Bantu languages. In the next section I will address the relation that may exist between information structure and morphosyntax that pervades across languages. This topic will be explored in detail in the following chapters.

3.6: ON THE RELATION BETWEEN INFORMATION STRUCTURE AND MORPHOSYNTAX

As previously mentioned, according to Rizzi (1997), the inclusion of pragmatic properties in syntax is necessary because they contain important information that should be considered during the derivation process. Rizzi's cartography offers a framework for mapping out the pragmatic aspects of language, such as clause types (affirmative, interrogative, etc.), relative pronouns, and topic and focus dynamics, within the left periphery. By "exploding" the complementizer layer (CP) into separate projections for Force, Topic, and Focus, Rizzi's proposal allows for a more in-depth analysis of various phenomena related to the C system. This includes examining adjacency and non-adjacency effects involving elements of the C system, as well as the different types of fillers in the subject position (overt DP, PRO, pro, trace/copy). As previously noted, Rizzi argues that A' movement within the C system is driven by Criteria, or interpretive features (WH, Negation, Focus, Topic, etc.) that have an impact on the interpretation of the category carrying them and its constituents.

According to Aboh (2010), who builds his work on Rizzi's analysis, the starting point of information structure is at the numeration. Aboh's conclusions are based on the analysis of various languages, with a focus on Fongbe and Gungbe. In addition, Aboh examines phenomena like question-answer pairs and overt particles that express Topic and Focus functions. It is argued that discourse-related lexical items, which drive the derivation, enter the computation at the

numeration step and encode information structure properties. These particles, which encode features such as Force, Topic, and Focus, project functional categories such as ForceP, TopicP, and FocusP in the syntax. Similar to optional formal features like Case features, these features are added arbitrarily when the lexical item enters the numeration. According to this perspective, the interaction between information structure and syntax occurs indirectly through PF and possibly LF. The author suggests that the selection of a linguistic expression is influenced by discourse factors outside of syntax. However, it is important to note that the chosen expression retains precise syntactic properties. Therefore, information structure properties should be derived by the computation.

Thus, I assume in the present work that information structure-related formal features enter the derivation via Numeration and must be then manipulated by the computation in order to derive the linguistic expression with all its properties.

Information structure can be marked by syntactic (102), morphological (103), and phonological/prosodic (104) means, as can be observed in the following examples.

ENGLISH

SYNTACTIC MEANS

(102a) John broke the vase.
(102b) It was John that broke the vase.
(102c) It was the vase that John broke.

PRONOMINALIZATION

(103a) John broke the vase.
(103b) He broke the vase.
(103c) John broke it.

PROSODIC MARKING

(104a) John broke the vase.
(104b) JOHN broke the vase.
(104c) John broke the VASE.

The same strategies, namely syntactic changes, pronominalization, and prosodic marking can be employed in Cinyanja. See the examples below. (105) illustrates the syntactic changes that may occur, such as passivization and cleft constructions; (106) illustrates the pronominalization; and (107) shows prosodic marking.

CINYANJA

SYNTACTIC MEANS

(105a) *Jowáwo waphwánya m'phíka.*

Jowawo	*w-a-phwany-a*	*m'phika*
1-John	1SM-PAST-break-FV	3-vase

'John broke the vase.'

(105b) *Ndí Jówáwo améne waphwánya m'phíka.*

ndi *Jowawo* *a-mene* *w-a-phwany-a* *m'phika*
COP 1-John 1-COMP 1SM-PAST-break-FV 3-vase
'It was John who broke the vase.'

(105c) *Ndí m'phíka uméne Jowáwo wáphwánya.*

ndi *m'phika* *u-mene* *Jowawo* *w-a-phwany-a*
COP 3-vase 3-COMP 1-John 1SM-PAST-break-FV
'It was the vase that John broke.'

PRONOMINALIZATION

(106a) *Jowáwu waphwánya m'phíka.*

Jowawo *w-a-phwany-a* *m'phika*
1-John 1SM-PAST-break-FV 3-vase
'John broke the vase.'

(106b) *Iye waphwánya m'phíka.*

iye *w-a-phwany-a* *m'phika*
he 1SM-PAST-break-FV 3-vase
'He broke the vase.'

(106c) *Jowawu wawúphwánya.*

Jowawu *w-a-wu-phwany-a*
1-John 1SM-PAST-3OM-break-FV
'John broke it.'

PROSODIC MARKING

(107a) *Jowáwu waphwánya m'phíka*.

Jowawo *w-a-phwany-a* *m'phika*
1-John 1SM-PAST-break-FV 3-vase
'John broke the vase.'

(107b) *JÓWÁWU, waphwanya m'phika.*

Jowawu	*w-a-phwany-a*	*m'phika*
1-John	1SM-PAST-break-FV	3-vase

'JOHN broke the vase.'

(107c) *Jowáwu waphwánya, M'PHÍKA*.

Jowawu	*w-a-phwany-a*	*m'phika*
1-John	1SM-PAST-break-FV	3-vase

'John broke the VASE.'

The data above illustrate the different strategies that natural languages may employ to mark focus and topic functions. In the present dissertation I will explore the syntactic and prosodic means of marking focus in Cinyanja. I will propose that syntactically there is a dedicated position for marking focus, namely the Immediately After the Verb position, henceforth IAV position. Regarding prosody, I will propose that there is a "silent" copula which is the source of the high tone borne by the focused argument.

Next, I will investigate the relation between information packaging and morphosyntax involved in the realization of focused arguments. The next section will deal with the methodology of question-and-answer pairs.

3.7: ON QUESTION-AND-ANSWER PAIRS

Aboh (2010) argues that answer-question pairs show that there is an interaction between information structure and syntactic structure. The form of a question will determine the particular form of the answer. For instance, see the examples from Chinese (Li & Thompson 1975):

(108a) Yàoshi ne?
key INTER
'what about the key/s?'

(108b) ??Wǒ wàng le yàoshi
I forgot ASP key/s
'I forgot the key/s'

(108c) Yàoshi, wǒ wàng le
key/s I forgot ASP
'about the key/s, I forgot'
[topicalization]

(108d) Wǒ bǎ yàoshi wàng le.
I ba key/s forgot ASP
'I forgot the key/s'
[bǎ = discourse anaphor]

(108e) Wǒ wàng le
I forgot ASP
'I forgot (it/them)'
[topic ellipsis]

As the data above reveal, there must be some kind of correspondence between the question and the answer. In this case, the argument *yàoshi* 'key/s' is part of the presupposition in the question, therefore, in the answer the argument

must be realized via topicalization, ellipsis, or differential marking. Thus, question-and-answer pairs constitute a useful methodology of collecting data on the interaction between form and informational function, and on pragmatically structured propositions – especially topic-commentary and focus-presupposition.

In Cinyanja the form of the answer also corresponds to the form of the question. See the data below. Example (109) corresponds to a structure with default interpretation, that is, it is a proposition that is not pragmatically structured.

(109) *M'nyamata walumpha m'panda.*

m'nyamata	*wa-lumph-a*	*m'panda*
1-boy	1SM-jump-FV	3-fence

'The boy jumped the fence.'

Examples in (110), (111) and (112) correspond to question-and-answer pairs, in which the answer corresponds to a pragmatically structured proposition. In (110a) the agent argument *m'nyamata* 'boy' is part of the presupposition, and the answer (110b) consists in a predicate focus construction (topic-commentary).

(110a) *Cacitika ndi ciyani kwa m'nyamata?*

cacitika	ndi	ciyani	kwa	m'nyamata?
happened	COP	what	to	1-boy

'What happened to the boy?'

(110b) *M'nyamata walumpha m'panda.*

m'nyamata	wa-lumph-a	m'panda
1-boy	1SM-jump-FV	3-fence

'The boy jumped the fence.'

In (111a) the affected argument *m'panda* 'fence' is part of the presupposition, which also elicits a topic-commentary (predicate focus) construction in the answer (111b).

(111a) *Cacitika ndi ciyani ndi m'panda?*

cacitika	ndi	ciyani	ndi	m'panda?
happened	COP	what	to	3-fence

'What about the fence?'

(111b) *M'panda walumphidwa ndi m'nyamata.*

m'panda	wa-lumph-idw-a	ndi	m'nyamata
3-fence	3SM-jump-PASS-FV	by	1-boy

'The fence was jumped by the boy.'

In (112) there occurs an open question in which the agent argument is interrogated. The answer corresponds to a focus-presupposition construction, that is, the focus domain is the constituent.

(112a) *Ndani walumpha m'panda?*

nd(i)-ani	wa-lumph-a	m'panda?
COP-who	1SM.PAST-jump-FV	3-fence

'Who jumped the fence?'

(112b) *M'panda walumphidwa ndi n'nyamata.*

m'panda	wa-lumph-idw-a	ndi	n'nyamata
3-fence	3SM.PAST-jump-PASS-FV	by	1-boy

'The fence was jumped by the boy.'

In the examples above all the affirmative sentences (109), (110b), (111b) and (112b) have the same truth-value, represented below.

(113) $(\exists e)\ \text{jump}(e)\ \&\text{agent}(e, \text{boy})\ \&\ \text{theme}(e, \text{fence})$

The interpretation of these sentences is that there is an event of jumping, and two participants: the argument that performs the activity, and the argument which is somewhat "affected" by the event. However, the form of the answer may vary according to the context – the question. Thus, in (110b) we have the default SVO word order in which the agent argument functions as the topic and is realized as the subject. The theme argument figures as the object. However, in the examples (111b) and (112b), the theme argument occurs as the subject, and the passive extension appears in the verbal complex. This change happens because in questions (111a) and (112a) the theme argument is part of the presupposition and thus must occur in topic position in the answer. These

questions elicit the theme argument as the topic (111a) and the agent argument as the focus (112a).

The data above illustrate how the question-and-answer pairs provide the context for observing changes in the realization of the syntactic and morphological structure of a pragmatically structured proposition.

CHAPTER SUMMARY

In this chapter I explored the theoretical framework assumed in the present work. This dissertation is based on the Derivation by Phase model of grammar, which was proposed by Chomsky in the 1990-2000's. My work is also based on the assumptions of Distributed Morphology, in which underspecified roots and bundles of features are selected from the narrow lexicon, then enter the computation, and are finally supplied with phonological and morphological features after spell-out via correspondence rules.

I then discussed the notion of argument structure. I adopted Gallego's (2015) proposal of predicates as elements with uninterpretable features that must be satisfied by relations with the elements that can value these features, namely he arguments. I also presented the semantic analyses of this phenomenon, focusing on three levels of representation: predicate decomposition, thematic

roles, and Heim and Krazter's (1998) proposal of the semantic operations Event Identification and Functional Application, which are involved in argument realization.

Furthermore, I showed an account of Bantu agreement and word order patterns that are accounted for by Carstens (2005) as being the result of the operation AGREE via spec-head, as well as the presence of φ-features and EPP-features in C and T.

This work is also based on the assumptions of Lambrecht (1994), Rizzi (1997) and Aboh (2010), among others, regarding information structure notions. I presented Aboh's (2010) proposal that topic and focus features enter the derivation via numeration, being associated with arguments just like formal Case features. The author also proposes that Topic and Focus project the functional categories TopicP and FocusP in the course of the computation.

Finally, I briefly presented an overview of the methodology of question-and-answer pairs, and how the form of the question will motivate the form of the answer.

In the next chapter I will discuss data from the Nyanja language in order to assess their morphosyntactic and informational properties.

CHAPTER 4

THE POSITION OF FOCUS IN CINYANJA

The aim of this chapter is to present and discuss the empirical data that will serve to support the theoretical proposal, which will be discussed in chapter 5. My goal is to present how focused arguments are grammatically encoded in Cinyanja. More specifically, the aim is to describe the syntactic positions, in which focused arguments may occur in a given pragmatically structured proposition. Additionally, another purpose is to examine how Cinyanja fits in the typology proposed by Gibson et al (2017), as was previously shown in section 3.5.3. This discussion is important due to the intense debate in Bantu literature regarding the position of focus in languages of this family.

In this dissertation I will propose that identificational constituent focus is uniformly marked via prosodic and syntactic means in the Nyanja language. Morphologically, there is no marking on the term nor on the verb; there is no dedicated particle/morpheme to encode focus; and there is no conjoint-disjoint alternation either. Syntactically, in some cases, pragmatically structured propositions seem to be realized as the default SVOX structure, according to the thematic hierarchy; in other cases, strategies such as canonical cleft and

pseudocleft constructions are overtly employed by the speakers. Regarding prosodic marking, the focused argument bears a high tone and there occurs the insertion of a p-boundary at the edge of the focused constituent.

In sum, this chapter aims to present empirical data in order to support the hypothesis that there is a dedicated focus position in this language, namely the IAV position. In order to achieve this goal, I will present sets of question-and-answer pairs from a range of contexts. The open questions interrogate arguments in a given situation and the answers correspond to pragmatically structured propositions in which the focus domain is the DP constituent. Since the form of the question will determine the form of the answer, as argued by Aboh (2010) (cf. section 3.7), I employ this methodology in order to present the data which are relevant for the theoretical proposal I will develop in Chapter 5.

This chapter is organized into three sections. Section 4.1 discusses the status of the interrogative words employed by Cinyanja speakers to generate open questions. Section 4.2 presents the prosodic strategies used to mark constituent focus. Section 4.3 presents the syntactic strategies used to encode constituent focus.

The aim of the next section is to assess the status of interrogative words which occur in open questions in Cinyanja.

4.1 INTERROGATIVE WORDS

This section aims to investigate the distribution of question words in Cinyanja in order to present a brief account of the status of these interrogative operators. It is observed that the language in question makes use of several interrogative words (or operators), that is, functional words employed in asking open questions. These operators are variables which are bound to the relevant element of the answer.

According to Macalane (2013) and Biriate (2023), the interrogative words are constructed via the combination of bound pronominal morphemes with a variety of prefixes in order to refer to several entities, namely person, thing, place, manner, reason, time, and quantity. Consider the table below, which illustrates these morphemes and their referent.

Pronominal form	**Referent**
-ni ‘who’ / -ti ‘whom’	Person
-ni ‘what’ / -nji ‘what’	Thing
-nji ‘how’; ‘in what way’	Manner
-ti ‘where’	Place
-nji ‘for what reason’	Reason

-ti ‘when’; ‘in which day’ / -nji ‘when’	Time
-ngati ‘how many’	Quantity

Table 6: interrogative pronouns and their semantics (Macalane, 2013)

The table below illustrates the derived words used by the speaker who provided the empirical data I will discuss in this dissertation.

Interrogative word	**Referent**
ndáni	who / whom
ciyáni	what
kúti	where
motani	how
líti	when
cifukwa ciyáni	why (reason)

Table 7: interrogative words in Cinyanja

As can be seen in the examples below, in Cinyanja the interrogative words can occur either in sentence-initial or in sentence-final position. Thus, both questions (114a)-(114b) have the same interpretation.

(114a) *Ndáni bilidála wamangíla nyúmba?*

nd(i)-ani *bilidala* *w-a-mang-il-a* *nyumba*
COP-whom 1-bricklayer 1SM-PAST-build-APPL-FV 3-house
'(It was) for whom did the bricklayer build the house?'

(114b) *Bilidála wamangíla nyúmba ndáni?*

bilidala *w-a-mang-il-a* *nyumba* *nd(i)-ani*
1-bricklayer 1SM-PAST-build-APPL-FV 3-house COP-whom
'The bricklayer built the house for whom (it was)?'

Observe another instance in which the position of the interrogative word may vary between sentence-initial (115a) and sentence-final (115b). In both open questions below the affected argument is interrogated.

(115a) *Ciyáni n'nyámáta walúmpha?*

ciyani *n'nyámáta* *w-a-lúmph-a*
what 1-boy 1SM-PAST-jump-FV
'What did the boy jump?'

(115b) *N'nyámáta walúmpha ciyáni?*

n'nyámáta *w-a-lúmph-a* *ciyáni*
1-boy 1SM-PAST-jump-FV what
'Did the boy jump what?'

The interrogative words delineated above can optionally occur with the copula *ndí*, except for *ndáni* 'who/whom'. Based on Macalane's (2013) description of the derivation of the interrogative pronouns and on Duarte's

insights (personal communication), I propose that the word *ndáni* is the grammaticalized form of the syntactic object {*ndí*+*ani*}, as illustrated below.

(116) ndí + ani → nd(í)ani → ndáni

The open question below (117) corresponds to an example in which the copula *ndí* occurs immediately before the interrogative word *kúti* 'where'. This structure resembles a cleft construction, except for the fact that the complementizer is not realized.

(117) *Ndí kúti mwána wang'ámba kábúdúla?*

ndi	*kuti*	*mwana*	*w-a-ng'amb-a*	*kabudula*
COP	where	1-child	1SM-PAST-tear-FV	5-shorts

'Where it was (that) the child tore the shorts?'

The copula may optionally occur with other interrogative words, such *lítí* 'when', as seen below.

(118) *(Ndí) lítí páka waphwánya m'phíka?*

ndi	*liti*	*paka*	*w-a-phwany-a*	*m'phika*
COP	when	1-cat	1SM-PAST-break-FV	3-vase

'(It was) when did the cat break the vase?'

The copula *ndí* may also optionally appear when the interrogative word occurs in sentence-final position, as can be observed in the example below.

(119) *Bilidála wamangíla anéba ndí ciyáni?*

bilidála	*w-a-mang-il-a*	*aneba*	*ndi*
1-bricklayer	1SM-PAST-build-APPL-FV	1-neighbor	COP

ciyáni
what
'What was it that the bricklayer build for the neighbor?'

Thus, the properties of the interrogative pronouns in Cinyanja may be summarized as follows:

i. Interrogative words are constructed via the combination of a bound morpheme and affixes in order to refer to a number of entities, such as person, thing, local, time, manner, reason, and quantity;

ii. Syntactically, the interrogative words may occupy the following positions: sentence-initial sentence-medial, and sentence-final;

iii. The copula *ndí* may optionally appear before interrogative words, except for *ndáni* 'who/whom', which I propose is the result of amalgamation of the copula and the particle {*-ani*}.

In this section I present a brief account of the status of interrogative words employed by Cinyanja speakers to mark focus. In the next section I will discuss the prosodic strategies of encoding constituent identificational focus in the language in question.

4.2: PROSODIC STRATEGIES OF ENCODING IDENTIFICATIONAL FOCUS MARKING IN CINYANJA

Regarding prosodic marking of constituent focus, two effects take place: (i) the focused argument bears a high tone which may spread up to the penultimate vowel of the constituent; and (ii) a p-boundary is inserted flanking the focused expression. Thus, my hypothesis is that, prosodically, Cinyanja behaves just as proposed by Zubizarreta (2010) and Duarte and Langa (2025) (cf. section 3.5.4).

Consider the examples below. In sentence (120) the affected argument *mápázi* 'feet' is focused. The underlying tone of this noun in isolation is LLL, but in this pragmatically structured proposition it changes into a HHL tone pattern. There also occurs the insertion of a p-boundary flanking the focused constituent to the right, as indicated by the parentheses and comma, which represent a pause.

(120) (What did the girl soak in the sea?)

Ntsíkána, wanyówétsa MÁPÁZI, kúnyánja.

ntsíkána)	*w-a-nyow-ets-a*	*mapaazi)*	*ku-nyanja*
1-girl	1SM-PAST-soak-CAUS-FV	6-feet	LOC-sea

'The girl soaked her FEET in the sea.'

Applying Zubizarreta's (2010) analysis, the prosodic structure of this sentence is given below. The parentheses indicate the insertion of a p-boundary, which is manifested via a pause.

(121) (ntsíkána) (wanyowetsa MÁPÁZI) (kunyánja)

girl she-soaked her-feet in the sea

'The girl soaked her FEET in the sea.'

Observe now the example below. In this pragmatically structured proposition, the theme argument *mípíla* 'balls' is focused. This noun bears a LLL underlying tone pattern, but in sentence (122) it surfaces as HHL. Moreover, a p-boundary is inserted to the right of the focused constituent.

(122) (What did the boy kick in the soccer field?)

N'nyámáta, waménya MÍPÍLA, kugaláwúndi lá m'píla.

n'nyamata)	*w-a-meny-a*	*mipila)*
1-boy	1SM-PAST-kick-FV	4-balls

kugalawundi la m'pila
LOC-soccer field
'The boy kicked the BALLS in the soccer field.'

The prosodic representation of the constituents according to Zubizarreta (2010) is shown below.

(123)	(N'nyámáta)	(wáménya	MÍPÍLA)	(kúgáláwúndí lá m'píla)
	boy	he-kicked	balls	in the soccer field

'The boy kicked the BALLS in the soccer field.'

Observe now another example of a pragmatically structured proposition in which the agent argument is interrogated. This argument is realized as the subject of the sentence, according to the thematic hierarchy. In addition, the underlying tone of this constituent is LHL, but it surfaces as HHL in the answer. Furthermore, a p-boundary is also inserted to the right of this argument.

(124) (Who is scattering the books around the house?)

