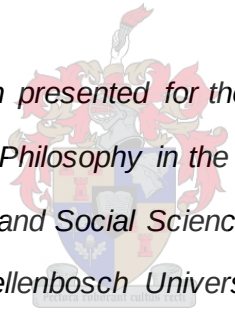
Nominal classification in Bantu revisited: The perspective from Chichewa

by

Peter Kondwani Msaka



Dissertation presented for the degree of

Doctor of Philosophy in the Faculty of

Arts and Social Sciences at

Stellenbosch University

Supervisor

Theresa Biberauer

December 2019

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2019

Signature Date

Copyright © 2019 Stellenbosch University

All rights reserved

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would like to express my sincere gratitude to all individuals and institutions that have assisted me throughout the course of this study. Specifically, I thank my supervisor, Theresa Biberauer for her insights, suggestions and constructive criticism, which have moulded the ideas in this thesis. Her selfless heart offered me a rare opportunity to learn from her expertise. I am also grateful to my former supervisor Johan Oosthuizen, with whom I began the PhD project.

I am grateful for the study funding awarded to me from Partnership for Africa's Next Generation of Academics (PANGeA) through the Graduate School - Faculty of Arts and Social Sciences, Stellenbosch University. I would also like to thank my employer, University of Malawi, for granting me study leave and for giving me financial assistance when my initial funding expired at the end of third year. I am also thankful to the Department of General Linguistics at Stellenbosch for according me the right mix of resources and people. I am equally indebted to colleagues at Chancellor College who took the burden of providing logistical support of various kinds. In a special way, I thank Prof. Edrinnie Elizabeth Lora-Kayambazinthu who has remained the fountain of wisdom and inspiration.

Some aspects of this work were presented to audiences at the *Pre-31st Comparative Germanic Syntax Workshop (CGSW31)*, Stellenbosch University; *Faculty of Humanities Monthly Seminar series*, Chancellor College, University of Malawi; *6th Southern African Microlinguistics Workshop (SAMWOP-6)*, Stellenbosch University and the *Department of Linguistics seminar*, University of Western Cape. I thank these audiences for their attendance, comments, questions, and feedback, which consistently improved my work.

For data access and general technical skills, I am thankful to the University of Malawi's Centre for Language Studies (CLS) for allowing me to participate in the technical section of the *Chichewa Monolingual Dictionary Project*, as a Voluntary Corpus Manager. I am grateful to Daniel Ridings who introduced me to computational lexicography and offered me direct access to the *Chichewa monolingual dictionary database* where I obtained the dataset. I am also grateful to Roland Misomali who honed my skills in using the MySQL database engine. In the same vein, I am also grateful to Steven Paas for his *Chichewa-English/English-*

Stellenbosch University https://scholar.sun.ac.za

Chichewa Dictionary (Paas, 2017),[1] from which I would get many English translations for the appendices. I also thank many well-wishers who assisted with linguistic judgments relating to Chichewa words and constructions.

I say *thank you* to the multitude of my friends: particularly, I would like to acknowledge the life-changing experience that I had with Chikumbutso Manthalu, ("Chiku"- as we would fondly call him) with whom I explored ideas, big and small. I only matured in my disposition because of his always well-thought-out counsel. I also appreciate the support from my inner-circle of friends, drawn from all corners of Africa: Amon Mwiine, Marion Chirwa-Kajombo, Pauline Liru, Maurine Amimo, Itai Makone, Douglas Kiyinikibi, Collins Jana, Dominick Makanjira, Norbert Basweti, Garspudus Mwombeki, Charmaine Manyani, Lucious Kwakwala, and many others.

I would not have remained in school if not for the self-sacrificing hearts of three families: First, that of my senior uncle, Joseph Msaka and his wife Annie Msaka-Chiusiwa who raised me from a very tender age. Second, that of my junior uncle, John Msaka and his wife Molly Msaka-Damba. Third, that of my grandparents, George Msaka and his wife Malita Sangu Msaka (RIP). I am grateful for their parental support, undaunted by resource scarcity. The true African *umunthu* spirit resides in them.

Finally, I owe a huge debt of gratitude to my family, especially my wife Tiwonge, my son Shepherd Withu and my daughter Nthanda. My absence from home impacted them in many agonising ways. They deserve a lion's share of all the successes of this study.

[1] See also chichewadictionary.org

Stellenbosch University https://scholar.sun.ac.za

Abstract

The present study focuses on the phenomenon of noun class and agreement systems in Chichewa, a Bantu language spoken in some parts of Eastern and Southern Africa. Noun classification is a strong areal feature of Bantu languages, however it remains one of the less well understood phenomena in this family (see i.a. Maho, 1999, chap. 6; Mchombo, 2004: 3; Katamba, 2006: 120). The generally accepted views about this phenomenon centre on the noun-class framework associated with Bleek (1862, 1869) and Meinhof (1899, 1906, 1932) – henceforth referred to as the *Bleek-Meinhof system*. The major problem with this system lies in the nature of the assumptions that are made regarding the principles underlying the classification system. In terms of the Bleek-Meinhof system, it is assumed that modern-day Bantu noun-class systems can all be understood as being based on a set of reconstructed noun prefixes. If we consider these modern-day systems, however, it emerges that these noun prefixes are not found on every noun and also not in every noun class. This is a distribution pattern that has been argued to have arisen due to language change, which has distorted the original regular morphological and semantic basis of the classes.

The present study is centrally concerned with Chichewa, a language whose present-day noun class system poses numerous challenges to the traditional Bleek-Meinhof system. I argue that this latter system can in fact be shown to have been constructed on the basis of a partial dataset. Drawing on a carefully sampled dataset from the *Chichewa monolingual dictionary*, I outline how many nominal classifications can be shown to be problematic if we adopt the Bleek-Meinhof system. Clearly setting out these facts, constitutes the thesis' descriptive objective. With the description in place, the second goal is to propose a novel noun-class system that is plausible from a child language acquisition perspective. I propose an agreement-based approach which classifies all expressions in the dataset according to the agreement markers that they control on adjectives, numerals, verbs, etc. Having taken this approach, I demonstrate that we can identify twelve agreement classes (ACs) in Chichewa, with each AC including a wide range of expressions, some of which are non-nominal in nature. In this regard, I propose two subsystems within the agreement classification system, one an *agreement-based noun-class* (ANC) subsystem and the other a *general agreement class* (GAC) subsystem. On the one hand, I will argue that in Chichewa, the ANC system has a primarily semantic basis, which gives rise to two super-

Stellenbosch University https://scholar.sun.ac.za

noun classes, one animate/agentive and the other non-agentive/inanimate. The inanimate/non-agentive is further subcategorised on a phonological basis, taking into account the word-initial element, into six other classes. The GAC, on the other hand, is triggered by on at first sight rather miscellaneous group of expressions, some of which have not traditionally been associated with nominal expressions such as complementiser phrases (CPs), prepositional phrases (PPs), and also underspecified agreement triggers. I also argue that the proposed underlying principles of the Chichewa ANC and GAC systems are at the core of various grammatical structures and language acquisition processes observed crosslinguistically.

Stellenbosch University https://scholar.sun.ac.za

Opsomming

Hierdie studie fokus op die naamwoordelike klas- (*noun class*) en grammatikale kongruensie (g*rammatical agreement*)-sisteem van Chichewa, 'n Bantoetaal wat in dele van oos- en suider-Afrika gepraat word. Naamwoordelike klassifikasie is 'n verskynsel wat sterk onder die Bantoetale figureer; nietemin bly dit een van die Bantoekenmerke wat linguiste vandag nog minder goed verstaan (sien o.a. Maho, 1999, hoofstuk 6; Mchombo, 2004: 3; Katamba, 2006: 120). Die perspektief op Bantoe naamwoordelike klassifikasie wat die algemeenste aanvaar word, sentreer op die raamwerk van Bleek (1862, 1869) en Meinhof (1899, 1906, 1932), waarna ons hierna sal verwys as die *Bleek-Meinhofsisteem.* Die hoofprobleem met hierdie sisteem lê in die aannames wat gemaak moet word aangaande die beginsels waarop die klassifikasiesisteem rus. In die konteks van die Bleek-Meinhofraamwerk word daar aangeneem dat hedendaagse Bantoe naamwoordelike klassifikasiestelsels almal gebaseer is op 'n stel gerekonstrueerde (*reconstructed*) naamwoordelike prefikse. Wanneer ons die hedendaagse stelsels bekyk, blyk dit egter dat daar naamwoorde is waar hierdie naamwoordelike prefikse ontbreek, en ook dat die verwagte prefikse nie by elke lid van 'n gegewe naamwoordelik klas gesien word nie. Hierdie distribusionele patroon word aan taalveranderig toegeskryf: die oorspronklike reëlmatige morfologiese en semantiese basis van die betrokke naamwoordelike klasse sou oor tyd vervorm het.