MPHÚNZÍTSI, akumwáza mabúkhu, munyúmba.

mphunzitsi)	*a-ku-mwaz-a*	*ma-bukhu)*	*mu-nyumba*
1-professor	1SM-PRES-scatter-FV	6-books	LOC-house

'The PROFESSOR is scattering the books around the house.'

The prosodic structure of this sentence is given below.

(125) (mphúnzítsi) (akumwáza mabúkhu) (munyúmba)
professor he-is-scattering the books around the house
'The PROFESSOR is scattering the books around the house.'

I propose in Chapter 5 that the high floating tone observed in the pragmatically structured propositions I presented above is the manifestation of a "silent" grammaticalized copula. Furthermore, the copula may appear segmentally as well, as seen in the examples below. Notice the high tone borne by the focused constituent in (126) below.

(126a) (What did the girl soak in the sea?)

Ntsikána, ányówétsa MÁPÁZI, kunyanja.

ntsikana)	*a-nyow-ets-a*	*mapazi)*	*ku-nyanja*
1-girl	1SM.PAST-soak-CAUS-FV	6-feet	LOC-sea

'The girl soaked (it was) her feet in the sea.'

(126b) *Ntsíkána, áyówétsa kunyanja, ndí MÁPÁZI.*

ntsikana)	*a-nyow-ets-a*	*kunyanja)*	*ndí*
1-girl	1SM.PAST-soak-CAUS-FV	LOC-sea	COP

mapazi
6-feet

‘The girl soaked it was her feet in the sea.’

In this section I presented the prosodic marking employed by speakers of Cinyanja to realize focused constituents. These effects can be summarized as follows.

i. Prosodic marking (on the constituent):
    a. High tone which spreads to the penultimate syllable
    b. P-boundary insertion

In this section I have shown data that illustrate the prosodic pattern associated with constituent focus marking in the Nyanja language. The goal of the next section is to present the syntactic strategies of marking focused constituents in pragmatically structured propositions.

4.3: SYNTACTIC STRATEGIES OF ENCODING CONSTITUENT FOCUS IN CINYANJA

In this section I investigate the syntactic strategies that are employed by the speakers of Cinyanja to encode identificational constituent focus. Two possible surface constructions emerge: (i) the default SVOX structure (in which

the copula and the complementizer are not segmentally realized), or (ii) an explicit cleft or pseudocleft construction (in which the copula and possibly the complementizer are segmentally realized). My proposal is that the SVOX order is only apparent and corresponds to a cleft or pseudocleft construction, given the prosodic effects of the grammaticalized copula on the focused constituent. I will begin with the canonical cleft and pseudocleft constructions.

4.3.1: CLEFT AND PSEUDOCLEFT CONSTRUCTIONS

Canonical clefts and pseudoclefts are used by speakers of Cinyanja to mark focused arguments in identificational constituent focus constructions. In these structures, the focused argument is realized immediately after a copular verb. Speakers of Cinyanja make use of two constructions, namely (i) the canonical cleft, and (ii) the pseudocleft. The canonical cleft consists in a construction in which the copula occurs in sentence-initial position licensing the focused constituent. The complementizer *méne* 'that' may be overt or covert, but when it is realized phonetically, it enters into an agreement relation with the focused argument. On the other hand, in pseudoclefts the complementizer may optionally occur overtly in sentence-initial position, whereas the syntactic object COPULA + FOCUSED ARGUMENT occurs in sentence-final position.

In order to illustrate the cleft and pseudocleft constructions, consider the following situation, denoted by a presentational sentence. In this construction there occurs the default SVOX word order.

(127) *Mbidzi yadya chíre kunkhalángo.*

mbidzi	*y-a-dy-a*	*chíre*	*ku-nkhalango*
9-zebra	9SM-PAST-eat-FV	?-grass	LOC-savannah

'The zebra ate grass in the savannah.'

Observe now an open question, in which the agent argument *mbidzi* 'zebra' is interrogated. In this construction, the copula *ndí* may optionally occur in sentence-initial position, immediately before the interrogative word.

(128) *(Ndí) ciyáni cádya chíre kunkhalángo?*

ndi	*ciyani*	*c-a-dy-a*	*chire*	*ku-khalango*
COP	what	7SM-PAST-eat-FV	5-grass	LOC-savannah

'(It was) what (that) ate the grass in the savannah?'

There are two possible answers to question (128). These consist in a canonical cleft (129) and a pseudocleft (130). In the canonical cleft (129) the syntactic object *ndí mbídzi* 'it was the zebra' occurs in sentence-initial position. On the other hand, in the pseudocleft (130), the copula and the focused constituent occur in sentence-final position. Notice that, in example (129), the

complementizer, which follows the focused argument, is optional; when it is overt then the noun class prefix of the focused constituent is merged with the complementizer, in a spec-head AGREE relation. Furthermore, in the pseudocleft (130) the complementizer may optionally appear in sentence-initial position. It is interesting to note that, in both (129) and (130), the focused constituent *mbídzi* 'zebra' surfaces with a HL tone pattern, even though the underlying tone of this argument is LL. This situation indicates that the high tone of the copula *ndí* spreads up to the penultimate syllable of the phonological phrase in which the focused argument is located. Additionally, in both constructions this focused argument is flanked by a p-boundary that marks the syntactic-phonological domain to which this focused DP is moved.

(129) *Ndí MBÍDZI, (imene) yadya chíre kunkhalángo.*

ndi	*mbidzi)*	*i-mene*	*y-a-dy-a*	*chire*
COP	9-zebra	9-COMP	9SM-PAST-eat-FV	5-grass

ku-nkhalango
LOC-savannah
'It was the ZEBRA that ate grass in the savannah.'

(130) *(Imene) yadya chíre kunkhalángo, ndí MBÍDZI.*

i-mene	*y-a-dy-a*	*chire*	*ku-nkhalango)*
9-COMP	9SM-PAST-eat-FV	5-grass	LOC-savannah

ndi *mbidzi*

COP 9-zebra
‘(What) ate the grass in the savannah it was the ZEBRA.’

Consider now another situation, which consists in a reason/purpose applicative construction, as seen in sentence (131) below.

(131) *Agogo aphikíla nyáma phwando.*

agogo	*a-phik-il-a*	*nyama*	*phwando*
1-old.woman	1SM.PAST-cook-APPL-FV	7-meat	5-party

‘The old woman cooked the meat for the party.’

The agent argument *agogo* ‘old woman’ may be interrogated, as in the question below.

(132) *Ndáni aphikíla nyáma phwando?*

nd(i)-ani	*a-phik-il-a*	*nyama*	*phwando*
COP-who	1SM.PAST-cook-APPL-FV	7-meat	5-party

‘Was it who (that) cooked the meat for the party?’

There are two possible answers to this question. Sentence (133) corresponds to a canonical cleft construction. Conversely, sentence (134) consists in a pseudocleft. In both sentences the complementizer is optional. In addition, the underlying tone pattern of the agent argument *agogo* ‘old woman’ is LLL, but it surfaces as HHL in both answers, owing to the tone rule spreading that applies

whenever a focused DP is moved to the focus position. Moreover, the focused constituent is flanked by a p-boundary in both (133) and (134), as follows.

(133) *Ndí ÁGÓGO, (améne) aphikíla nyáma, phwando.*

ndi *agogo* *a-mene* *a-phik-il-a*
COP 1-old.woman 1-COMP 1SM.PAST-cook-APPL-FV

nyama *phwando*
7-meat 5-party
'It was the OLD WOMAN (that) cooked the meat for the party.'

(134) *(Améne) waphika aphikíla nyáma phwando, ndí ÁGÓGO.*

a-mene *w-a-phik-il-a* *nyama* *phwando*
1-COMP 1SM-PAST-cook-APPL-FV 7-meat 5-party

ndi *agogo*
COP 1-old-woman
'(Who) cooked the meat for the party it was the OLD WOMAN.'

Observe now a question in which the reason argument *phwando* 'party' is interrogated by the use of the interrogative expression *cifúkwa ciyáni* 'for what reason', as shown in (135).

(135) *(Ndí) cifúkwa ciyáni agogo aphikíla nyáma?*

ndi *cifukwa* *ciyani* *agogo* *a-phik-il-a*
COP reason what 1-old.woman 1SM.PAST-cook-APPL-FV

nyama

7-meat
'For what reason did the old woman cook the meat?'

Sentences (136)-(137) below are both natural answers to the question above. The examples below illustrate the surface structure of the pseudocleft (136) and the canonical cleft (137). Notice the same prosodic and syntactic patterns occurring in these data, that is, the tone pattern of the focused argument is changed from LL to HL and a p-boundary is inserted to the left of this constituent, as well as the optional segmental presence of the copula *ndí* (in (136)) and the complementizer (in both sentences). When the copula and the complementizer are not segmentally realized in (136) then the default SVOX order surfaces. I propose, however, that this situation is only apparent, given the spread of the high floating tone to the penultimate syllable of the focused constituent and the agreement relation established between the overt complementizer and the focused argument, manifested by the prefix *li-* being merged with the complementizer *méne*. Hence, my proposal is that sentence (136) is syntactically structured as a pseudocleft, even when the copula and complementizer are not realized segmentally.

(136) *(Liméne) agógo aphikíla nyáma, (ndí)* PHWÁNDO.

(li-mene)	*agogo*	*a-phik-il-a*	*nyama*
5-comp	1-old.woman	1SM.PAST-cook-APPL-FV	7-meat

(ndi) *phwando*
COP 5-party
‘(What) the old woman cooked the meat it was for the PARTY.’

(137) *Ndí PHWÁNDO, (liméne) agogo aphikíla nyáma.*

ndi *phwando* *li-mene* *agogo*
COP 5-party 5-COMP 1-old.woman

a-phik-il-a *nyama*
1SM.PAST-cook-APPL-FV 7-meat
‘It was for the PARTY (that) the woman cooked the meat.’

Now consider another example of a presentational sentence, in which the transitive verb is modified by a locative phrase. Notice that the tone pattern of the agent argument *mphunzítsi* ‘professor’ is LHL when out of context.

(138) *Mphunzítsi akumwáza mabúkhu munyúmba.*

mphunzitsi *a-ku-mwaz-a* *mabukhu* *mu-nyumba*
1-professor 1SM-PRES-scatter-FV 6-books LOC-house
‘The professor is scattering the books around the house.’

The agent argument *mphunzítsi* ‘professor’ may be interrogated by the use of the interrogative word *ndáni* ‘is who’, as shown in the open question below.

(139) *Ndáni akumwáza mabúkhu munyúmba?*

ndani *a-ku-mwaz-a* *mabukhu* *mu-nyumba*
who 1SM-PRES-scatter-FV 6-books LOC-house
'Who is scattering the books around the house?'

The examples below correspond to two possible answers to question (139) above. Notice that, in both pragmatically structured propositions, the tone pattern borne by the focused constituent changes to HHL and a p-boundary (pause) is inserted flanking the focused argument. In sentence (140) the copular verb *ndí* and the complementizer *méne* are both optional; when they are not segmentally realized then the default SVOX order surfaces. However, when the complementizer is overtly realized then it enters an agreement relation with the focused constituent, as shown by the noun class 1 prefix {*a-*} merged with the complementizer *méne*. On the other hand, in (141) the syntactic object *ndí mphúnzítsi* 'is the professor' appears in sentence-final position, which corresponds to a pseudocleft construction. In this context the complementizer may optionally appear in sentence-initial position.

(140) (*Ndí) mphúnzítsi, (améne) akumwáza munyúnba mabúkhu.*

ndi *mphunzitsi)* *amene* *a-ku-mwaz-a* *mu-nyumba*
cop 1-professor comp 1sm-pres-scatter-fv loc-house

mabukhu
6-books
'(Is the professor (that) is scattering the books around the house.'

(141) *(Améne) akumwáza munyúnba mabúkhu, ndí mphúnzítsi.*

amene	*a-ku-mwaz-a*	*mu-nyumba*	*mabukhu*
COMP	1SM-PRES-scatter-FV	LOC-house	6-books

ndi *mphunzitsi*
cop 1-professor
‘(who) scattering around the house

Hence, based on the data above, I argue that, in both canonical clefts and pseudoclefts, there is a dedicated syntactic position for focused constituents, i.e. the IAV position. I also argue that, even in situations in which the default SVOX order appears the focused constituent is moved into the specifier of a low Focus projection in the matrix clause of the cleft/pseudocleft, as shown in the diagram below.

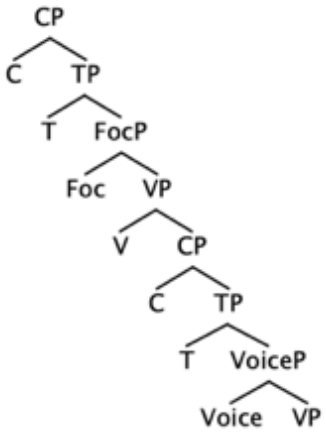


Figure 6: Syntactic structure of a pragmaticlly structured proposition in Cinyanja

Prosodically, the high tone borne by the copula *ndí* spreads to the focused argument, and a p-boundary is inserted to the right or left of the focused constituent. However, as the attentive reader might have noticed, the high floating tone spreading rule only applies when the focused XP equates with identificational focus, but not in contexts such as interrogative words in open questions. This prediction is confirmed by questions (142)-(143) below, insofar as the interrogative operators *ciyáni* and *cifúkwa* do not bear a high tone in the first syllable even though they follow the copula *ndí*.

(142) *Ndí ciyáni n'nyámáta walúmpha?*

ndi	*ciyani*	*n'nyámáta*	*w-a-lúmph-a*
COP	what	1-boy	1SM-PAST-jump-FV

'(It was) what did the boy jump?

(143) *Ndí cifúkwa ciyáni mulími adagwilizitsíra nkhúmba?*

ndi	*cifukwa*	*ciyani*	*mulimi*	*a-dagwiliz-its-il-a*
COP	reason	what	1-farmer	1SM.PAST-stab-CAUS-APPL-FV

nkhumba
9-pig
'(It was) for what reason did the farmer stab the pig?

The agreement pattern between the focused argument and the overt complementizer is another piece of evidence that the focused DP moves from the thematic position to the border of the phase heads Voice and C of the subordinate clause c-selected by the copula *ndí*. More to the point, my proposal is that there occurs cyclical movement of the focused constituent into the specifier of the low Focus projection of the matrix clause, which is headed by the copular verb. More details of this derivation will be given in Chapter 5.

In this section I have discussed the usage of cleft and pseudocleft constructions, both in identificational and informational Focus constructions. The aim of the next section is to present examples in which the default SVOX order appears. The proposal I will advocate is that the focused constituent is moved into the specifier of low FocusP even when the copular verb and the complementizer do not occur segmentally.

4.3.2: FOCUSED ARGUMENT IN IAV POSITION

In Cinyanja, focused arguments may apparently remain *in situ*, including subjects, objects, and locatives. However, as will become evident below, these constructions emerge when the subordinated predicate is raised to a position before the copula *ndí*. A derivational analysis of how this word order is achieved

will be detailed in Chapter 5. In these contexts, the focused constituent is marked prosodically, via the spreading of a high floating tone and the insertion of a p-boundary. The copular verb may optionally appear segmentally. Consider the example below, which consists of a presentational sentence with a benefactive applicative[14].

(144) *Nkángo waphéla mbidzi osáka.*

nkango	*w-a-ph-el-a*	*mbiidzi*	*osaka*
3-lion	3SM-PAST-kill-APPL-FV	9-zebra	2-hunters

'The lion killed the zebra for the hunters.'

Notice that the patient argument *mbíidzi* 'zebra' may be interrogated by means of a cleft construction, as in the question below. In such a context, the copula *ndí* may occur immediately before the interrogative pronoun *ciyáni* 'what'. Even though the copular verb is realized segmentally, the high tone does not spread to the first syllable of the interrogative operator. As was mentioned before, the high tone spreading rule only applies in identificational focus contexts but not in informational focus constructions. In this regard, observe the example below.

[14] As seen in section 2.3, in applicative constructions it is expected that the applied argument should be realized immediately after the verb. However, according to Rocha and Bewala (to appear) in Cinyanja the order of the objects seems to be free. This fact led the authors to propose that this language is partially asymmetric according to the typology posited by Bresnan and Moshi (1990).

(145) *Ndí ciyáni nkángo waphéla osáka?*

ndi	*ciyani*	*nkango*	*w-a-ph-el-a*	*osaka*
COP	what	3-lion	3SM-PAST-kill-APPL-FV	2-hunters

'Was it what that the lion killed for the hunters?'

A natural answer to this question is given below. In this pragmatically structured proposition, the focused argument *mbíidzi* 'zebra' is realized in sentence-final position, just like in presentational sentence (144). Strong evidence that this constituent is not *in situ*, but moved into a focus position, has to do with the fact that this argument is affected by the high floating tone of the copula. In this sense, regarding sentence (146), observe that the focused constituent is affected by a high tone spreading of the "silent" copula *ndí* from left to right, up to its penultimate vowel, changing its tone pattern from LL to HL. In addition, notice that a p-boundary is inserted to the left of the focused argument *mbídzi* 'zebra'.

(146) *Nkángo, waphéla osáka,* MBÍDZI.

nkango)	*w-a-ph-el-a*	*osaka)*	*mbidzi*
3-lion	3SM-PAST-kill-APPL-FV	2-hunters	9-zebra

'The lion killed for the hunters the zebra.'

Another piece of evidence in favor of my proposal that the focused DP above is not *in situ*, but actually it is moved into a focus position, comes from the fact that the copula *ndí* may precede it, as the example below illustrates.

(147) *Nkángo, waphéla osáka, ndí MBÍDZI.*

nkango)	*w-a-ph-el-a*	*osaka)*	*ndí*
3-lion	3SM-PAST-kill-APPL-FV	2-hunters	COP

mbidzi
9-zebra
'The lion killed for the hunters the zebra.'

Consider now another situation expressed by a presentational sentence, which corresponds to a double object construction.

(148) *Mwamúna apátsa Jowawu magalasi.*

mwamuna	*a-pats-a*	*Jowawu*	*magalasi*
1-man	1SM.PAST-give-FV	1-John	6-glasses

'The man gave John the glasses.'

The theme argument *magalasi* 'glasses' may be interrogated, as observed in the open question below. Again, notice that the high tone of the "silent" copula does not spread to the interrogative operator *ciyáni* 'what' in question (149), as opposed to what happens with the identificational focused constituent *magalasi*

'glasses' as shown in answer (150) below. The focused constituent *mágálási* 'glasses' has an underlying LLLL tone pattern as shown in (148), but in the pragmatically structured proposition (150) it changes to a HHHL pattern. Moreover, a p-boundary is inserted to the left of the focused argument. Compare the examples below.

(149) *Ciyáni mwamúna apátsa Jowáwu?*

ciyani	*mwamuna*	*a-pats-a*	*Jowawu*
what	1-man	1SM.PAST-give-FV	1-John

'What did the man give to John?'

(150) *Mwamúna, wapátsa Jowáwu, mágálási.*

mwamuna)	*w-a-pats-a*	*Jowawu)*	*magalasi*
1-man	1SM-PAST-give-FV	1-John	6-glasses

'The man gave John the glasses.'

In addition to the examples above, there is another construction that exhibits another applicative construction which semantically expresses reason/ purpose. This reason/purpose applied argument *phwando* 'party' in sentence (151) may be interrogated by means of the question operator *cifúkwa* as in the open question (152). Compare the data below.

(151) *Mulími adagwilizítsíla nkhúmba phwando.*

mulimi *a-dagwiliz-íts-íl-a* *nkhumba* *phwando*
1-farmer 1SM.PAST-stab-CAUS-APPL-FV 9-pig 5-party
'The farmer stabbed the pig for the party.'

(152) *Cifúkwa ciyáni mulímí adagwilizitsíra nkhúmba?*

cifukwa *ciyani* *mulimi* *a-dagwiliz-its-il-a*
reason what 1-farmer 1SM.PAST-stab-CAUS-APPL-FV

nkhumba
9-pig
'For what reason did the farmer stab the pig?'

Example (153) below is an appropriate answer to the question above. Notice the p-boundary to the left of the focused argument *phwándo* 'party'. Moreover, although the underlying tone of the reason/purpose constituent is LL, it surfaces as HL in (153).

(153) *Mulímí adagwilizitsíra nkhúmba, phwándo.*

Mulimi *a-dagwiliz-its-ir-a* *nkhumba* *phwando*
1-farmer 1SM.PAST-stab-CAUS-APPL-FV 9-pig 5-party
'The farmer stabbed the pig for the party.'

Another presentational sentence, consisting of a transitive verb modified by a locative argument is given below. Notice that the agent argument *mphunzitsi* 'professor' may be interrogated in the question below if one uses the interrogative operator *ndáni*. As mentioned previously, the hypothesis I am contending here is

that *ndáni* originates from *ndí+ani* by means of a phonological process that elides the final vowel /i/ of the copula *ndí* when it is merged with the interrogative particle *ani*, thereby giving rise to the form *ndáni*. The high tone of the copula is manifested in the first vowel of this interrogative word, as can be observed in (155).

(154) *Mphúnzítsi akumwáza mabúkhu munyúmba.*

mphunzitsi	*a-ku-mwaz-a*	*mabukhu*	*mu-nyumba*
1-professor	1SM-PRES-scatter-FV	6-books	LOC-house

'The professor is scattering the books around the house.'

(155) *Ndáni akumwáza mabúkhu munyúmba?*

ndani	*a-ku-mwaz-a*	*mabukhu*	*mu-nyumba*
Who	1SM-PRES-scatter-FV	6-books	LOC-house

'Who is scattering the books around the house?'

A possible answer to the question above makes use of the SOVX word order, in which the focused constituent *mphúnzítsi* 'professor' bears the HHL tone pattern, even though its underlying tone is LHL. This tone spreading is directly linked to the fact that this focused argument is moved into an IAV focus position. Furthermore, the focused DP is flanked by a p-boundary to its right, as indicated by the comma.

(156) *Mphúnzítsi, akumwáza mabúkhu, munyúmba.*

mphunzitsi)	*a-ku-mwaz-a*	*mabukhu)*	*mu-nyumba*
1-professor	1SM-PRES-scatter-FV	6-books	LOC-house

'The professor is scattering the books around the house.'

In the data above the focused argument is marked by a high floating tone which spreads to the penultimate syllable of the focused constituent, and by the insertion of a p-boundary. My proposal is that the tone of the focused constituent *mphunzitsi* 'professor' is changed from LHL to HHL due to the high floating tone originated from the "silent" copula. Strong evidence in favor for this analysis comes from the fact that the copula *ndí* may, in fact, be realized segmentally, as shown in the answer below.

(157) *Ndí mphúnzítsi, akumwáza mabúkhu, munyúmba.*

ndi	*mphunzitsi*	*a-ku-mwaz-a*	*mabukhu*	*mu-nyumba*
cop	1-professor	1SM-PRES-scatter-FV	6-books	LOC-house

'(It is) the professor is scattering the books around the house.'

A final piece of argument in favor of the analysis above comes from double object constructions. In this sense, consider the presentational sentence below.

(158) *Jowáwu wapátsa mwana kabudula*.

Jowawu	*w-a-pats-a*	*mwana*	*kabudula*
1-John	1SM-PAST-give-FV	1-child	5-shorts

'John gave the child the shorts.'

The theme argument *kabudula* 'shorts' may be interrogated by means of the question operator *ciyáni* 'what', as shown in example (159) below. The copula *ndí* is optional in this context, a situation that confirms my hypothesis that the copula *ndí* may be either segmentally realized or be elided. The option for the realization or elision of the copula is due to informational and stylistic considerations when the speaker is conceptualizing the utterance (at the Numeration stage, as I will propose in Chapter 5.

(159) *(Ndí) ciyáni Jowawu wapátsa mwana?*

ndi	*ciyani*	*Jowawu*	*w-a-pats-a*	*mwana*
COP	what	1-John	1SM-PAST-give-FV	1-child

'(Was it) what did John give the child?'

The following pragmatically structured proposition corresponds to an answer to (159) above. Notice that in this sentence the copula *ndí* occurs immediately before the focused argument *kábúdúla* 'shorts', and there also occurs the insertion of a p-boundary flanking the focused constituent to its left. In this regard, compare the sentences below. The focused argument may be realized in sentence-final position (160) or in sentence-medial position (161).

However, when the copula is segmentally realized the structure in which the focused constituent occurs in sentence-medial position is considered odd (161).

(160) *Júwáwú, wápátsá mwána, ndí KÁBÚDÚLA.*

Juwawu	*w-a-pats-a*	*mwana)*	*ndi*	*kabudula*
1-John	1SM-PAST-give-FV	1-Mary	COP	5-shorts

'John gave the child it was the SHORTS.'

(161) *?? Jówáwú, wápátsá, ndí KÁBÚDÚLA, mwána*.

Jowawu)	*w-a-pats-a)*	*ndi*	*kabudula)*	*mwana*
1-John	1SM-PAST-give-FV	COP	5-shorts	1-child

'John gave it was the SHORTS the child.'

In addition to the possibilities above, it is observed that the copula *ndí* may even be erased as shown in sentence (162) below. In such contexts, evidence that the copula is "silent" comes from the tone borne by the focused argument *kábúdúla* 'shorts'; the underlying tone of this DP is LLLL, but it surfaces as HHHL when focused.

(162) *Júwáwú, wápátsá mwána, KÁBÚDÚLA.*

Juwawu	*w-a-pats-a*	*mwana)*	*kabudula*
1-John	1SM-PAST-give-FV	1-child	5-shorts

'John gave Mary (it was) the SHORTS.'

In sum, the example above serves as an additional piece of evidence in favor of my analysis, according to which the focused DP is not really *in situ*, but it is moved into a low focus position immediately after the copular verb in the matrix clause of a complex cleft-like sentence.

One of the goals of Chapter 5 is to provide the reader with a generative approach in order to explain why the focused DP in (162) can appear in sentence-final position. As will become clearer in the next chapter, my hypothesis is that sentences that exhibit focused DP in such positions correspond to pseudocleft constructions in which the whole predicate (=CP) is moved into spec-InflP of the matrix clause. The side effect of this analysis is that, when the copula *ndí* appears segmentally before the focused constituent, the interpreptation of the sentence is akin to a pseudocleft construction, in the sense that it is derived by means of the movement of the embedded clause (CP) to spec-InflP of the higher clause.

In sum, the pieces of evidence discussed in this section do confirm my hypothesis according to which the high floating tone spreading rule occurs only in contexts of identificational focus constructions, but not in open interrogative sentences. The analysis also shows that a 'silent' copular verb may be inferred in contexts in which the focus argument is affected by a high floating tone, regardless of whether that these constituents appear in sentence-initial, sentence-medial, or in sentence-final position.

A derivation proposal will be presented in the next chapter, in order to explain the reason why focused constituents appear in different syntactic positions, although all of them are uniformly moved into the specifier of a low Focus projection.

CHAPTER SUMMARY

In this chapter I presented the grammatical strategies of expressing constituent focus in Cinyanja.

Two main strategies to mark focused constituents in Cinyanja were identified: prosody marking, and syntactic marking. Regarding prosody marking, the focused argument bears a high floating tone and is flanked by a p-boundary. As for syntactic marking, the pragmatically structured proposition may be realized as a canonical cleft, a pseudocleft, or the default SVOX structure. It is interesting to note that both the copula *ndí* and the complementizer *-mene* are optional depending on the construction.

Thus, the aim of the next chapter is to develop the theoretical proposal, based on the data discussed so far. My hypothesis is that clefts and pseudoclefts are widely used to mark focused constituents, even when the SVOX structure

emerges, due to the presence of a “silent” copula which is diagnosed by means of a high floating tone that always affects identificational focused constituents. The difference between the two constructions has to do with the fact that there is an interaction of different movements. Thus, the focused argument is systematically moved into spec-FocusP. However, in the pseudocleft, but not in the canonical cleft, the subordinate CP is moved into spec-IP of the matrix clause. I will discuss this hypothesis in detail in the next chapter.

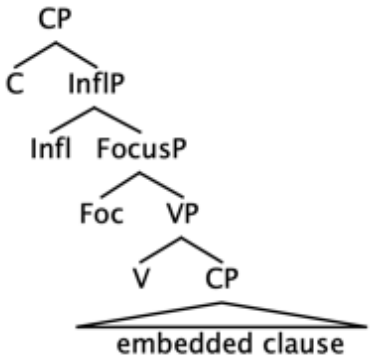
CHAPTER 5

THEORETICAL PROPOSAL: DERIVING THE REALIZATION OF FOCUSED ARGUMENTS IN CINYANJA

The aim of this chapter is to provide a minimalist account of the distribution of focused constituents in Cinyanja. As seen in chapter 4, focused arguments occupy the following syntactic positions: in sentence-initial, sentence-medial, and sentence-final position, or immediately after a segmentally realized copular verb, that is, the IAV position. The goal of this chapter is to propose that this situation is only apparent. More to the point, the hypothesis I am advancing in this dissertation is that, syntactically, the Focus head always projects in a low IAV position, and the focused constituent is always moved into spec-FocusP. In sum, my proposal is that FocusP is systematically projected below Inflection Phrase[15] (InflP) and above VP, as depicted in the syntactic structure below.

[15] In this dissertation I will assume a simple InflP, even though the verbal prefixes in Cinyanja consist in many heads such as Aspect, Tense, Mood, Negation, and φ-AGREE, as well as pronoun incorporation. This discussion over the split-InflP is out of the scope of this dissertation. I will explore this question in further work.

(1)



The proposal above is based on five pieces of evidence:

(i) the copula *ndí* bears a high tone which spreads to the focused argument;

(ii) the copula *ndí* may optionally occur immediately before the focused constituent;

(iii) the focused DP bears a high floating tone even when the copula is not segmentally realized;

(iv) the complementizer *méne* 'that' is also optional in clefts and pseudoclefts;

(v) when it is overt, the complementizer enters an agreement relation with the focused constituent.

Regarding the surface structure of a pragmatically structured proposition in Cinyanja, the default word order is SVOX[16], as one can notice in the example below, which consists of a presentational sentence. Notice that the agent argument *mphunzítsi* 'professor' bears an underlying LHL tone pattern.

(163) *Mphunzítsi akumwáza mabúkhu munyúmba.*

mphunzitsi	*a-ku-mwaz-a*	*mabukhu*	*mu-nyumba*
1-professor	1SM-PRES-scatter-FV	6-books	LOC-house

'The professor is scattering the books around the house.'

Consider now the question-and-answer pair below, in which the agent argument *mphunzitsi* 'professor' is interrogated by means of the question operator *ndáni* 'is who', as follows.

(164a) *Ndáni akumwáza mabúkhu munyúmba?*

ndani	*a-ku-mwaz-a*	*ma-bukhu*	*mu-nyumba*
is.who	1SM-PRES-scatter-FV	6-books	LOC-house

'Who is scattering the books around the house?'

(164b) *MPHÚNZÍTSI, akumwáza munyúnba mabúkhu.*

mphunzitsi)	*a-ku-mwaz-a*	*mu-nyumba*	*mabukhu*
1-professor	1SM-PRES-scatter-FV	LOC-house	6-books

'The professor is scattering the books around the house.'

[16] This order occurs with transitive verbs. With ditransitive verbs the order is SV(IO)(DO) or SV(DO)(IO), as shown in 2.3.

Notice that in answer (164b) the agent argument is realized in subject position. It is important to point out that this SVOX order is, however, only apparently default. In this regard, one of the theses I propose in this dissertation is that, in such contexts, there is a "silent" copula immediately before the focused constituent *mphunzitsi* 'professor' above. A piece of evidence in favor of this proposal comes from prosody. Pursuing these lines of reasoning, I argue that the focused DP, although its underlying tone is LHL as seen in (163) above, surfaces with a HHL tone pattern, owing to a high floating tone spreading that originates in the "silent" copular verb *ndí*, which bears a high tone. Moreover, it is observed that there is a phonological boundary (p-boundary) to the right of the focused argument. Additionally, the copula *ndí* and the complementizer *méne* may optionally occur as indicated by the parentheses in (165) below, giving rise to an overt cleft construction. Notice that the focused constituent and the complementizer establish an agreement relation, which is evidenced by the occurrence of the prefix {*-a*} of class 1 on the complementizer stem.

(165) (*Ndí) MPHÚNZÍTSI, (améne) akumwáza munyúnba mabúkhu.*

ndi	*mphunzitsi)*	*a-mene*	*a-ku-mwaz-a*	*mu-nyumba*
COP	1-professor	1SM-COMP	1SM-PRES-scatter-FV	LOC-house

mabukhu

6-books
'(It is) the PROFESSOR (that) is scattering the books around the house.'

It is also important to point out that the agreement that occurs between the focused constituent and the complementizer *méne* serves as evidence that the focused constituent has crossed through spec-CP before landing in the specifier of the focus projection, i.e. spec-FocusP. In other contexts, the focused constituent occurs in sentence-final position, as shown by the example below. Notice that the complementizer may optionally occur in sentence-initial position, as indicated by the parentheses.

(166a) Who is scattering the books around the house?

(166b) (*Améne) akumwáza mabúkhu munyúnba, ndí mphúnzítsi*.

a-mene	*a-ku-mwaz-a*	*mabukhu*	*mu-nyumba*	*ndi*
1-COMP	1SM-PRES-scatter-FV	6-books	LOC-house	COP

mphunzitsi
1-professor
'Who is scattering the books around the house is the professor.'

Taking into account the empirical data shown thus far, the main theoretical problem I will explore in this chapter may be summarized as follows:

i. How to account for the apparently different syntactic positions a focused constituent may occupy in a given pragmatically structured proposition in Cinyanja?

ii. Are there more than one focus positions in Cinyanja clauses?

Based on the data presented in Chapter 4 and Duarte and Langa's (2025) analysis, I propose that in fact there is only one syntactic position for focused arguments in this language, namely the IAV position, which in turn corresponds to the specifier of a low focus position, as depicted in the syntactic structure (1) above.

Following Aboh (2010), I will assume that Focus features are present in the narrow lexicon and enter the derivation via the Numeration step. Features such as [*i*FOC], [*U*FOC:__], and EPP are assigned at the beginning of the derivation, similarly to other formal features such as Case. I will also assume, following Rizzi (1997), that Focus features are Criteria – features which bear semantic import and drive Ā-movement. Hence, the Focus head bears the interpretable focus feature [*i*FOC] which values the uninterpretable focus feature [*u*FOC:__] of the focused constituent. Additionally, the Focus head should bear an EPP feature in order to attract the focused DP into its specifier position. Moreover, the Focus head functions as a probe, which enters an AGREE relation with the goal – the focused argument. Therefore, Cinyanja conforms to Carstens'

(2005) proposal that AGREE in Bantu requires a spec-head configuration (cf. section 3.3).

The two tree diagrams below illustrate the syntactic structure of a pragmatically structured proposition in Cinyanja. Notice that the embedded CP may occupy two positions: *in situ* (2), or in spec-InflP (3). This happens due to the nature of the Inflection head. In this sense, I propose that, in the lexicon, there are three different heads which bear TAM, agreement, and negation features: the first head bears a strong EPP feature and φ-features, as well as an interpretable Case feature [NOM], which allows this Infl head to attract arguments; the second head bears a strong EPP feature but no Case nor φ-features[17] and thus it attracts the embedded clause of the sentence it is part of; and the third head bears no Case feature, nor φ-features, neither EPP features, and therefore does not project a specifier position. The choice between the three Infl heads occurs at the Numeration step of the derivation, when the speaker selects the roots and bundles of features (heads) according to stylistic or informational considerations, i.e. the conceptualization of the content of the sentence at the time of utterance.

[17] Duarte (personal communication) points out that, in Bantu languages such as Changana, the copular verb bears φ-features, given that this verb establishes an agreement relation with the different grammatical persons that is realized morphologically. Whether the same is true for Cinyanja is out of the scope of this dissertation and will be addressed in further research.

(2) (3)

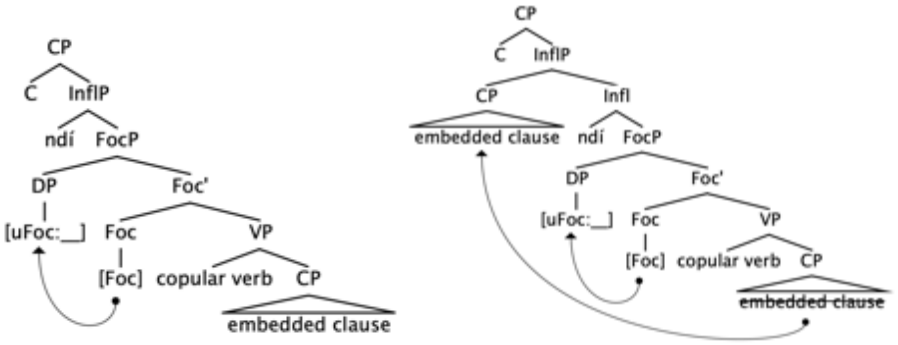


Figure 7: Syntactic structures of a pragmatically structured proposition in Cinyanja

Regarding the typology of focus marking proposed by Gibson et al (2017) (cf. section 3.5.3), I propose that Cinyanja expresses focus syntactically in the IAV position. There is no morphological marking on the verb nor on the argument, but the constituent is marked via prosody effects (due to the presence of the copula). Consider the table below.

Table 8: Typology of focus marking in Cinyanja

typology	initial	IBV	IAV	final
on term	X	X	√	X
on verb	X	X	X	X

In sum, the typology of focus marking in Cinyanja can be summarized as follows:

i. Prosodic marking:

   a. The copular high tone that spreads up to the penultimate syllable of the focused DP

   b. P-boundary insertion to the right or left of the focused argument

ii. Syntactic marking:

   a. Cleft and pseudocleft constructions

This chapter is organized as follows: section 5.1 will discuss the derivation of open questions in Cinyanja. Section 5.2 will discuss the derivation of the canonical cleft employed by Cinyanja speakers to mark constituent focus syntactically. Section 5.3 will present the derivation of the pseudocleft construction.

5.1: DERIVATION OF OPEN QUESTIONS

As demonstrated in Chapter 4, question operators such as *ciyáni* 'what' and *ndáni* 'who' may occur either in sentence-initial, sentence-medial, or sentence-final position. Notice that the copula *ndí* may optionally occur before

the interrogative word. In order to reinforce that this proposal is really on the right track, compare the examples below.

(167a) *(Ndí) ciyáni bilidála wamangíla anéba?*

ndi	*ciyani*	*bilidala*	*w-a-mang-il-a*	*aneba*
COP	what	1-bricklayer	1SM-PAST-build-APPL-FV	1-neighbor

'What did the bricklayer build for the neighbor?'
'It was what did the bricklayer build for the neighbor?' (lit.)

(167b) *Bilidála wamangíla anéba (ndí) ciyáni?*

bilidala	*w-a-mang-il-a*	*aneba*	*ndí*
1-bricklayer	1SM-PAST-build-APPL-FV	1-neighbor	COP

ciyáni
what
'The bricklayer built for the neighbor what?'
'The bricklayer built for the neighbor was what?' (lit.)

Notice that the first vowel of the interrogative operator *ciyáni* 'what' does not bear a high tone, a fact that is not expected given the interaction with the copula. As already discussed in Chapter 4, I propose that the expansion of the high floating tone of the copula only applies to identificational constituent focus and not to information focus, so that question operators are not affected by the high tone, but only DPs in identificational constituent focus context.

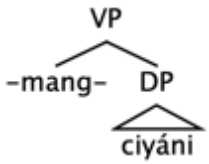
Pursuing the theoretical assumptions raised in Chapter 3, the derivation of question (167a) above can proceed as follows. The first step of any given

derivation is the Numeration. In this step the roots and bundles of features are selected from the narrow lexicon and are then available to be merged with each other in order to establish syntactic relations.

N = {$\sqrt{\text{bilidala}}$, $\sqrt{\text{mang}}$, $\sqrt{\text{aneba}}$, ciyáni, $\text{Case}_{[\text{NOM}]}$, $\text{Case}_{[\text{ACC}]}$, $\text{Case}_{[\text{DAT}]}$, [uCASE:__], [FOC], [uFOC:__], [PAST], [PERFECTIVE], AgrS}

The next step of the derivation consists in the categorization of the roots, in what Arad (2003) proposes to be the first (micro) phase. Then, the interrogative word *ciyáni* 'what' corresponds to the internal argument of the verb and is, thus, merged with the verbal root in order to derive the VP projection.

(4)

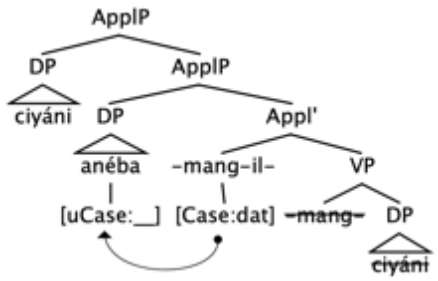


In the next step of the derivation, the VP is taken as the complement of the applicative head, which licenses the benefactive argument *aneba* 'neighbor';

the uninterpretable Case[18] feature of this applied DP is valued by the [CASE:DAT] feature of the applicative head in a spec-head relation. The verb is then merged with the applicative morpheme. Since the applicative head is phasic, according to McGinnis (2008), it bears an EPP feature and projects an "escape hatch", that is, an extra specifier position. This position is then occupied by the interrogative word *ciyáni* that is raised to this position out of its thematic position. Insofar as this question operator bears both a WH-feature, an uninterpretable [UFOC:__] feature, and an uninterpretable [UCASE:__] feature, these features need to be valued and deleted in the next stages of the derivation. At this point, the domain of the applicative projection comprised of the VP is then sent to spell-out. Notice that the edge of the phase head remains visible for further computation. Given the theme argument, which is realized by the question operator *ciyáni*, occupies this edge position, it is still active and is then accessible to higher probes, so that it can get its uninterpretable features to be valued by the Voice head, the Focus head and the complementizer head.