Hierdie studie se sentrale fokus is Chichewa, 'n taal wie se huidige naamwoordelike klassifikasiestelsel talle uitdagings skep vir die tradisionele Bleek-Meinhofsisteem. Ek dui aan hoe hierdie wydaanvaarde sisteem in feite op grond van 'n gedeeltelike datastel opgestel is. Met behulp van 'n datastel wat sorgvuldig op grond van die *Chichewa monolingual dictionary* saamgestel is, demonstreer ek hoeveel aspekte van die bestaande klassifikasie in die geval van Chichewa problematies is. Om die relevante feite duidelik uiteen te sit verteenwoordig die tesis se deskriptiewe oogmerk. Met dié beskrywing in plek is die tesis se tweede doelwit om 'n hersiene naamwoordelike klassifikasiesisteem uit te lê wat veral ook vanuit 'n taalverwerwingsperspektief geloofwaardig sou kon wees. Ek stel 'n kongruensie (*agreement*)-gebaseerde benadering voor in terme waarvan al die naamwoordelike items in die datastel op grond van die kongruensiemerkers wat hulle op byvoeglike naamwoorde, telwoorde, werkwoorde, ens. bepaal, geklassifiseer word. As hierdie benadering gevolg word, kan daar vir Chichewa twaalf kongruensieklasse (*agreement classes = ACs*) onderskei word. Elke AC bevat 'n wye verskeidenheid items,

Stellenbosch University https://scholar.sun.ac.za

wat onder andere ook nie-naamwoordelike items insluit. Ek onderskei twee sub-sisteme binne die AC-klassifikasie: enersyds ‘n kongruensie-gebaseerde naamwoordelike klas (*agreement-based noun-class* = *ANC*) sub-sisteem, en andersyds ‘n *algemene kongruensie klas* (*general agreement class* = *GAC*) sub-sisteem. In hierdie verband wys ek dat die ANC in Chichewa ‘n hoofsaaklik semantiese basis het, wat aanleiding gee tot twee super-naamwoordklasse, die ene lewend (*animate*)/agentief en die ander nie-lewend/nie-agentief. Die nie-lewende/nie-agentiewe klas word dan op fonologiese basis in ses verdere klasse onderverdeel, met die kenmerke van die woordinisiële element wat telkens deurslaggewend is. Die GAC, daarenteen, word deur ‘n groep uitdrukkings bepaal wat met die eerste oogopslag nogal uiteenlopend is, met van die bepalers wat nie tradisioneel as naamwoordelike uitdrukkings beskou word nie, onder meer komplementeerderfrases (CPs), preposisionele frases (PPs), en ondergespesifiseerde kongruensiebepalers. Om af te sluit, toon ek dat die beginsels waarop die voorgestelde Chichewa ANC- en GAC-stelsels gebaseer is ook ‘n kernrol speel by ‘n verskeidenheid grammatikale strukture en taalverwerwingsprosesse wat in onverwante tale aangetref word.