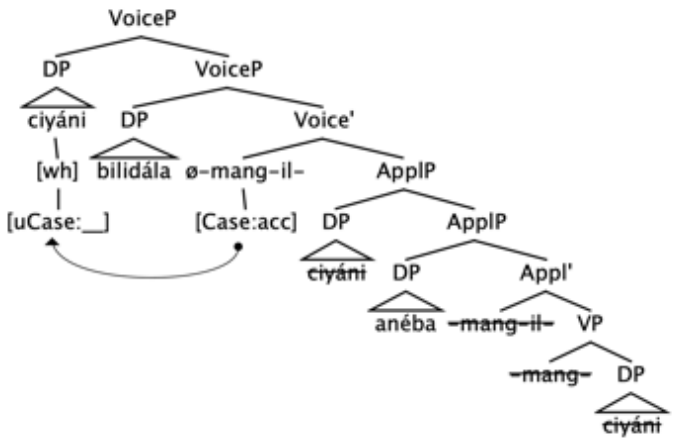
[18] It is important to note that Case in Cinyanja is not marked morphologically. However, abstract Case considerations such as the Case filter still apply in this language. I am assuming Chomsky's (2012) proposal, according to which Case pervades in every language, the difference being the way it is morphologically encoded in a particular language. In some languages Case is realized morphologically, while in others there are no particular morphemes to encode it. Cinyanja is a language of the latter type.

(5)



In the next step, ApplP is then taken as the complement of an active Voice head, which introduces the external agent argument *bilidála* 'bricklayer'. Since this is also a phase head, it also bears an EPP feature and projects an escape hatch. As a result, in order to value the EPP feature of the Voice head, the theme argument realized as the question operator *ciyáni* 'what' is moved into this extra specifier position. By doing so it values the EPP feature of Voice and gets its uninterpretable Case feature valued as [CASE:ACC]. Notice that the question operator *ciyáni*, as it is located in an outer specifier of VoiceP, can be accessible to higher probes, namely C and Focus. At this point of the derivation, the domain of the Voice head is sent to spell-out, except for the interrogative pronoun and the agent argument, since they are both on the edge of the Voice projection, as depicted by the syntactic derivation below.

(6)

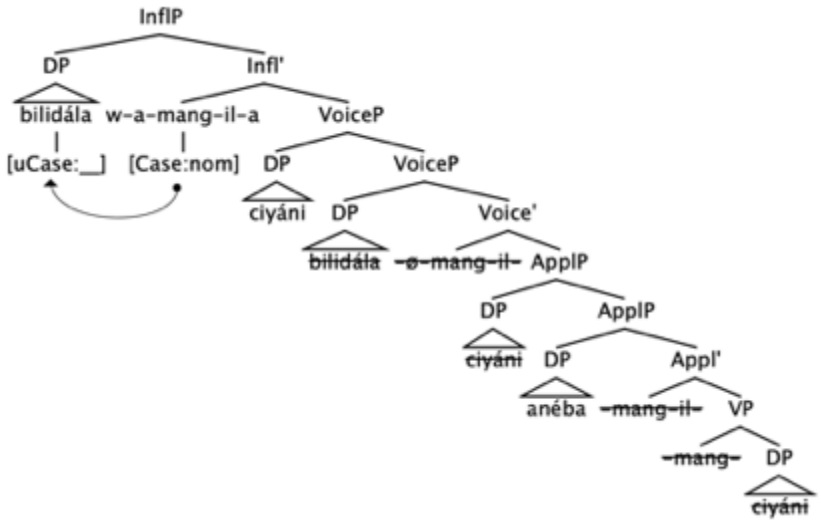


VoiceP is then taken as the complement of an Inflection head[19], which bears Case features, φ-features, and EPP features. In the structure of the verb in Cinyanja there are prefixes that encode negation, tense, aspect, mood, subject agreement, and pronominal object incorporation. I will conflate all of these heads into a syncretic head Inflection (InflP), since the realization of the different verbal prefixes does not seem to interact with focus dynamics. Hence, the Infl head attracts the highest argument that still bears an uninterpretable [uCASE:__] feature – in this context the agent argument *bilidála* 'bricklayer'. The uninterpretable Case feature of the interrogative word *ciyáni* 'what' is valued by the Voice head, thus it becomes invisible to the Case-assigning Probe, the Infl

[19] As mentioned before, the nature of the Infl head will not be explored in this dissertation since it is out of the scope of the research.

head. This explains the reason why minimal link condition (MLC) is not violated when the agent crosses the theme argument that is sitting in the outer specifier of VoiceP at the moment it is moved into spec-InflP.

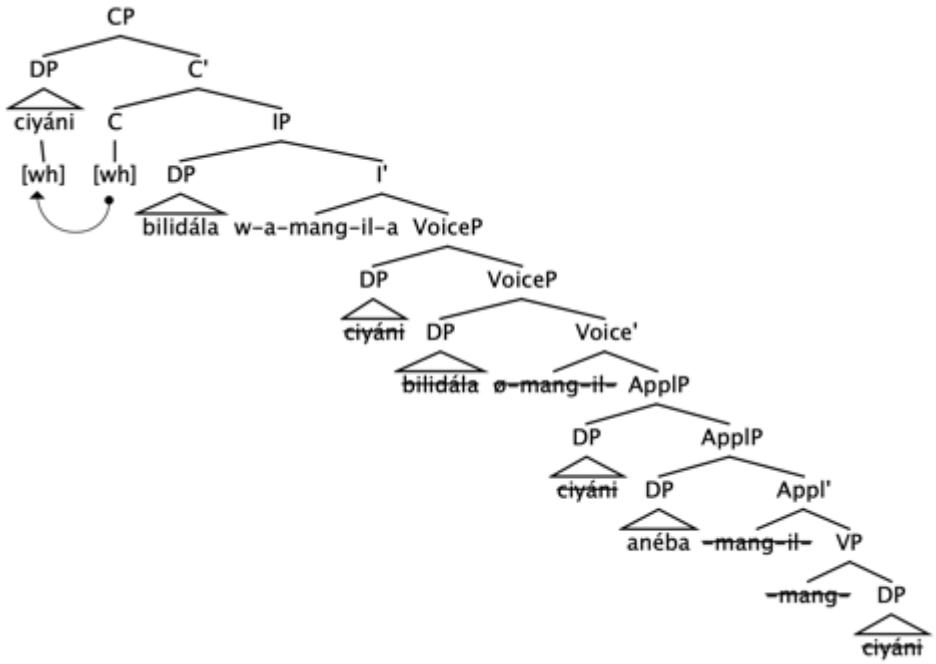
(7)



Notice that the focused constituent *ciyáni* 'what' is left behind in spec-VoiceP, at the moment when the agent is raised to spec-InflP. Moreover, it is important to point out that this question operator bears both a WH-feature and an uninterpretable feature [uFOC:__] that need to be valued respectively by the complementizer probe and the Focus probe. A solution to this problem can be found if I propose that, after the complementizer head C c-selects InflP, the question operator must be moved into the specifier position of CP via Ā-

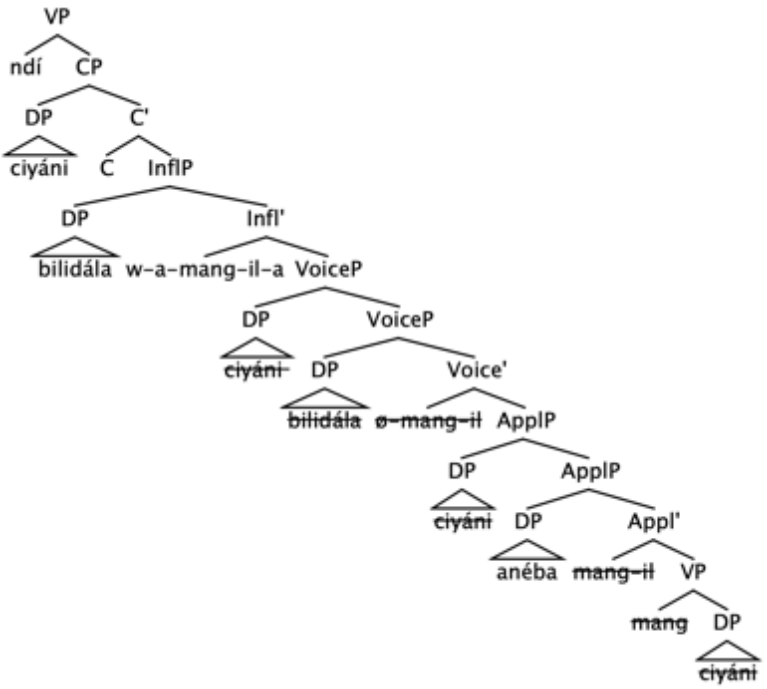
movement. This movement is driven by the WH-feature/criteria and EPP features borne by the complementizer head. The WH-feature/criteria of the interrogative word is valued by the WH-feature of C, via spec-head relation, resulting in the interrogative character of the sentence. On the other hand, this question operator satisfies the EPP feature of C. In this sense, Cinyanja conforms to Carstens' (2005) proposal that, in Bantu languages, the complementizer head C bears EPP features.

(8)



In turn, CP is taken as the complement of a copular verb, which may optionally occur segmentally as *ndí*.

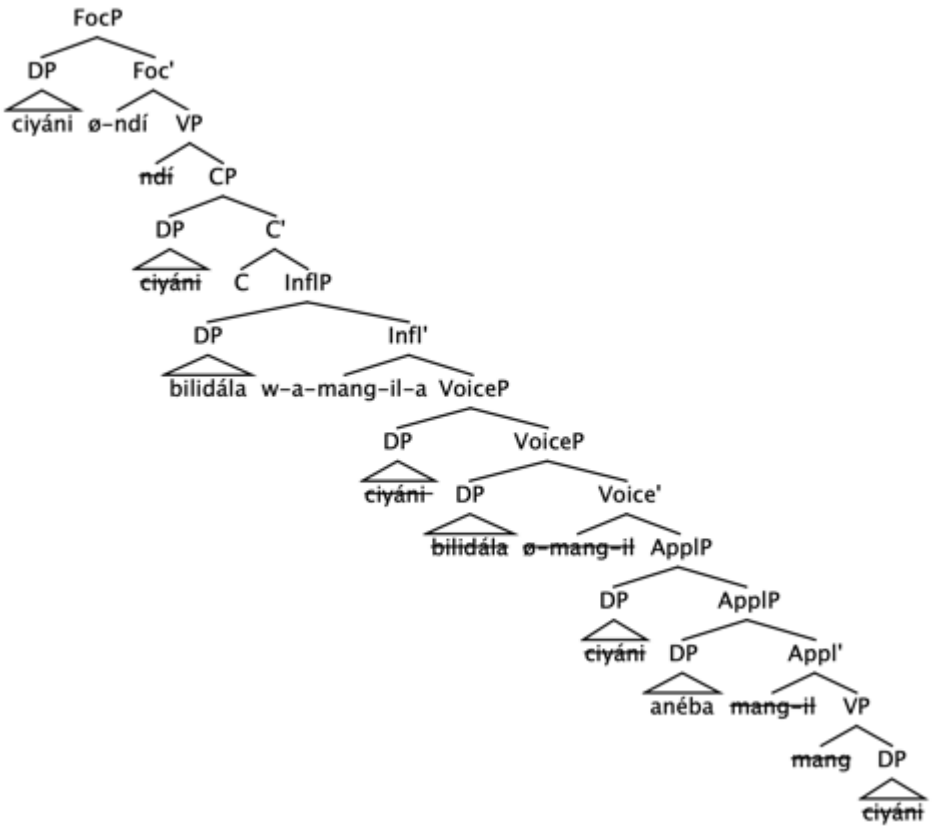
(9)



In the next step, the copular verb is then taken as the complement of the Focus head. As this head bears an uninterpretable EPP feature and an interpretable [FOC] feature, we conclude that the Ā-movement of the interrogative pronoun into spec-FocusP values the EPP feature of Focus and, in this syntactic position, the DP gets its uninterpretable focus feature [UFOC:__]

valued by the Focus head. In addition, the copular verb *ndí* must move to the focus head, as is depicted in the syntactic structure below.

(10)



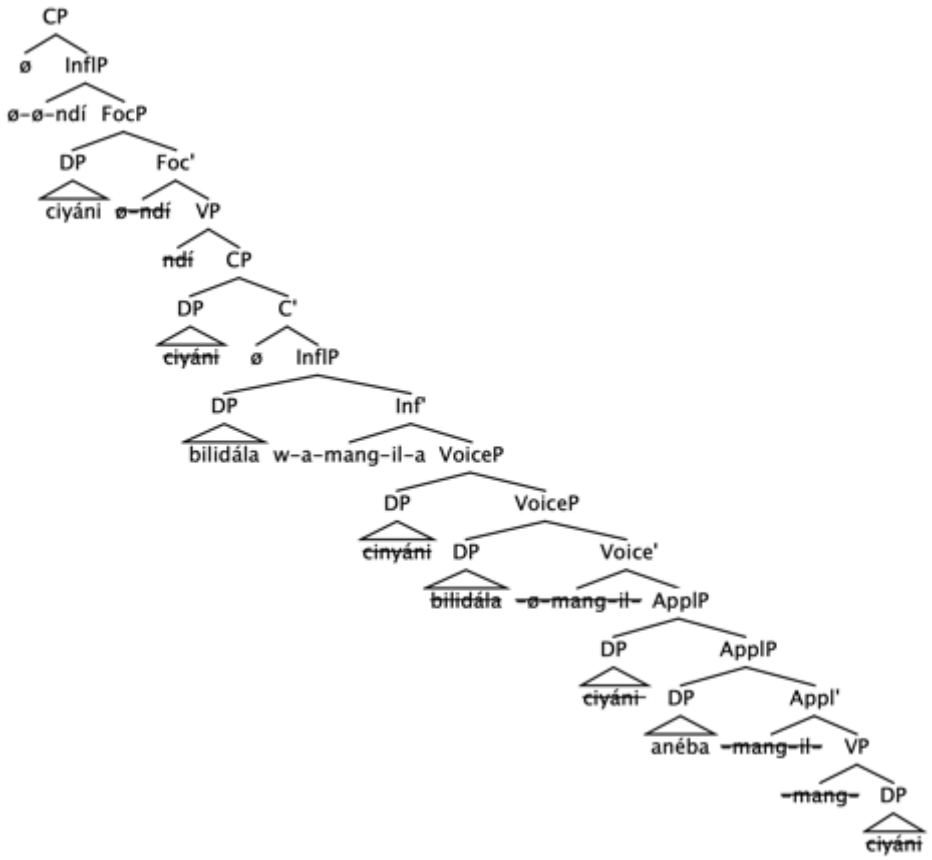
In the next step we assume that FocusP is c-selected by the Inflection head. Then, the copular verb is merged into Infl, but the focused constituent does not move to spec-InflP, for two reasons: (i) this Infl head is defective and hence

does not bear an EPP feature, and (ii) the question operator is not visible to further computation, given the fact that all of its uninterpretable features have already been valued and deleted. The result of the syntactic operations delineated above is the canonical cleft construction. Additionally, this Infl head is defective in the sense that its specifier position is not projected. As previously argued, the choice for the appropriate Infl head is due to stylistic or informational properties, that is, how the speaker conceptualizes the content of the utterance. In the context of the canonical cleft, the Infl head of the embedded clause bears Case, EPP, and φ-features, and hence it attracts the highest argument in the structure which bears uninterpretable Case features. On the other hand, the Infl head of the matrix clause does not bear EPP, Case, nor φ-features, and thus it does not project a specifier position and does not attract any XP.

(11)

Then, InflP is merged with C. At this point the computation is complete, and all the structure built so far is sent to spell-out. The result of this process is a canonical cleft construction, in which the focused argument (the interrogative pronoun) occupies the Immediately After the (copular) Verb (IAV) syntactic position, as follows.

(12)



As mentioned in the previous chapter, the interrogative pronoun may also occur in sentence-final position. In such contexts, I propose that the embedded CP is Ā-moved into spec-InflP, thereby giving rise to the pseudocleft construction (reverse cleft in Kato's (2017) sense). Hence, in this case the Infl head which occurs in the main clause bears an EPP-feature, but not φ-features nor Case features. More precisely, the thesis that I propose is that CP moves to spec-InflP,

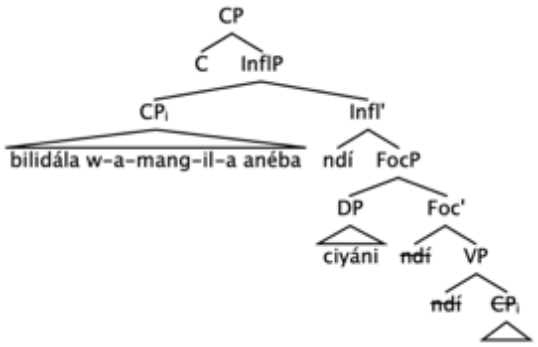
and, as a consequence, it leaves a silent copy in its base position. In this sense, the sentence (167b), repeated below as (168), in which the question operator is realized in sentence-final position, has the derivation depicted in the syntactic structure (13) below.

(168) *Bilidála wamangíla anéba (ndí) ciyáni?*

bilidala	*w-a-mang-il-a*	*aneba*	*ndí*
1-bricklayer	1SM-PAST-build-APPL-FV	1-neighbor	COP

ciyáni
what
'The bricklayer built for the neighbor what?'
'The bricklayer built for the neighbor was what?' (lit.)

(13)



The same syntactic process that derives open questions in Cinyanja may also be applied to the correspondent answer since the form of the question determines the form of the answer.

Based on these assumptions, in the next section I will deal with the context of focused agent arguments that occur in sentence-initial position. As will become clearer, I will assume that the same derivation process as the one outlined above can be adopted to derive identificational focus constructions. I will claim that, in these contexts, the copular verb may be realized either segmentally or via prosody effects, that is, by a high floating tone, thereby rendering a construction which resembles the canonical cleft.

5.2: DERIVATION OF THE CANONICAL CLEFT CONSTRUCTIONS

In Chapter 4, I showed that Cinyanja makes use of canonical cleft constructions to encode constituent focus. As demonstrated previously, the canonical cleft takes the form of COPULA + FOCUSED ARGUMENT + COMPLEMENTIZER + EMBEDDED CLAUSE. Consider the example below. In this sentence the copula *ndí* and the complementizer *méne* are both optional, as indicated by the parentheses. Hence my proposal is that the pragmatically structured proposition (169b) below corresponds to a canonical cleft

construction, even in contexts in which the copula is not realized segmentally and the SVOX word order occurs.

(169a) *Ndáni aphikíla nyáma phwando?*

ndani	*a-phik-il-a*	*nyama*	*phwando*
is.who	1SM.PAST-cook-APPL-FV	7-meat	5-party

'Who cooked the meat for the party?'

(169b) (*Ndí) ÁGÓGO, (améne) aphikíla nyáma, phwando.*

ndi	*agogo)*	*a-mene*	*a-phik-il-a*
COP	1-old.woman	1-COMP	1SM.PAST-cook-APPL-FV

nyama)	*phwando*
7-meat	5-party

'(It was) the OLD WOMAN (that) cooked the meat for the party.'

Regarding the informational properties of the pragmatically structured proposition in (169b) above, they can be described as follows.

Q: Who cooked the meat for the party?

A: *Ndí* [AGOGO]$_{FOC}$, *améne aphikíla nyáma, phwando*.

It was the OLD WOMAN that cooked the meat for the party.

- Presupposition: 'x cooked the meat for the party'

- Assertion: ‘x = old woman’
- Focus: old woman
- Focus domain: DP

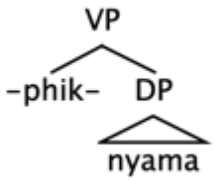
Concerning the syntactic structure of answer (169b), it is compounded of a complex sentence comprising two clauses: a main clause headed by the copular verb and an embedded clause from which the focused argument is raised. The derivation of this sentence occurs in cyclical stages in which the syntactic objects are merged into new syntactic objects. The derivation below illustrates my proposal.

The first step of the derivation is the Numeration, as follows.

N = {$\sqrt{\text{agogo}}$, $\sqrt{\text{nyama}}$, $\sqrt{\text{phwando}}$, $\sqrt{\text{phik}}$, $\text{Case}_{[\text{NOM}]}$, $\text{Case}_{[\text{ACC}]}$, $\text{Case}_{[\text{DAT}]}$, [uCASE:__], [FOC], [uFOC:__], $\text{V}_{[\text{COP}]}$, $\text{Infl}_{[\text{PAST, PERF, 1SM}]}$

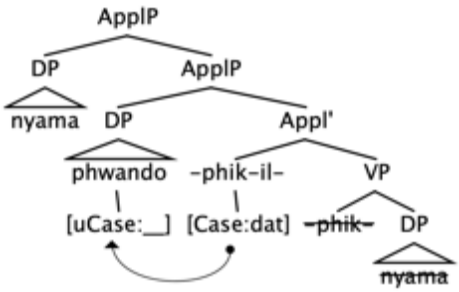
Then there occurs the categorization of the roots. The next step merges the verb root $\sqrt{phik}$ ‘cook’ with the patient argument *nyama* ‘meat’, deriving the VP projection.

(14)



The VP is later c-selected by the Applicative head, which introduces the reason/purpose argument *phwando* 'party'. This head assigns dative Case to the applied argument in spec-head configuration. Additionally, Appl is phasic, so it projects an escape hatch at its edge that is occupied by the patient DP *nyáma* 'meat'. Its domain, i.e. VP, is then sent to spell-out, except for the edge.

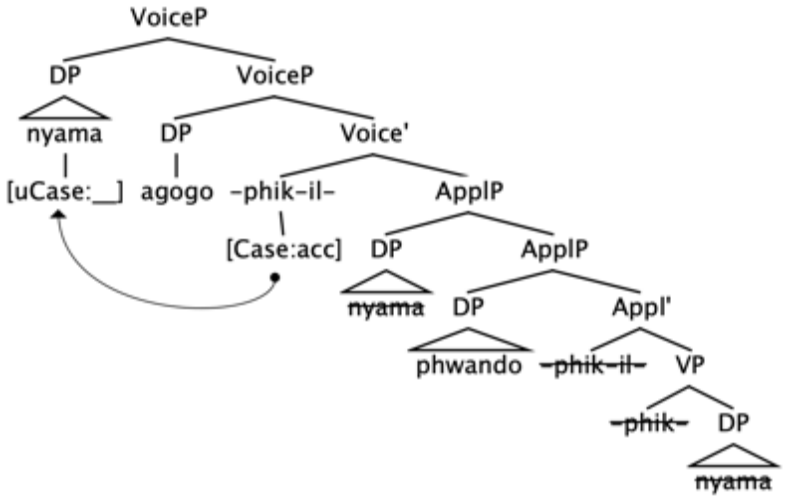
(15)



The applicative head is in turn merged as the complement of an active Voice projection that introduces the agent argument. This phase head projects an

escape hatch into which the patient argument *nyáma* ‘meat’ is moved. In this specifier position the patient DP is assigned accusative Case by the Voice head via spec-head agreement. At this stage of the derivation, the domain of VoiceP that corresponds to ApplP, is sent to spell-out.

(16)



In the next step of the derivation the Inflection head takes VoiceP as its complement. The agent argument *ágógo* ‘old woman’ is moved into spec-InflP in order to value EPP and φ-features of Infl and get its uninterpretable Case feature valued by Infl. Notice that, at this point of the computation, the agent DP *ágógo* ‘old woman’ has to skip the patient DP *nyáma* ‘meat’ that is sitting at the edge of VoiceP. Hence, in this context, the agent argument seems to violate the

Minimal Link Condition (MLC)[20]. However, due to the fact that the agent DP still bears an uninterpretable [uFOC:__] feature/criteria and an uninterpretable Case feature which must be valued and deleted, it is thus available for the Move operation. All the formal features such as Case features of the patient constituent are already valued and, therefore, it is invisible to the next stages of the derivation. Thus, the MLC is not violated because the closest argument that bears an unvalued uninterpretable Case feature and therefore may be probed by the active Infl head is the agent.

The structure below illustrates the step, in which there occurs the merge of the agent argument into spec-InflP and the feature valuation relations that take place in this syntactic position, namely EPP, φ-features, and Case features.

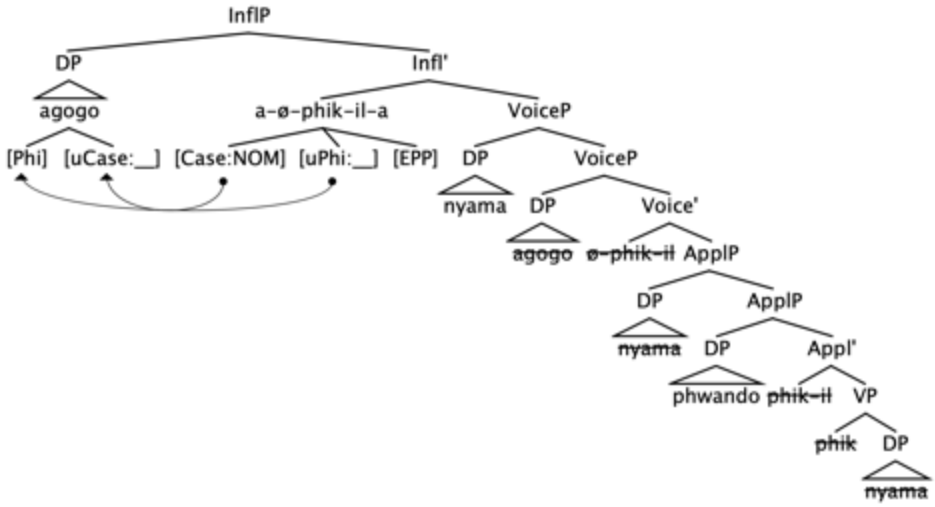
---

[20] The Minimal Link Condition is a locality constraint which is formulated as a condition on category movement. Ishii (2000:309) provides the following definition:

(i) α can raise to target K only if there is no legitimate operation Move-β targeting K, where β is closer to K than α (where α and β are formal features).

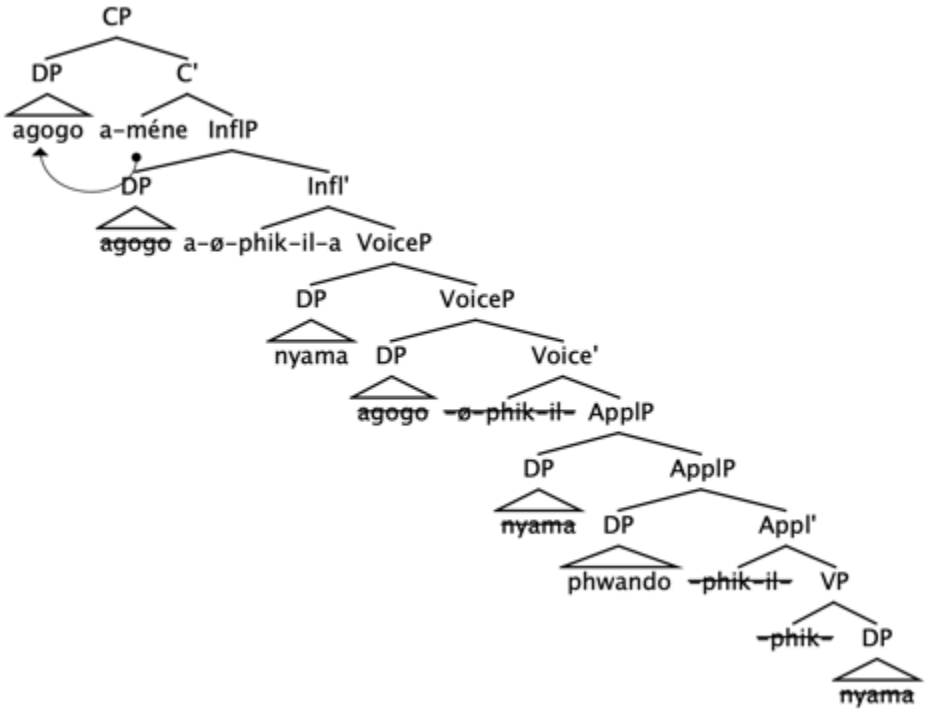
Thus, MLC predicts that the closest category which can enter a featural relation with a probe will be chosen for the complex operation Move (external Merge). Recall my assumption in Chapter 3 that the lexicon contains bundles of features which are categorized in the course of the computation.

(17)



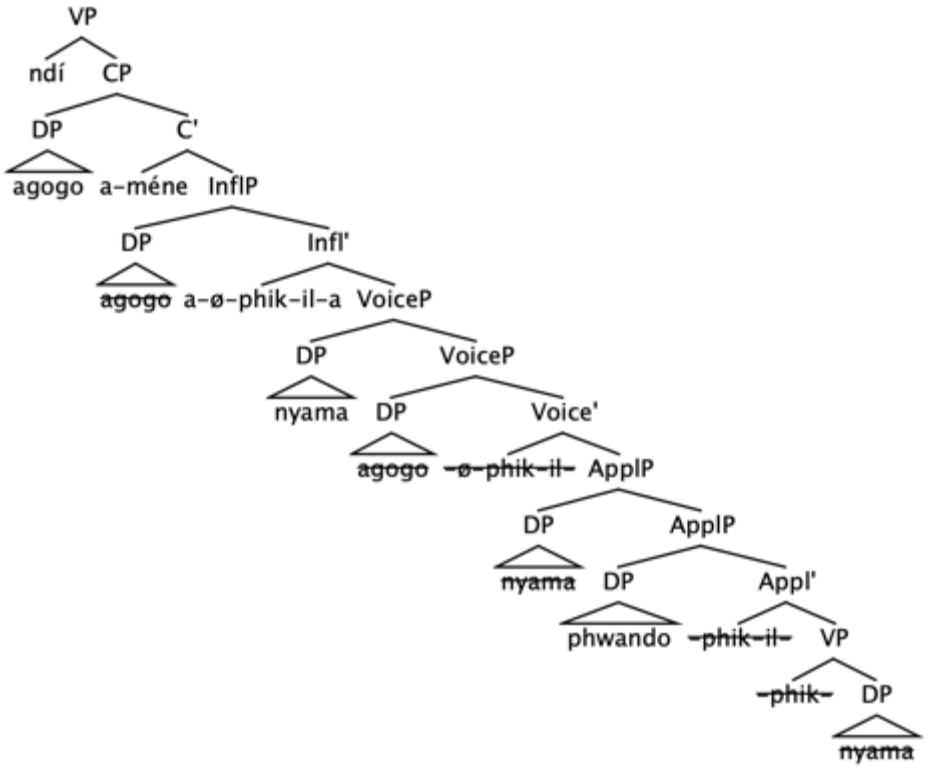
Finally, IP is taken as the complement of CP. In this step the focused argument is attracted by the EPP feature of C. Recall that this Ā-movement is available because the focused DP bears an uninterpretable feature that needs to be valued. Then, the focused constituent and the complementizer head enter a φ-AGREE relation in a spec-head configuration, as the reader can observe from the fact that, as a result of the agreement operation between C and the focused DP, the prefix of class 1 {-a} must appear merged with the complementizer stem. The structure below illustrates this relation.

(18)



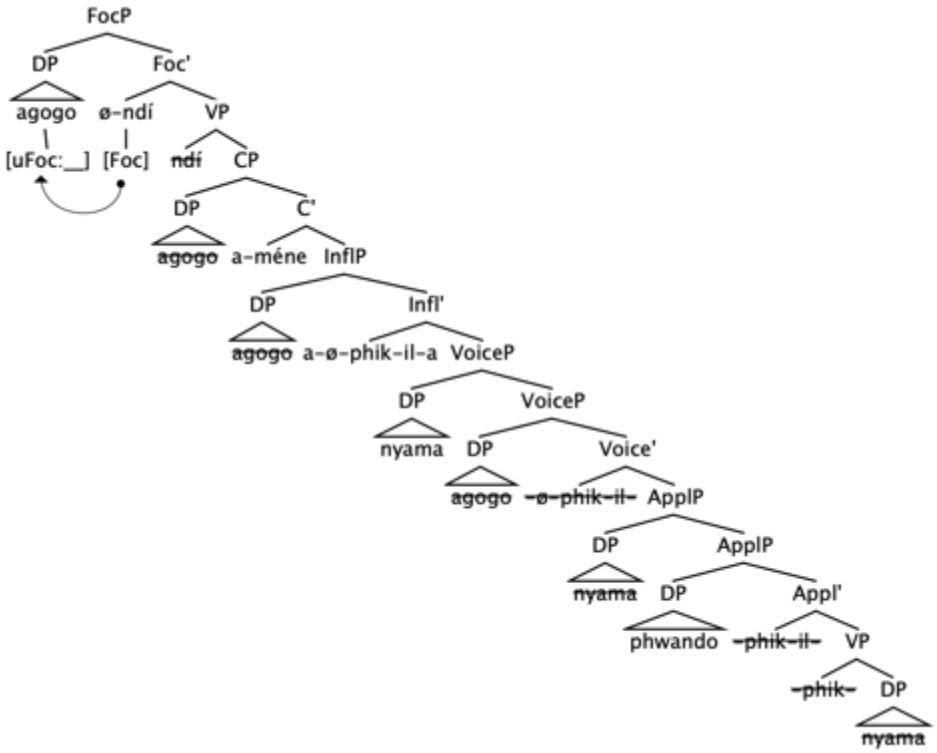
Because C is a phase head, its domain (InflP) is sent to spell-out, leaving the edge (the focused constituent) still accessible to the next steps of the derivation. The next step of the computation, when the CP phase is closed, this syntactic projection is taken as the complement of a copular verb, which is realized as *ndí* in Cinyanja. The structure below illustrates this stage of the computation.

(19)



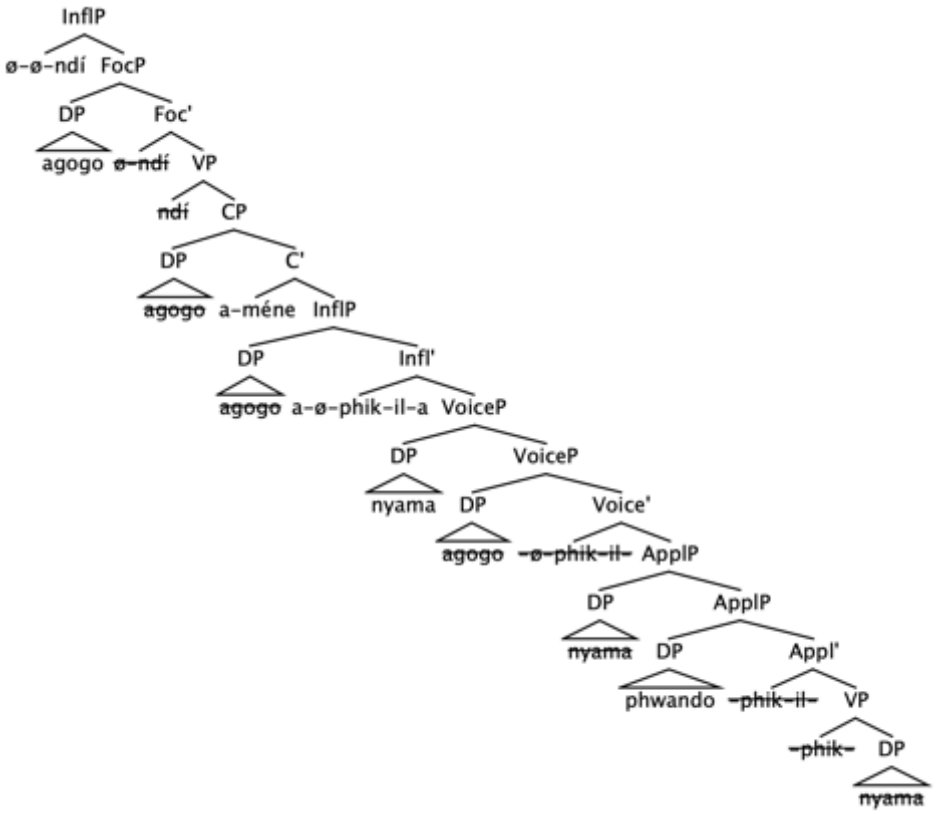
Then the Focus head takes the copular verb as its complement. This head bears an EPP feature that attracts the DP which bears the uninterpretable Focus feature [uFOC:__], namely the agent argument *ágógo* 'old woman'. As a consequence, the [FOC] feature of the Focus head values [uFOC:__] of the focused constituent in a spec-head configuration, so that the uninterpretable focus feature of the DP is deleted in the narrow syntax. The syntactic structure below illustrates this step of the derivation.

(20)



The Focus head is then merged as the complement of the Infl head. At the next step of the derivation, the copular verb is merged into the Infl head. Since this head is defective and does not bear an EPP feature, it does not attract any XP, as shown in the structure below.

(21)



Finally, the inflection projection is taken as the complement of C, followed by the merge of the CP projection in the structure. The resulting construction is sent to spell-out and the derivation is completed, as depicted in the syntactic structure below.

(22)

As seen in Chapter 4, the focused argument may also be realized in sentence-final position. In order to explain this word order, I assume that this clausal pattern corresponds to a pseudocleft construction, in the sense that the embedded clause headed by a CP projection is Ā-moved into spec-InflP of the main clause. The next section will provide a detailed account of the derivation of this construction.

5.3: DERIVATION OF THE PSEUDOCLEFT CONSTRUCTIONS

In Cinyanja, the focused constituent may also occur in sentence-final position. I propose that, in such contexts, there occurs a pseudocleft construction, which is similar to the canonical cleft, the difference being that in the pseudocleft the CP clausal projection (embedded clause) is fronted, that is, it is linearized to the left of the focus constituent. Hence, I will advocate that, in this case, the whole embedded clause is moved into spec-InflP, thereby stranding the focused constituent in sentence-final position. As a result of this operation, the CP is fronted to a position before the focused argument, rendering the word order (COMPLEMENTIZER) + EMBEDDED CLAUSE + COPULAR VERB + FOCUSED CONSTITUENT. Consider the pragmatically structured proposition below.

(170a) Who shaped the waterpot for the child?

(170b) *Wawúmbíla mbíya ya mádzi mwáná, ndí N'ZÍMÁYI.*

w-a-wumb-il-a	*mbiya ya madzi*	*mwana*	*ndi*
1SM-PAST-shape-APPL-FV	3-pot of water	1-child	COP

n'zimayi
1-woman
'Shaped the waterpot for the child was the WOMAN.' (lit.)

In the sentence above the focused constituent *n'zímáyi* 'woman' occurs in sentence-final position immediately after the copular verb *ndí*. A strong piece of evidence in favor of this analysis comes from the sentence (171b) below. Notice that in this sentence the overt complementizer and the focused argument *mbidzi* 'zebra' clearly enter an AGREE relation at the moment the focused subject DP moves across the specifier position of the phase C head before it lands in spec-FocusP and before all the CP is fronted into spec-Infl of the matrix clause. The piece of evidence of this agreement operation has to do with the occurrence of the prefix of class 9 {-y} on the complementizer stem, as observed below.

(171a) *Ciyáni cádya chíre kunkhalángo?*

ciyáni	*cádya*	*chire*	*ku-nkhalango*
what	7SM-PAST-eat-FV	5-grass	LOC-savannah

'What did the zebra eat at the savannah?'

(171b) *Yméne yadya chíre kunkhalángo, ndí MBÍDZI.*

y-mene	*y-a-dy-a*	*chire*	*ku-nkhalango*
9-COMP	9SM-PAST-eat-FV	5-grass	LOC-savannah

ndi	*mbidzi*
COP	9-zebra

'Who ate grass in the savannah was the ZEBRA.'

Regarding the syntactic derivation of pseudocleft sentences, I propose that in this construction the whole embedded clause is moved into spec-InflP,

while the copula *ndí* and the focused constituent are stranded in the sentence-final position.

As for the informational properties of the pseudocleft (171b) above, they can be summarized as follows.

Q: Who ate grass in the savannah?

A: *Ymene yadya chíre kunkhalángo, ndí* [MBÍDZI]$_{FOC}$.

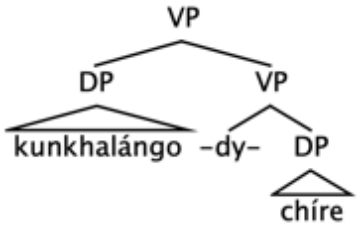
Who ate grass in the savanna was the ZEBRA.

- Presupposition: 'x ate grass in the savannah'
- Assertion: 'x = zebra'
- Focus: zebra
- Focus domain: DP

Thus, I will assume hereafter that, syntactically, sentence (171b) corresponds to a pseudocleft. Based on this, I propose the following derivation, starting with the Numeration step. The main steps of the derivation were already mentioned in the previous sections. I will summarize the most important steps of the derivation that are crucial to our analysis.

The derivation starts at the Numeration. Then, the verb and noun roots are categorized, followed by the first merge of the verbal root *dy* 'eat' with the patient argument *chire* 'grass'. The locative argument *kunkhalángo* 'in the savannah' is then merged with the verbal projection. This argument bears a lexical locative Case assigned by the locative prefix[21].

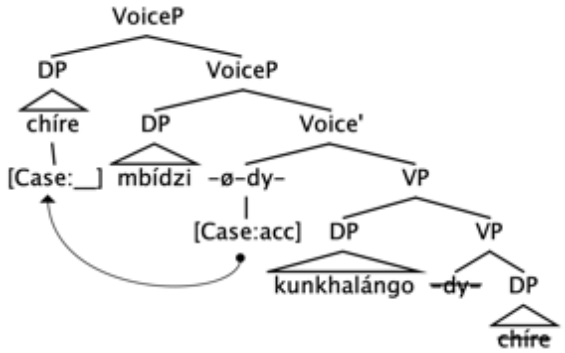
(23)



Active Voice c-selects the VP as its complement, and introduces the agent argument *mbidzi* 'zebra'. The patient argument is raised into the escape hatch of Voice, which assigns accusative Case to this DP in a spec-head configuration. Hence, the patient constituent becomes invisible to further computation. At this point of the derivation the domain of VoiceP is sent to spell-out. The agent argument *mbidzi* 'zebra', by virtue of being at the edge of the phase, is still visible for further operations.

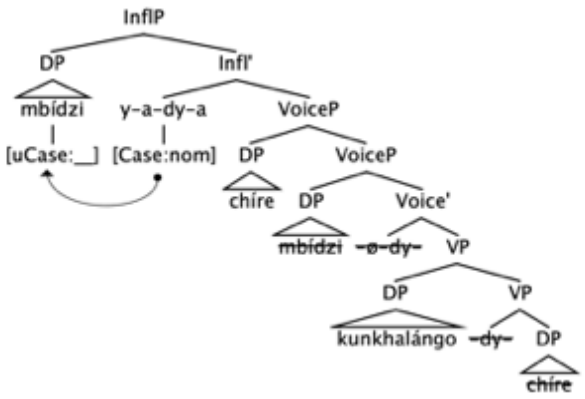
[21] This prefix is akin to a locative preposition, but it behaves as a noun prefix.

(24)



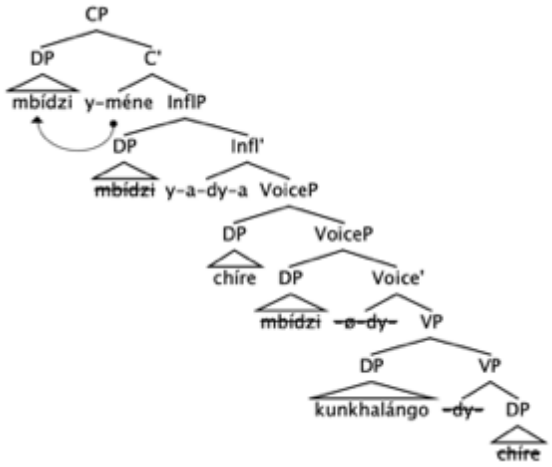
Next, VoiceP is taken as the complement of the Inflection head. Infl bears a [CASE:NOM] feature and an EPP feature. Hence this head attracts the argument that has unvalued uninterpretable Case features. In such situation, it is the agent argument *mbidzi* 'zebra' that bears an interpretable Case feature to be valued and deleted. Therefore, the agent DP must be raised into spec-InflP, where the Infl head values the Case feature of this constituent. This movement also allows for the valuing of the EPP and φ-features of Infl.

(25)



At the next step of the derivation, the complementizer head C c-selects InflP. The focused constituent (the agent *mbídzi* 'zebra') is moved into spec-CP and enters an AGREE relation with the complementizer *méne*. As mentioned above, clear evidence of this agreement comes from the realization of the prefix of class 9 {y} on the complementizer stem. Since CP is a phase head, its domain (=InflP) is sent to spell-out. However, since the focused constituent is at the edge of the C phase, and bears an uninterpretable [uFOC:__] feature, it is still active and visible to higher probes, as depicted below.

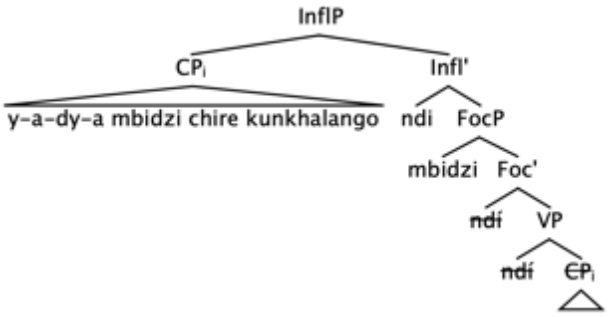
(26)



Next, CP is c-selected by a functional copular verb *ndí*, which is in turn selected by the Focus head. Focus attracts the focused DP into its specifier position due to its EPP feature. In this spec-head configuration, Focus values the uninterpretable [uFOC:__] feature of the focused DP. Then, FocusP is merged as the complement of the Inflection head.

At this point of the derivation, the whole embedded CP *yméne yadya chíre kunkhalángo* 'who ate the grass in the savannah' is Ā-moved into spec-InflP, leaving a silent copy in its base position, and rendering the structure of the pseudocleft, represented by the diagram below.

(27)



It is important to point out that the structure proposed above is the derivation of the following sentence.

(172a) Who ate the grass in the savannah?

(172b) *Yméne yadya chíre kunkhalángo, ndí MBÍDZI.*

y-mene	*y-a-dy-a*	*chire*	*ku-nkhalango*
9-COMP	9SM-PAST-eat-FV	5-grass	LOC-savannah

ndi *mbidzi*
COP 9-zebra
'Who ate grass in the savannah was the ZEBRA.'

As shown in chapter 4, focused arguments may also occur in sentence-medial position, as in the examples below. Notice that the copula *ndí* cannot appear segmentally in sentence-medial position (174).

(173) (What did the girl soak in the sea?)

Ntsíkána, wanyówétsa MÁPÁZI, kúnyánja.

ntsíkána)	*w-a-nyow-ets-a*	*mapazi)*	*ku-nyanja*
1-girl	1SM-PAST-soak-CAUS-FV	6-feet	LOC-sea

'The girl soaked her FEET in the sea.'

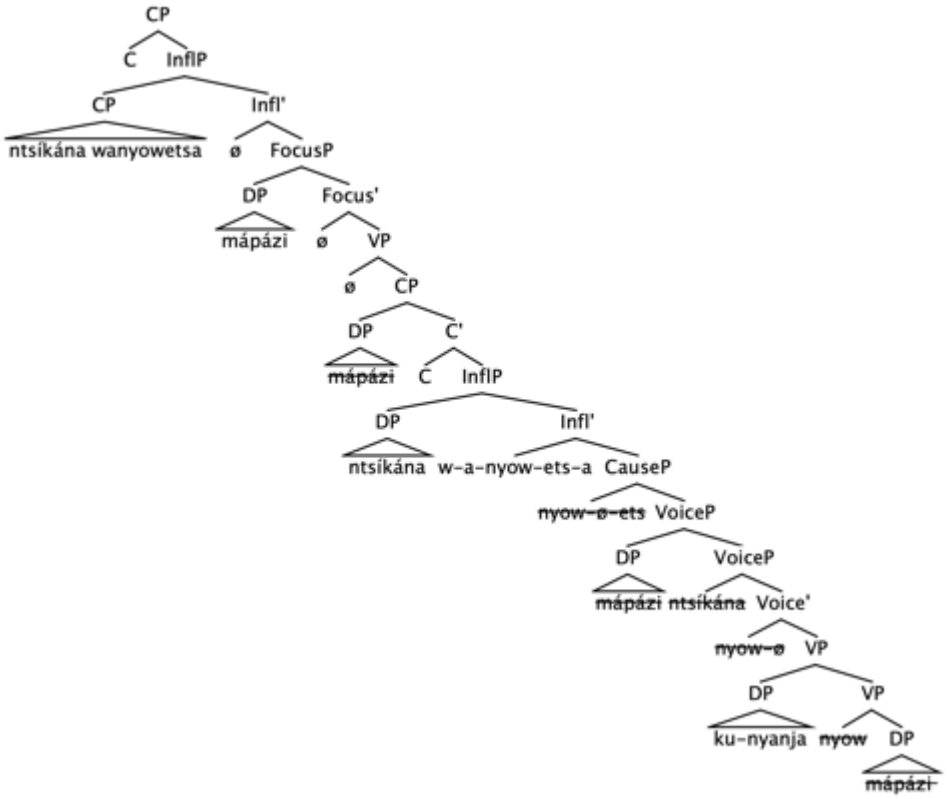
(174) ??*Ntsíkána, wanyówétsa ndí MÁPÁZI, kúnyánja.*

ntsikana	*w-a-nyow-ets-a*	*ndi*	*mapazi*
1-girl	1SM-PAST-soak-CAUS-FV	COP	6-feet

ku-nyanja
LOC-sea
'The girl soaked her FEET in the sea.'

I propose that this example is another instance of a pseudocleft, in which the predicate is fronted but the locative argument stays *in situ*. This situation is analogous to what happens in Tenetehára, according to Duarte (2012, 2018), a language in which part of the predicate can be fronted but adjuncts may stay behind. In this context the locative argument stays *in situ* whereas the rest of the predicate is fronted. Due to some language-particular constraint, the copula *ndí* cannot be segmentally realized in this construction, as shown in example (174) above, but the high floating tone and the insertion of a p-boundary still affect the focused argument. The following figure illustrates this structure.

(28)



In sum, the analysis advocated above has the advantage to allow for the proposal that all focused constituents in the data examined thus far occupy a uniform syntactic position, more precisely the IAV position, which corresponds to a low Focus projection. As already mentioned above, a strong piece of evidence in favor of this analysis comes from the fact that the AGREE syntactic relation is obligatory between the focused constituent and the overtly realized complementizer *méne*. As a conclusion, this in turn signals that the final-sentence

focus argument is not exactly at the end of the sentence but in the specifier of a low Focus projection, that has been stranded in final position (or immediately before a locative argument) after the embedded CP is fronted into spec-InflP of the matrix clause. In short, this fronting case blinds the fact that the focused DP is stranded in spec-FocusP. Furthermore, the fact that the complementizer triggers the appropriate agreement prefix in the sentences above makes it clear that the central thesis of this dissertation is on the right track.

Chapter summary

In this chapter I provided the theoretical formal analysis of the derivation of informational and identificational constituent focus constructions in Cinyanja. As shown in Chapter 4, focus is marked via prosody and syntactic strategies. The prosody effects consist in the presence of a high floating tone which is derived from the occurrence of the copula *ndí* that is elided in many contexts, and in the insertion of p-boundaries flanking the focused argument. Regarding syntactic strategies, this language makes use of a canonical cleft construction and a pseudocleft construction, giving rise to the IAV position. Hence, I propose that all focused arguments, at least in the constructions investigated thus far, occur immediately after a copular verb. We contend that this copula is manifested

segmentally or via prosody effects. I also propose that Focus is projected between InflP and VP in the main clause of a cleft or pseudocleft construction. In all cases shown here, the Focus head takes v*P as its complement; the manifestation of this copular verbal head is *ndí*.

The next chapter will present the final remarks.

CHAPTER 6

FINAL REMARKS

The aim of this dissertation is to investigate the encoding of informational and identificational constituent focus in Cinyanja and propose a minimalist analysis of these constructions. The following questions guided the research put forth here.

i. How to map the position of the focus projection in a sentence that encodes a pragmatically structured proposition in Cinyanja?

ii. How to represent the interaction between syntax and information structure expression in Cinyanja?

iii. To what extent do nonsemantic factors such as information structure and heaviness govern argument realization?

iv. What are the prosodic and syntactic properties of focused constituents in Cinyanja according to Zubizarreta's (2010) framework?

v. How to assess the way Cinyanja fits in the typology of focus expression in Bantu languages proposed by Gibson et al (2017)?

The hypothesis developed in this dissertation is that Cinyanja makes use of a dedicated low Focus position, namely the Immediately After the Verb (IAV)

position, and that the prosodic effects on the focused constituent are due to the presence of a grammaticalized copular verb. I identified five pieces of evidence that my hypothesis is on the right track:

(i) the copula *ndí* bears a high tone which spreads onto the identificational focused argument;

(ii) the copula *ndí* may optionally occur immediately before the focused constituent;

(iii) the focused DP bears a high floating tone, even when the copula is not segmentally realized;

(iv) the complementizer {-*mene*} 'that' is also optional in clefts and pseudoclefts;

(v) when it is overt, the complementizer enters an agreement relation with the focused constituent.

My hypothesis is that the high floating tone borne by the focused argument originates in a grammaticalized copular verb that loses its segmental phonological content in some contexts. I propose that this copular verb is taken as the complement of a functional Focus projection that is linked to phonological

boundaries (p-boundaries) flanking the focused argument. Hence, my hypothesis is confirmed.

The answer to question (i) may be expressed as follows: as shown in Chapter 4, focused arguments may appear either in sentence-initial, sentence-medial, or sentence-final position. I argue that these positions are only superficial. The core of this dissertation is that, in fact, these arguments occur in IAV position, and the different positions are the result of syntactic processes. That is, focused arguments are systematically raised into the specifier of a low Focus projection, which is mapped above a copular verb and below the Inflection projection. My hypothesis is that, in some instances, the copula *ndí* lost segmental phonological material and is realized via prosody effects such as high floating tone spreading.

The answer to question (ii) may be expressed as follows: based on Aboh (2010) and Rizzi (1997), I assume that informational features are already present in the lexicon and project in the syntax. These features drive syntactic operations such as movement and agreement. Moreover, given the fact that the form of the question determines the form of the answer, I propose that conceptualization of the information to be encoded in the sentence occurs during the Numeration step of the derivation, that is, when the speaker selects the appropriate roots and bundles of features to organize the utterance.

The answer to question (iii) is hinted at in the answer to question (ii), that is, I propose that the fact that informational features are manipulated in the computation accounts for the way arguments are realized. Features such as Focus may have an EPP component which motivates XP movement in the course of the derivation, for instance.

Regarding the answer to question (iv), I argue that the prosodic properties of focused constituents in Cinyanja occur just as proposed by Zubizarreta (2010), that is, the argument is marked prosodically via high floating tone spreading and insertion of a p-boundary. The high floating tone comes from the copular verb, while the p-boundaries flanking the DP come from the functional Focus projection. Syntactically, this language corroborates the proposal according to which there is a dedicated low Focus position in Bantu languages (cf. Duarte and Langa (2025), among others).

As for question (v), regarding the typology of Focus expression in Bantu proposed by Gibson et al (2017) (cf. section 3.5.3), Cinyanja presents the following properties: the focused DP occurs in Immediately After the Verb (IAV) position, and although there is no morphological marking on either the verb nor the constituent, the focused argument is marked prosodically, via the insertion of a phonological boundary (pause) and the occurrence of a high floating tone that

spreads to the penultimate syllable of the focused DP. Thus, we conclude that Cinyanja is an IAV language with marking on the constituent (cf. Table 8).

I propose that the difference between the cleft and pseudocleft constructions is due to the existence of three distinct Inflection heads in the lexicon, according to their formal features: (i) a full active Infl that bears EPP, Case, and φ-features and attracts DPs to its specifier position; (ii) a partially defective Infl which bears EPP features but no Case nor φ-features, and thus may attract the embedded CP to its specifier position; and (iii) a full defective Infl which does not bear EPP, Case nor φ-features and therefore does not project a specifier position. The choice between the different constructions used to mark identificational constituent focus in Cinyanja happens in the Numeration step of the derivation, as mentioned above.

The hypothesis advanced here also confirms Carstens' (2005) proposal that AGREE in Bantu occurs in spec-head relations. Additionally, I assume, together with Carstens, that functional heads such as FocusP and CP bear formal features that force the raising of XPs to their specifier positions, i.e. they consist in EPP features as well. However, as mentioned previously, I propose the existence of a defective Inflection head which does not bear EPP, nor φ-features, and therefore does not project a specifier position.

Regarding the MLC, I propose the following explanation for the perceived violations when two arguments occupy the specifier of phase heads such as Voice and Appl: since movement is driven by particular features such as Case, Focus, or φ-features, the argument that bears the appropriate feature will be probed by the appropriate head, regardless of its position at the edge of the phase, if it is still active (visible). Thus, when the agent argument bears an uninterpretable Case feature it must be raised into spec-InflP in order to value this feature; arguments which do not bear an uninterpretable Case feature will not be visible to this probe. On the other hand, the focused argument must be always available to probes such as C and Focus, due to its uninterpretable Focus feature that must be valued and deleted.

The final goal of this dissertation is to contribute to the knowledge and documentation of Cinyanja as spoken in Mozambique. As my work brings original data and a new analysis of the informational and syntactic properties of this language, I contend that I help shed light into the structure of Cinyanja.

REFERENCES

ABOH, Enoch Oladé. Information structuring begins with the numeration. **Iberia: An International Journal of Theoretical Linguistics**, 2010.

ALMEIDA, J. C. **Um estudo das estruturas clivadas do Português.** Brasília: Universidade de Brasília. 2011.

ARAD, Maya. Locality constraints on the interpretation of roots: The case of Hebrew denominal verbs. **Natural language & linguistic theory**, v. 21, n. 4, p. 737–778, 2003.

BAKER, Mark C. **Incorporation: A theory of grammatical function changing**. Chicago: University of Chicago Press, 1988.

BEAVER, David I.; CLARK, Brady Z. **Sense and sensitivity: How focus determines meaning**. *[S.l.]*: John Wiley & Sons, 2009.

BIRIATE, Mário. A sintaxe das orações relativas em Cinyanja. 2023.

BLEEK, Wilhelm HI. **A comparative grammar of South African Languages. 1, The noun: 1. Phonology**. *[S.l.]*: Trübner & Company, 1869.

BRESNAN, Joan; KANERVA, Jonni M. Locative inversion in Chichewa: A case study of factorization in grammar. **Linguistic inquiry**, p. 1–50, 1989.

BRESNAN, Joan; MOSHI, Lioba. Object asymmetries in comparative Bantu syntax. **Linguistic inquiry**, v. 21, n. 2, p. 147–185, 1990.

CANÇADO, M; AMARAL, L.; MEIRELLES, L. **VerboWeb: classificação sintático-semântica dos verbos do português brasileiro.** Banco de dados lexicais. UFMG. Disponível em: http://www.letras.ufmg.br/verboweb. 2017.

CANÇADO, M.; GODOY, L.; AMARAL, L. **Catálogo de verbos do português brasileiro:** classificação verbal segundo a decomposição de predicados (Vol. I - Verbos de mudança). 2a Edição revisada. 2017.

CANÇADO, Márcia; AMARAL, Luana; MEIRELLES, Letícia Lucinda. Verboweb: Uma proposta de classificação verbal. **Revista da Anpoll**, v. 1, n. 46, p. 123–141, 1 out. 2018.

CANÇADO, Márcia; AMARAL, Luana. **Introdução à semântica lexical: papéis temáticos, aspecto lexical e decomposição de predicados**. *[S.l.]*: Editora Vozes Limitada, 2016.

CARSTENS, Vicki. Agree and EPP in Bantu. **Natural Language & Linguistic Theory**, v. 23, n. 2, p. 219–279, 2005.

CASTANEDA, H. Comments on Donald Davisnons The Logical Form of Action Sentences. The Logic of Decision and Action, ed. by Licholas Rescher, 104-112. 1967.

CHAFE, Wallace. Cognitive constraints on information flow. Coherence and grounding in discourse, ed. by Russell Tomlin, 21-51. **Amsterdam & Philadelphia: John Benjamins**, 1987.

CHAFE, Wallace. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. **Subject and topic**, 1976.

CHOMSKY, N. Chomsky's linguistics. *In*: GRAFF, Peter; VAN URK COPPE (Orgs.). 2012.

CHOMSKY, N. Derivation by Phase. *In*: KENSTOWICZ, M. (Org.). **Ken Hale: A Life in Language**. Cambridge, MA: MIT Press, 2001. p. 1–52.

CHOMSKY, N. Minimalist inquiries: the framework. *In*: MARTIN, R.; MICHAELS, D.; URIAGEREKA, J. (Orgs.). **Step by Step: Essays on**

Minimalist Syntax in Honor of Howard Lasnik. Cambridge, MA: MIT Press, 2000. p. 89–155.

CHOMSKY, N. **The Minimalist Program**. Cambridge, MA: MIT Press, 1995.

CHOMSKY, Noam. Beyond explanatory adequacy. **Structures and beyond: The cartography of syntactic structures**, v. 3, p. 104–131, 2004.

CHOMSKY, Noam. Conditions on Transformations. An Annotated Syntax Reader. p. 58, 1973.

CHOMSKY, Noam. Lectures on government and binding. Dordrecht: Foris.-. 1986. Knowledge of language: Its nature, origin, and use. **New York: Praeger..(1993)" A minimalist program for linguistic theory". Kenneth Hale and Samuel J. Keyser, eds. The view from building**, v. 20, p. 1–52, 1981.

CHOMSKY, Noam. On phases. **Current Studies in Linguistics Series**, v. 45, p. 133, 2008.

DAVIDSON, D. **The logical form of active sentences**. In: Nicholas Rescher (ed.) *The logic of decision and action*. University of Pittsburgh Press, Pittsburgh, 1967. pp. 81-95.

DOWNING, Laura J.; MTENJE, Alfred James Dailex. **The phonology of Chichewa**. *[S.l.]*: Oxford University Press, 2017.

DOWTY, David R. On the semantic content of the notion of 'thematic role'. *In*: **Properties, types and meaning: Volume II: Semantic issues**. *[S.l.]*: Springer, 1989. p. 69–129.

DOWTY, David. Thematic proto-roles and argument selection. **language**, v. 67, n. 3, p. 547–619, 1991.

DUARTE, Fábio Bonfim. Clausal recursion, predicate raising and head-finality in Tenetehára. **Recursion across borders and beyond**, p. 143–165, 2018.

DUARTE, Fábio Bonfim. Tenetehára: A predicate-fronting language. **Canadian Journal of Linguistics/Revue canadienne de linguistique**, v. 57, n. 3, p. 359–386, 2012.

DUARTE, F. B.; LANGA, D. A. S. **Disjoint/conjoint alternation and the low focus position in Xichangana (S53).** 2025.

DUIGNAN, Brian. **Occam's razor**. In: *Encyclopedia Britannica*, 20 Mar. 2025, https://www.britannica.com/topic/Occams-razor. Accessed 22 June 2025.

ERTESCHIK-SHIR, Nomi. Wh-Questions and Focus. **Linguistics and Philosophy**, v. 9, n. 2, p. 117–149, 1986.

FILLMORE, Charles J. The case for case. 1967.

GALLEGO, Ángel J. Roots and phases. In: **The Syntax of Roots and the roots of syntax**. Artemis Alexiadou, Hagit Borer, Florian Schäfer (eds.) 2015.

GIBSON, Hannah *et al.* Locating the Bantu conjoint/disjoint alternation in a typology of focus marking. In: **The Conjoint/Disjoint Alternation in Bantu**. Jenneke van der Wal, Larry M. Hyman (eds). 2017.

GUESSER, Simone. Redução e concordância em sentenças clivadas do Português Brasileiro. NAVES, Rozana Reigota; LIMA-SALLES, Heloisa Maria. **Estudos Formais da Gramática das Línguas Naturais. Goiânia: Cânone Editorial**, 2008.

GUTHRIE, M. The classification of the Bantu languages. 1948.

GUTHRIE, Malcolm. The status of radical extensions in Bantu languages. **Journal of African Languages**, v. 1, n. 3, p. 202–220, 1962.

GUTHRIE, Malcolm. Variations in the range of classes in the Bantu languages. **La classification nominale dans les langues négro-africaine**, p. 341–348, 1967.

HAJIČOVÁ, Eva. **Issues of sentence structure and discourse patterns**. *[S.l.]*: Charles University, 1993. v. 2

HALE, Ken; KEYSER, Samuel Jay. On argument structure and the lexical expression of syntactic relations. *In*: HALE, Ken; KEYSER, Samuel Jay (Orgs.). **The view from the building 20**. Cambridge: MIT Press, 1993. p. 312–327.

HALE, Ken; KEYSER, Samuel Jay. **Prolegomenon to a theory of argument structure**. *[S.l.]*: MIT press, 2002.

HALLE, Morris *et al.* Distributed morphology and the pieces of inflection. **1993**, p. 111–176, 1993.

HALLE, Morris; MARANTZ, Alec. Some key features of Distributed Morphology. **MIT working papers in linguistics**, v. 21, n. 275, p. 88, 1994.

HEIM, Irene; KRATZER, Angelika. Semantics in generative grammar. 1998.

HIRAIWA, Ken. **Dimensions of symmetry in syntax: Agreement and clausal architecture**. PhD Thesis—*[S.l.]*: Massachusetts Institute of Technology, 2005.

HYMAN, Larry M. Phonology: theory and analysis. **(No Title)**, 1975.

ISHII, Toru. The Minimal Link Condition and the Theory of Movement. v. 17, p. 305–329, 1 jan. 2000.

KANERVA, Jonni M. Focusing on phonological phrases in Chichewa. **The phonology-syntax connection**, v. 145, p. 145–163, 1990.

KATO, Mary A. TWO FOCUS POSITIONS IN THE HISTORY OF BRAZILIAN. 2015.

KATO, Mary. A mudança na posição estrutural de foco no português brasileiro. **Revista Linguíʃtica**, v. 13, n. 2, p. 158, 1 set. 2017.

KISS, Katalin É. Identificational focus versus information focus. **Language**, v. 74, n. 2, p. 245–273, 1998.

KRAMER, Ruth T. **The morphosyntax of gender**. *[S.l.]*: Oxford University Press, 2015.

KRAMER, Ruth. Gender in Amharic: A morphosyntactic approach to natural and grammatical gender. **Language sciences**, v. 43, p. 102–115, 2014.

KRATZER, Angelika. Severing the external argument from its verb. **Phrase structure and the lexicon**, p. 109–137, 1996.

KRIFKA, Manfred; MUSAN, Renate. **The expression of information structure**. *[S.l.]*: Mouton de Gruyter Berlin/New York, 2012.

LAKOFF, George. Irregularity in Syntax. 1970.

LAMBRECHT, Knud. A framework for the analysis of cleft constructions. 2001.

LAMBRECHT, Knud. **Information structure and sentence form: Topic, focus, and the mental representations of discourse referents**. *[S.l.]*: Cambridge university press, 1994b. v. 71

LEGATE, J. A. Morphological and abstract case. **Linguistic Inquiry**, v. 39, p. 55–101, 2008.

LEVIN, Beth; HOVAV, Malka Rappaport. **Argument realization**. *[S.l.]*: Citeseer, 2005. v. 10

LEWIS, David. Scorekeeping in a language game. **Journal of philosophical logic**, v. 8, p. 339–359, 1979.

LI, Charles N.; THOMPSON, Sandra. The semantic function of word order: a case study in Mandarin. 1975.

MACALANE, Geraldo L. **A variação paramétrica das interrogativas parciais em Cinyanja**. *[S.l.]*: Centro de Estudos Africanos (CEA)-UEM, 2013. v. 9

MARANTZ, A. **On the Nature of Grammatical Relations**. Cambridge, MA: MIT Press, 1984.

MARANTZ, Alec. On the nature of grammatical relations. 1981.

MATTHEWSON, Lisa. On the methodology of semantic fieldwork. **International journal of American linguistics**, v. 70, n. 4, p. 369–415, 2004.

MATEU FONTANALS, Jaume. **Argument structure: Relational construal at the syntax-semantics interface**. *[S.l.]*: Universitat Autònoma de Barcelona, 2002.

MCCAWLEY, James D. Lexical insertion in a transformational grammar without deep structure. 1968.

MCGINNIS, Martha. Applicatives. **Language and Linguistics Compass**, v. 2, n. 6, p. 1225–1245, 2008.

MCGINNIS, Martha. UTAH at Merge: Evidence from multiple applicatives. **MIT Working Papers in Linguistics**, v. 49, p. 183–200, 2005.

MCGINNIS, Martha. Variation in the phase structure of applicatives. **Linguistic variation yearbook**, v. 1, n. 1, p. 105–146, 2001.

MCHOMBO, Sam. **The syntax of Chichewa**. *[S.l.]*: Cambridge University Press, 2004.

MEEUSSEN, Achille Emiel. Bantu grammatical reconstructions. In: Africana linguistica III, p 79-121. Annalen van het Koninklijk Museum voor Midden-Afrika, menselijke wetenschappen, n 61. Tervuren. 1967.

MYERS, Scott P. **Tone and the structure of words in Shona**. *[S.l.]*: University of Massachusetts Amherst, 1987.

MYERS, Scott. OCP effects in Optimality Theory. **University of Texas, Austin**, 1994.

NGUNGA, A.; FAQIR, O. G. Padronização da ortografia de línguas moçambicanas: relatório do III Seminário. *[S.d.]*.

NGUNGA, Armindo; MANUEL, Carlos; LANGA, David; MACHUNGO, Inês; CÂMARA, Crisófia L. da (eds). **Padronização da ortografia de línguas moçambicanas: relatório do IV seminário**. Maputo: Universidade Eduardo Mondlane (UEM), Centro dos Estudos Africanos (CEA). 2023.

NGUNGA, Armindo; SIMBINE, Madalena Citia. **Gramática descritiva da língua changana**. *[S.l.]*: Centro de Estudos Africanos, Universidade Eduardo Mondlane, 2012.

NGUNGA, Armindo. Introdução à linguística bantu. 2004.

ODDEN, David. On the role of the Obligatory Contour Principle in phonological theory. **Language**, p. 353–383, 1986.

PARK, Sangbaek. Situational Analysis of Primary Education in Mozambique. **Journal of International Development Cooperation**, v. 12, n. 3, p. 137–162, 2017.

PARSONS, Terence. Events in the semantics of English: A study in subatomic semantics. 1990.

PESETSKY, David. Language-particular processes and the Earliness Principle. Ms. **Ca, bridge, Mass.: MIT**, 1989.

PIKE, Kenneth L. Tone languages. **University of Michigan Publications in Linguistics**, v. 4, 1948.

PRINCE, Ellen F. Toward a taxonomy of given-new information. **Radical pragmatics**, 1981.

PYLKKÄNEN, Liina. **Introducing Arguments**. Cambridge: MIT press, 2008. v. 49

RIZZI, Luigi. The fine structure of the left periphery. *In*: **Elements of grammar: Handbook in generative syntax**. Dordrecht: Springer, 1997. p. 281–337.

ROCHA, B. G. **Aplicativos: definição e propriedades**. Monograph. Belo Horizonte: Universidade Federal de Minas Gerais. 2014.

ROCHA, B. G.; BEWALA, K. **The informational properties of applicatives in Cinyanja (Bantu).** To appear.

RODRIGUES, Gabriel Roisemberg; MENUZZI, Sérgio de Moura. Estrutura informacional. **Manuscrito. Michighan State University**, 2009.

ROOTH, Mats. Association with focus. 1985.

SITOE, Bento; NGUNGA, Armindo. Relatório do II seminário sobre a padronização da ortografia de línguas moçambicanas. **Maputo: NELIMO/UEM**, 2000.

SVENONIUS, Peter. On the edge. *In*: **Peripheries: Syntactic edges and their effects**. *[S.l.]*: Springer, 2004. p. 259–287.
the Roots of Syntax. Artemis Alexiadou, Hagit Borer, Florian Schäfer (eds.). Oxford University Press. 2015, pp. 192–208.

VALLDUVÍ, Enric; ENGDAHL, Elisabet. The linguistic realization of information packaging. 1996.

VAN DER WAL, Jenneke; FUCHS, Suzanna. Gender on n in Bantu DP structure: from root derived nominals to locatives. **Paper presented at CamCoS 8**, 2019.

VAN DER WAL, Jenneke. Bantu Syntax. *In*: OXFORD HANDBOOKS EDITORIAL BOARD (Org.). **Oxford Handbook Topics in Linguistics**. *[S.l.]*: Oxford University Press, *[S.d.]*. p. 0.

WILLIAMS, Edwin. Argument structure and morphology. 1981.

ZUBIZARRETA, M. Luisa. The syntax and prosody of focus: the Bantu-Italian connection. **Iberia: An International Journal of Theoretical Linguistics**, 2010.

ANNEX 1 – QUESTIONNAIRE

Questionário – relações entre estrutura argumental e estrutura informacional

Bárbara Guimarães Rocha
LALIAFRO/LETRAS/UFMG

1. Introdução

A relação entre a forma das sentenças e a codificação dos referentes e a estruturação da informação no discurso vem sendo investigada há várias décadas. Destaca-se o trabalho de Lambrecht (1994), um manual muito didático e compreensivo sobre os mecanismos de tal relação, onde se vê as noções de representação, princípios discursivos e noções como tópico e foco.

Lambrecht (1994) indiretamente relaciona a noção de tópico com várias propriedades que são consideradas evidências de relações argumentais e de concordância, tais como marcas de objeto que podem ser analisadas como pronominalização em línguas Bantu (Givón 1976), e a possibilidade de apagamento de constituintes. Tal relação merece ser investigada com mais profundidade, de forma a tentar relacionar a sintaxe e a estrutura informacional.

Lambrecht (1994) argumenta que essa relação entre sintaxe e estrutura informacional deve ser investigada através do epifenômeno das alossentenças – variações formais do mesmo valor de verdade que são mais apropriadas em certos contextos discursivos. De forma a captar essas variações, o instrumento aqui apresentado busca elicitar respostas para diferentes perguntas que perfazem o mesmo valor de verdade, porém com estruturações informacionais distintas. Cada pergunta busca elicitar um tipo de resposta de acordo com a informação que se pede: alguns elementos fazem parte da pressuposição (estão presentes na pergunta), enquanto outros perfazem a informação que se pede em foco (constitui na variável que deve ser apresentada como “nova” no discurso).

A recolha de dados de diversas línguas que permita investigar tal relação faz-se imperativa. Para tanto, desenvolvi um instrumento de coleta que busca facilitar tal recolha de dados e aprofundar o conhecimento das línguas em análise, assim como permitir a criação de um banco de dados compreensivo e que possa ser acessado por outros pesquisadores.

2. O instrumento de coleta de dados

O presente questionário pretende encontrar relações entre como argumentos são introduzidos em uma sentença e a estrutura informacional dessas sentenças.

O objetivo é encontrar as sentenças que são mais adequadas em cada situação discursiva, por meio de perguntas e respostas, e encontrar os mecanismos de introdução e referenciação dos argumentos de cada resposta.

Serão apresentadas 66 sentenças apresentacionais, na forma ativa, organizadas em 20 classes de acordo com as propriedades sintáticas e semânticas dos verbos envolvidos. Os verbos foram selecionados na plataforma Verboweb (Cançado, Amaral e Meirelles, 2017), e são classificados de acordo com as autoras com base em suas propriedades semânticas e sintáticas. Foram coletados verbos das seguintes classes:

Classe 1	inergativos
Classe 2	instrumento incorporado
Classe 3	contato mediado por instrumento
Classe 4	contato mediado pelo corpo
Classe 5	meteorológicos
Classe 6	mudança de estado opcionalmente agentivos
Classe 7	mudança de estado não-agentivo

Classe 8	mudança de estado locativo
Classe 9	mudança de estado de posse
Classe 10	locativos
Classe 11	locatum
Classe 12	transferência do tipo locatum
Classe 13	criação
Classe 14	mudança de estado
Classe 15	mudança de lugar
Classe 16	mudança de posse
Classe 17	mudança de estado mental
Classe 18	psicológicos complexos
Classe 19	psicológicos
Classe 20	existenciais

As perguntas buscam elicitar como os argumentos são organizados na sentença que é adequada em cada contexto. Assim, busca-se ver como são codificados elementos que já estão na pergunta (pressuposição/tópico) ou que constituam o "cerne" da resposta (comentário/foco). A hipótese assumida nesse questionário prevê que diferentes estruturas sintáticas são adequadas a cada contexto específico, e diferentes estratégias sintáticas serão utilizadas em cada resposta.

O objetivo do questionário é que tais sentenças apresentacionais sejam traduzidas para a língua alvo, e em seguida, que as perguntas sobre a sentença sejam respondidas também na língua alvo. Escolha o verbo que for apropriado para a situação, mesmo que seja diferente do exemplo dado. Por favor, utilize sentenças completas para cada resposta (realizando todos os argumentos). Favor também indicar os tons nas sentenças. É desejável que seja feita a glosa com a separação dos morfemas.

Veja abaixo um exemplo de como os dados devem ser apresentados:

1. O leão quebrou a cerca para as zebras.

Mkángo	unathyólétsá	mbidzí	mpánda
Mkángo	u-na-thyól-éts-á	mbidzí	mpánda
3-leão	3SM-PST-quebrar-APPL-FV	10-zebras	3-cerca

‘O leão quebrou a cerca para as zebras’

a. O que aconteceu com o leão?

R: O leão quebrou a cerca para as zebras.

R: Ele quebrou a cerca para as zebras.

b. O que aconteceu com a cerca?

R: A cerca foi quebrada pelo leão para as zebras.

R: Ela foi quebrada pelo leão para as zebras.

c. O que aconteceu com as zebras?

R: Para as zebras, o leão quebrou a cerca.

R: Para elas o leão quebrou a cerca.

d. Para quem o leão quebrou a cerca?

R: O leão quebrou a cerca para as zebras.

R: Para as zebras, o leão quebrou a cerca.

Perceba que em (1c) *as zebras* faz parte da pergunta, enquanto em (1d) busca-se que a resposta mostre esse constituinte como uma informação “nova”, ou a identificação de um elemento de um conjunto proposto pela pergunta. Espera-se que a codificação do constituinte *zebras* seja diferente nas duas sentenças, ou que a estrutura sintática seja distinta de forma a refletir a organização da informação nas sentenças, ou seja, a identificação dos elementos tópicos ou focalizados.

Segue abaixo o questionário. Não se esqueça de escrever respostas completas para as perguntas e indicar os tons de cada sentença. Caso possível, forneça a separação dos morfemas.

Classe 1

1. **O menino pulou a cerca.**
    a. O que aconteceu com o menino?
    b. O que aconteceu com a cerca?
    c. O que o menino pulou?
    d. Quem pulou a cerca?

2. **O vizinho andou até o mercado.**
    a. O que aconteceu com o vizinho?
    b. O que aconteceu com o mercado?
    c. Quem andou até o mercado?
    d. Até onde o vizinho andou?

3. **A mãe sorriu para o filho.**
    a. O que aconteceu com a mãe?
    b. O que aconteceu com o filho?
    c. Quem sorriu para o filho?
    d. Para quem a mãe sorriu?

4. **A menina dança balé.**
    a. O que acontece com a menina?
    b. O que acontece com o balé?
    c. Quem dança balé?
    d. O que a menina faz?

5. **O pai sonha com os filhos.**
    a. O que acontece com o pai?
    b. O que acontece com os filhos?
    c. Quem sonha com os filhos?
    d. Com o que o pai sonha?

Classe 2

1. **A criança está escovando o cachorro.**
    a. O que está acontecendo com a criança?
    b. O que está acontecendo com o cachorro?
    c. Quem a criança está escovando?
    d. Quem está escovando o cachorro?

2. **O fazendeiro apunhalou o porco para a festa.**
    a. O que aconteceu com o fazendeiro?
    b. O que aconteceu com o porco?
    c. Para que o fazendeiro apunhalou o porco?
    d. O que o fazendeiro fez com o porco?

Classe 3

1. **O marido moeu os grãos para a esposa.**
    a. O que aconteceu com o marido?
    b. O que aconteceu com a esposa?
    c. O que aconteceu com os grãos?
    d. Quem moeu os grãos?
    e. O que o marido moeu?
    f. Para quem o marido moeu os grãos?

2. **O jardineiro podou os arbustos para o cliente.**
    a. O que aconteceu com o jardineiro?
    b. O que aconteceu com os arbustos?
    c. O que aconteceu com o cliente?
    d. O que o jardineiro podou para o cliente?
    e. Para quem o jardineiro podou os arbustos?
    f. Quem podou os arbustos para o cliente?

Classe 4

1. **O pai abraçou a filha.**
    a. O que aconteceu com o pai?
    b. O que aconteceu com a filha?
    c. Quem o pai abraçou?
    d. Quem abraçou a filha?

2. **O gato mordeu o dono.**
    a. O que aconteceu com o gato?
    b. O que aconteceu com o dono?
    c. Quem mordeu o dono?
    d. Quem o gato mordeu?

3. **A mãe estapeou o filho.**
    a. O que aconteceu com a mãe?
    b. O que aconteceu com o filho?
    c. Quem estapeou o filho?
    d. Quem a mãe estapeou?

Classe 5

1. **Choveu forte ontem em Maputo.**
    a. O que aconteceu ontem?
    b. O que aconteceu em Maputo?
    c. Quando aconteceu a chuva forte em Maputo?
    d. Onde aconteceu a chuva forte?

2. **Amanhã vai ventar forte em Maputo.**
    a. O que vai acontecer amanhã?
    b. O que vai acontecer em Maputo?
    c. Onde vai ventar forte amanhã?
    d. Quando vai ventar forte em Maputo?

Classe 6

1. **O João quebrou o vaso com o martelo.**
    a. O que aconteceu com João?
    b. O que aconteceu com o vaso?
    c. O que aconteceu com o martelo?
    d. Quem quebrou o vaso?
    e. Que instrumento foi usado para quebrar o vaso?
    f. O que João fez com o vaso?

2. **O limão coagulou o leite.**
    a. O que aconteceu com o limão?
    b. O que aconteceu com o leite?
    c. O que coagulou o leite?
    d. O que o limão coagulou?

3. **A mulher derreteu chocolate para o bolo.**
    a. O que aconteceu com a mulher?
    b. O que aconteceu com o chocolate?
    c. O que aconteceu com o bolo?
    d. Quem derreteu o chocolate para o bolo?
    e. O que a mulher derreteu para o bolo?

4. **O professor espalhou livros pela casa.**
    a. O que aconteceu com o professor?
    b. O que aconteceu com os livros?
    c. O que aconteceu na casa?
    d. Quem espalhou os livros pela casa?
    e. O que o professor espalhou pela casa?
    f. Onde o professor espalhou os livros?

5. **O anfitrião trancou a porta com a chave.**
    a. O que aconteceu com o anfitrião?
    b. O que aconteceu com a porta?
    c. O que aconteceu com a chave?
    d. Quem trancou a porta com a chave?
    e. O que o anfitrião trancou com a chave?

Classe 7

1. **O pescador resfriou o peixe no freezer.**
    a. O que aconteceu com o pescador?
    b. O que aconteceu com o peixe?
    c. O que aconteceu no freezer?

d. Quem resfriou o peixe no freezer?
e. O que o pescador resfriou no freezer?

2. A prova cansou os alunos.

a. O que aconteceu com a prova?
b. O que aconteceu com os alunos?
c. Quem se cansou com a prova?
d. Com o que os alunos se cansaram?

3. O remédio cicatrizou a ferida rápido.

a. O que aconteceu com a ferida?
b. O que aconteceu com o remédio?
c. Como a ferida cicatrizou?

Classe 8

1. O empresário depositou o dinheiro no cofre.

a. O que aconteceu com o empresário?
b. O que aconteceu com o dinheiro?
c. O que aconteceu no cofre?
d. Quem depositou o dinheiro no cofre?
e. O que o empresário depositou no cofre?
f. Onde o empresário depositou o dinheiro?

2. A criança largou os brinquedos no chão.

a. O que aconteceu com a criança?
b. O que aconteceu com os brinquedos?
c. O que aconteceu no chão?
d. O que a criança largou no chão?

e. Quem largou os brinquedos no chão?
f. Onde a criança largou os brinquedos?

3. **O homem prendeu os cachorros no canil.**
   a. O que aconteceu com o homem?
   b. O que aconteceu com os cachorros?
   c. O que aconteceu no canil?
   d. Onde o homem prendeu os cachorros?
   e. Quem prendeu os cachorros no canil?
   f. O que o homem prendeu no canil?

Classe 9

1. **A chuva ensopou a roupa no varal.**
   a. O que aconteceu com a chuva?
   b. O que aconteceu com a roupa?
   c. O que aconteceu no varal?
   d. Onde a chuva ensopou a roupa?
   e. Quem ensopou a roupa no varal?
   f. O que a chuva ensopou?

2. **O estudante entupiu a lixeira com papel.**
   a. O que aconteceu com o estudante?
   b. O que aconteceu com a lixeira?
   c. O que aconteceu com o papel?
   d. O que o estudante fez com a lixeira?
   e. Quem entupiu a lixeira com papel?
   f. Com o que o estudante entupiu a lixeira?

3. **O barro sujou a roupa no varal.**
    a. O que aconteceu com o barro?
    b. O que aconteceu com a roupa?
    c. O que aconteceu no varal?
    d. O que o barro sujou no varal?
    e. O que sujou a roupa no varal?

Classe 10

1. **O operário enlatou a sardinha.**
    a. O que aconteceu com o operário?
    b. O que aconteceu com a sardinha?
    c. O que o operário enlatou?
    d. Quem enlatou a sardinha?

2. **O caçador engaiolou o passarinho.**
    a. O que aconteceu com o caçador?
    b. O que aconteceu com o passarinho?
    c. Quem engaiolou o passarinho?
    d. O que o caçador engaiolou?

3. **A bibliotecária encaixotou os livros na biblioteca.**
    a. O que aconteceu com a bibliotecária?
    b. O que aconteceu com os livros?
    c. O que aconteceu na biblioteca?
    d. Quem encaixotou os livros na biblioteca?
    e. O que a bibliotecária encaixotou na biblioteca?
    f. Onde a bibliotecária encaixotou os livros?

Classe 11

1. **A vendedora embrulhou o presente para a cliente.**
    a. O que aconteceu com a vendedora?
    b. O que aconteceu com o presente?
    c. O que aconteceu com a cliente?
    d. O que a vendedora embrulhou para a cliente?
    e. Quem embrulhou o presente para a cliente?
    f. Para quem a vendedora embrulhou o presente?

2. **A criança rabiscou o caderno.**
    a. O que aconteceu com a criança?
    b. O que aconteceu com o caderno?
    c. Quem rabiscou o caderno?
    d. O que a criança rabiscou?

3. **O bispo coroou a rainha com a tiara de ouro branco.**
    a. O que aconteceu com o bispo?
    b. O que aconteceu com a rainha?
    c. O que aconteceu com a tiara?
    d. Quem o bispo coroou com a tiara de ouro branco?
    e. Quem coroou a rainha com a tiara de ouro branco?
    f. O que o bispo usou para coroar a rainha?

Classe 12

1. **O juiz condenou o réu.**
    a. O que aconteceu com o juiz?
    b. O que aconteceu com o réu?

c. Quem o juiz condenou?

d. Quem condenou o réu?

2. **O concurso premiou o vencedor com uma estátua de bronze.**

a. O que aconteceu com o concurso?

b. O que aconteceu com o vencedor?

c. O que aconteceu com a estátua de bronze?

d. Quem premiou o vencedor com uma estátua de bronze?

e. Quem o concurso premiou com uma estátua de bronze?

f. Com o que o concurso premiou o vencedor?

3. **O padre abençoou o doente.**

a. O que aconteceu com o padre?

b. O que aconteceu com o doente?

c. Quem abençoou o doente?

d. Quem o padre abençoou?

Classe 13

1. **A costureira bordou o vestido para a filha.**

a. O que aconteceu com a costureira?

b. O que aconteceu com o vestido?

c. O que aconteceu com a filha?

d. Quem bordou o vestido?

e. O que a costureira bordou?

f. Para quem a costureira bordou o vestido?

2. **O ferreiro fabricou uma espada para o soldado.**

a. O que aconteceu com o ferreiro?

b. O que aconteceu com a espada?
c. O que aconteceu com o soldado?
d. Quem fabricou a espada?
e. O que o ferreiro fabricou?
f. Para quem o ferreiro fabricou a espada?

3. A mulher moldou o pote de água para a criança.

a. O que aconteceu com a mulher?
b. O que aconteceu com a criança?
c. O que aconteceu com o pote de água?
d. Quem moldou o pote de água para a criança?
e. O que a mulher moldou para a criança?
f. Para quem a mulher moldou o pote de água?

4. O pai montou o brinquedo para a filha.

a. O que aconteceu com o pai?
b. O que aconteceu com o brinquedo?
c. O que aconteceu com a filha?
d. O que o pai montou para a filha?
e. Quem montou o brinquedo para a filha?
f. Para quem o pai montou o brinquedo?

5. O pedreiro construiu a casa para o vizinho.

a. O que aconteceu com o pedreiro?
b. O que aconteceu com a casa?
c. O que aconteceu com o vizinho?
d. Quem construiu a casa?
e. O que o pedreiro construiu?
f. Para quem o pedreiro construiu a casa?

Classe 14

1. **O sol amadureceu as frutas.**
    a. O que aconteceu com o sol?
    b. O que aconteceu com as frutas?
    c. O que amadureceu as frutas?
    d. O que o sol amadureceu?

2. **O remédio sarou a criança em benefício dos pais.**
    a. O que aconteceu com o remédio?
    b. O que aconteceu com a criança.
    c. O que aconteceu com os pais?
    d. O que sarou a criança?
    e. Quem o remédio sarou?
    f. Em benefício de quem o remédio sarou a criança?

3. **O calor azedou o leite.**
    a. O que aconteceu com o calor?
    b. O que aconteceu com o leite?
    c. O que azedou o leite?
    d. O que o calor azedou?

4. **O tempo envelheceu a avó.**
    a. O que aconteceu com o tempo?
    b. O que aconteceu com a avó?
    c. O que envelheceu a avó?
    d. O tempo envelheceu quem?

5. **O leão morreu no zoológico.**
    a. O que aconteceu com o leão?
    b. O que aconteceu no zoológico?
    c. Quem morreu no zoológico?
    d. O leão morreu onde?

Classe 15

1. **A criança caiu no parque.**
    a. O que aconteceu com a criança?
    b. O que aconteceu no parque?
    c. Quem caiu no parque?
    d. Onde a criança caiu?

2. **A menina chegou em casa tarde.**
    a. O que aconteceu com a menina?
    b. O que aconteceu em casa?
    c. O que aconteceu tarde?
    d. Aonde a menina chegou tarde?
    e. Quem chegou em casa tarde?
    f. Quando a menina chegou em casa?

3. **O estudante saiu da escola cedo.**
    a. O que aconteceu com o estudante?
    b. O que aconteceu na escola?
    c. O que aconteceu cedo?
    d. Quem saiu da escola cedo?
    e. De onde o estudante saiu cedo?
    f. Quando o estudante saiu da escola?

Classe 16

1. **A criança ganhou um brinquedo do pai.**
    a. O que aconteceu com a criança?
    b. O que aconteceu com o brinquedo?
    c. O que aconteceu com o pai?
    d. Quem ganhou um brinquedo do pai?
    e. O que a criança ganhou do pai?
    f. De quem a criança ganhou o brinquedo?

2. **O filho herdou o anel do avô.**
    a. O que aconteceu com o filho?
    b. O que aconteceu com o anel?
    c. O que aconteceu com o avô?
    d. Quem herdou o anel do avô?
    e. O que o filho herdou do avô?
    f. De quem o filho herdou o anel?

3. **O aniversariante recebeu um presente da convidada.**
    a. O que aconteceu com o aniversariante?
    b. O que aconteceu com o presente?
    c. O que aconteceu com a convidada?
    d. O que a aniversariante recebeu da convidada?
    e. Quem recebeu o presente da convidada?
    f. De quem o aniversariante recebeu o presente?

4. **A mãe deu dinheiro para o filho.**
    a. O que aconteceu com a mãe?

b. O que aconteceu com o dinheiro?
c. O que aconteceu com o filho?
d. Quem deu o dinheiro para o filho?
e. O que a mãe deu para o filho?
f. Para quem a mãe deu dinheiro?

Classe 17

1. **A estudante entendeu a questão da prova.**
   a. O que aconteceu com a estudante?
   b. O que aconteceu com a questão da prova?
   c. O que a estudante entendeu?
   d. Quem entendeu a questão da prova?

2. **O aluno aprendeu a matéria.**
   a. O que aconteceu com o aluno?
   b. O que aconteceu com a matéria?
   c. Quem aprendeu a matéria?
   d. O que o aluno aprendeu?

3. **A vítima reconheceu o criminoso.**
   a. O que aconteceu com a vítima?
   b. O que aconteceu com o criminoso?
   c. Quem reconheceu o criminoso?
   d. Quem a vítima reconheceu?

Classe 18

1. **A criança aborreceu a tia.**

a. O que aconteceu com a criança?
b. O que aconteceu com a tia?
c. Quem aborreceu a tia?
d. Quem a criança aborreceu?

2. **A chuva frustrou os planos dos caçadores.**
   a. O que aconteceu com a chuva?
   b. O que aconteceu com os planos dos caçadores?
   c. O que frustrou os planos dos caçadores?
   d. A chuva frustrou o quê?

3. **O filho preocupa a mãe.**
   a. O que aconteceu com o filho?
   b. O que aconteceu com a mãe?
   c. Quem preocupa a mãe?
   d. Quem o filho preocupa?

4. **O balanço sossegou o bebê.**
   a. O que aconteceu com o balanço?
   b. O que aconteceu com o bebê?
   c. O que sossegou o bebê?
   d. O balanço sossegou quem?

Classe 19

1. **O estudante admira o professor.**
   a. O que acontece com o estudante?
   b. O que acontece com o professor?
   c. Quem admira o professor?

d. Quem o estudante admira?

2. **A esposa ama o marido.**
   a. O que acontece com a esposa?
   b. O que acontece com o marido?
   c. Quem ama o marido?
   d. Quem a esposa ama?

3. **Eu quero chocolate quente.**
   a. O que acontece comigo?
   b. Que acontece com o chocolate quente?
   c. Quem quer chocolate quente?
   d. O que eu quero?

4. **O crente teme o inferno.**
   a. O que acontece com o crente?
   b. O que acontece com o inferno?
   c. Quem teme o inferno?
   d. O que o crente teme?

Classe 20

1. **Existe uma casa amarela naquela rua.**
   a. O que acontece com a casa?
   b. O que acontece naquela rua?
   c. A casa existe onde?
   d. O que existe naquela rua?

Referências

CANÇADO, M.; AMARAL, L.; MEIRELLES; L. **Verboweb.** Disponível em: http://www.letras.ufmg.br/padrao_cms/?web=verboweb&lang=1&page=&menu=&tipo=1 Acessado 17/07/2022.

GIVÓN, Talmy. **Topic, pronoun, and grammatical agreement.** In: *Li* 1976:149-88.

LAMBRECHT, Knud. **Information structure and sentence form**: Topic, focus, and the mental representations of discourse referents. Vol. 71. Cambridge University Press, 1996.

ANNEX 2: QUESTIONNAIRE

Questionário – foco em Nyanja

(1) O menino está pulando a cerca na fazenda.
Quem está pulando a cerca na fazenda?
O que o menino está pulando na fazenda?
Onde o menino está pulando a cerca?

O menino está pulando a cerca na fazenda.
N'nyamata akulumpha mpanda kufamu.

Quem está pulando a cerca na fazenda?
Ndani akulumpha m'panda kufamu?

N'nyamata akulumpha mpanda kufamu. (primeira opção)
M’panda walumphidwa ndi n’nyamata kufamu. (segunda opção)

O que o menino está pulando na fazenda?
Ciyani n'nyamata akulumpha kufamu?

N'nyamata akulumpha mpanda kufamu. (primeira opção)
M’panda walumphidwa ndi n’nyamata kufamu. (mais natural)
M’panda walumphidwa kufamu ndi n’nyamata. (terceira opção)

Onde o menino está pulando a cerca?
Kuti n'nyamata akulumpha mpanda?
N'nyamata akulumpha mpanda kufamu.

(2) O menino sempre pula a cerca na fazenda.

Quem sempre pula a cerca na fazenda?

O que o menino sempre pula na fazenda?

Onde o menino sempre pula a cerca?

O menino sempre pula a cerca na fazenda.

N'nyamata nthawi zonse amalumpha mpanda kufamu.

Quem sempre pula a cerca na fazenda?

Ndani nthawi zonse amalumpha mpanda kufamu?

O que o menino sempre pula na fazenda?

Ciyani n'nyamata nthawi zonse amalumpha kufamu?

Onde o menino sempre pula a cerca?

Kuti n'nyamata nthawi zonse amalumpha mpanda?

N'nyamata amalumpha mpanda nthawi zonse kufamu

ou

Mpanda umalumpidwa nthawi zonse ndi n'nyamata kufamu

(para todas as perguntas)

(3) O pai está sonhando com os filhos.

Quem está sonhando com os filhos?

Com quem o pai está sonhando?

(4) O pai sempre sonha com os filhos.
Quem sempre sonha com os filhos?
Com quem o pai sempre sonha?

(5) A criança sempre escova o cachorro.
Quem sempre escova o cachorro?
O que a criança sempre escova?

(6) O marido sempre mói os grãos para a esposa no moinho.
Quem sempre mói os grãos para a esposa no moinho?
O que o marido sempre mói para a esposa no moinho?
Para quem o marido sempre mói os grãos no moinho?
Onde o marido sempre mói os grãos para a esposa?

(7) O marido está moendo os grãos para a esposa no moinho.
Quem está moendo os grãos para a esposa no moinho?
O que o marido está moendo para a esposa no moinho?
Para quem o marido está moendo os grãos no moinho?
Onde o marido está moendo os grãos para a esposa?

(8) João está quebrando o vaso com o martelo.
Quem está quebrando o vaso com o martelo?
O que o João está quebrando com o martelo?
Com o que João está quebrando o vaso?

(9) João sempre quebra o vaso com o martelo.
Quem sempre quebra o vaso com o martelo?

O que o João sempre quebra com o martelo?

Com o que o João sempre quebra o vaso?

(10) O professor está espalhando os livros pela casa.

Quem está espalhando os livros pela casa?

O que o professor está espalhando pela casa?

Onde o professor está espalhando os livros?

O professor está espalhando os livros pela casa.

Mphunzitsi akumwaza mabukhu munyumba.

Quem está espalhando os livros pela casa?

Ndani akumwaza mabukhu munyumba?

O que o professor está espalhando pela casa?

Ciyani mphunzitsi akumwaza munyumba?

Onde o professor está espalhando os livros?

Kuti mphunzitsi akumwaza mabukhu?

Mphunzitsi akumwaza mabukhu munyumba

Ou

Mabukhu akumwazidwa munyumba ndi mphunzitsi

[vale para todas as perguntas]

(11) O professor sempre espalha os livros pela casa.

Quem sempre espalha os livros pela casa?

O que o professor sempre espalha pela casa?

Onde o professor sempre espalha os livros?

O professor está espalhando os livros pela casa.

Mphunzitsi akumwaza mabukhu munyumba.

Quem está espalhando os livros pela casa?

Ndani akumwaza mabukhu munyumba?

O que o professor está espalhando pela casa?

Ciyani mphunzitsi akumwaza munyumba?

Onde o professor está espalhando os livros?

Kuti mphunzitsi akumwaza mabukhu?

Mphunzitsi nthawi zonse amamwaza mabukhu munyumba

Ou

Mabukhu amamwazidwa munyumba nthawi zonse ndi mphunzitsi

[vale para todas as perguntas]

(12) O anfitrião está trancando a porta com a chave.

Quem está trancando a porta com a chave?

O que o anfitrião está trancando com a chave?

Com o que o anfitrião está trancando a porta?

(13) O anfitrião sempre tranca a porta com a chave.
Quem sempre tranca a porta com a chave?
O que o anfitrião sempre tranca com a chave?
Com o que o anfitrião sempre tranca a porta?

(14) A mulher está moldando o pote para a criança.
Quem está moldando o pote para a criança?
O que a mulher está moldando para a criança?
Para quem a mulher está moldando o pote?

(15) A mulher sempre molda o pote para a criança.
Quem sempre molda o pote para a criança?
O que a mulher sempre molda para a criança?
Para quem a mulher sempre molda o pote?

(16) O pedreiro está construindo a casa para o vizinho.
Quem está construindo a casa para o vizinho?
O que o pedreiro está construindo para o vizinho?
Para quem o pedreiro está construindo a casa?

(17) O pedreiro sempre constrói a casa para o vizinho.

Quem sempre constrói a casa para o vizinho?
O que o pedreiro sempre constrói para o vizinho?
Para quem o pedreiro sempre constrói a casa?

(18) Suponha que tem um leão e um tigre num zoológico.
O leão morreu. O tigre não morreu.

O que/quem morreu no zoológico?
(resposta esperada: o leão morreu, e não o tigre)

Onde o leão morreu?
(resposta esperada: o leão morreu no zoológico, e não na savana)

(19) A criança está caindo no parque.
Quem está caindo no parque?
Onde a criança está caindo?

(20) A criança sempre cai no parque.
Quem sempre cai no parque?
Onde a criança sempre cai?

(21) A menina está chegando cedo em casa.
Quem está chegando cedo em casa?

Aonde a menina está chegando cedo?

Quando a menina está chegando em casa?

(22) A menina sempre chega cedo em casa.

Quem sempre chega cedo em casa?

Aonde a menina sempre chega cedo?

Quando a menina sempre chega em casa?

(23) A criança está ganhando um brinquedo do pai.

Quem está ganhando um brinquedo do pai?

O que a criança está ganhando do pai?

De quem a criança está ganhando o brinquedo?

(24) A criança sempre ganha um brinquedo do pai.

Quem sempre ganha um brinquedo do pai?

O que a criança sempre ganha do pai?

De quem a criança sempre ganha um brinquedo?

(25) A mãe está dando dinheiro para o filho no banco.

Quem está dando dinheiro para o filho no banco?

O que a mãe está dando para o filho no banco?

Para quem a mãe está dando dinheiro para o filho no banco?

Onde a mãe está dando dinheiro para o filho?

(26) A mãe sempre dá dinheiro para o filho no banco.

Quem sempre dá dinheiro para o filho no banco?

O que a mãe sempre dá para o filho no banco?
Para quem a mãe sempre dá dinheiro no banco?
Onde a mãe sempre dá dinheiro para o filho?

(27) A vítima está reconhecendo o criminoso na rua.
Quem está reconhecendo o criminoso na rua?
Quem a vítima está reconhecendo na rua?
Onde a vítima está reconhecendo o criminoso?

(28) A vítima sempre reconhece o criminoso na rua.
Quem sempre reconhece o criminoso na rua?
Quem a vítima sempre reconhece na rua?
Onde a vítima sempre reconhece o criminoso?

(29) A chuva está frustrando os planos dos caçadores.
O que está frustrando os planos dos caçadores?
O que a chuva está frustrando?
De quem a chuva está frustrando os planos?

(30) A chuva sempre frustra os planos dos caçadores.
O que sempre frustra os planos dos caçadores?
O que a chuva sempre frustra?
De quem a chuva sempre frustra os planos?

(31) O estudante está admirando o professor.
Quem está admirando o professor?

Quem o estudante está admirando?

(32) O estudante sempre admira o professor.
Quem sempre admira o professor?
Quem o estudante sempre admira?

(33) A esposa sempre ama o marido.
Quem sempre ama o marido?
Quem a esposa sempre ama?

(34) O fazendeiro está apunhalando o porco para a festa.
Quem está apunhalando o porco para a festa?
O que o fazendeiro está apunhalando para a festa?
Para que o fazendeiro está apunhalando o porco?

(35) O fazendeiro sempre apunhala o porco para a festa.
Quem sempre apunhala o porco para a festa?
O que o fazendeiro sempre apunhala para a festa?
Para que o fazendeiro sempre apunhala o porco?