Stellenbosch University https://scholar.sun.ac.za

Chidule cha kafukufuku uno

Kafukufuku uno akuyang'ana za magulu a maina ndi agwirizanitsi awo mChichewa, chiyankhulo cha chiBantu chimene chimayankhulidwa mmaiko angapo a ku mmawa ndi ku mwera kwa Afrika. Magulu a maina ndi chizindikiro chachikulu cha ziyankhulo za chiBantu, komabe ndi gawo limodzi la ziyankhulo zimenezi limene silinamvetsetseke (Onani mwachitsanzo Maho, 1999, mutu 6; Mchombo, 2004: 3; Katamba, 2006: 120). Magulu a mainawa amagawidwa kutsatira ndondomeko imene anafotokoza aBleek (1862; 1869) ndi aMeinhof (1899; 1906; 1932) – mwachidule, *ndondomeko ya Bleek-Meinhof*. Vuto lalikulu pa nkhaniyi ndi yokhudza malamulo amene amapanga magulu a mainawa. Motsatira ndondomeko ya Bleek-Meinhof, ambiri amakhulupilira kuti magulu a mainawa amatsatira magulu aphatikirammbuyo amene amaganiziridwa kuti anali mwadongosolo mmake wa ziyankhulo za Chibantu. Vuto ndi lakuti, aphatikirammbuyowa sapezeka pa dzina lililonse kapena mmagulu onse a maina. Kusokonekeraku amanena kuti kunadza kamba ka kusintha kwa ziyankhulozi, kumene kunachititsa kusintha kwa aphatikirammbuyowa ndi matanthauzo awo.

Kafukufuku uno akuyang'ana kwambiri pa Chichewa, chiyankhulo chimene magulu ake amaina sangafotokozedwe momveka bwino ndi ndondomeko ya Bleek-Meinhof. Ndionetsa kuti ndondomeko ya Bleek-Meinhof inapangidwa ndi umboni osakwanira bwino. Pogwiritsa ntchito umboni osanjidwa bwino omwe ukuchokera mu *Mtanthauzira Mawu wa Chichewa*, ndionetsa kuti magulu a maina ambiri sangafotokozedwe momveka bwino ngati tingatsatire maganizo a aBleek ndi aMeinhof. Kotero, cholinga choyamba cha kafukufuku uno ndi kutulutsa umboni oyenerera kuchokera mu *Mtanthauzira Mawu wa Chichewa*. Ndikatulutsa umboniwu, cholinga chachiwiri ndi kupereka ndondomeko ya tsopano ya magulu a maina ndi mawu ena ofuna agwirizanitsi mChichewa. Ndionetsa magulu atsopanowa pogwiritsa ntchito agwirizanitsi amene dzina lililonse limafuna pa afotokozi, afotokozi owerenga, aneni, ndi mawu ena. Pogwiritsa ntchito njira imeneyi, ndionetsa kuti mChichewa muli magulu a maina okwana khumi ndi awiri. Kuonjezera apo, ndionetsanso kuti gulu lililonse lili ndi maina osiyanasiyana, kuphatikizapo mawu ena amene simaina. Kamba ka izi, ndidzagawa magulu a mawu ofuna agwirizanitsiwa pawiri, gulu loyamba likhala la *magulu a maina opezeka potsatira agwirizanitsi* (MOPA) ndi gulu lachiwiri gulu la *magulu amawu potsatira agwirizanitsi* (MAPA). Ndionetsanso kuti gulu la MOPA limatsatira matanthauzo a maina zimene zitipatsa magulu ena awiri, lina la maina onena za ochita/zamoyo ndi lina lotenga

Stellenbosch University https://scholar.sun.ac.za

maina otanthauza ochitiridwa/zopanda-moyo. Gulu lotenga maina a ochitiridwa/zopanda-moyo likugawidwanso pasanu potsatira kamvekedwe ka liwu loyambirira la dzinalo. Pamene gulu la MAPA, limatenga magulu a mawu opangidwa podzera mnjira zophatikiza mawu monga alozi, ziganizo zosaima pazokha komanso mziganizo zina zimene agwirizanitsi alibe mawu ogwirizana nawo. Pomaliza ndionetsa kuti malamulo omanga magulu a maina mChichewa amapezekanso mmalamulo a ziyankhulo zina zosiyanasiyana komanso mndondomeko zimene ana amaonetsa kuti amatsata pophunzira ziyankhulo.

Stellenbosch University https://scholar.sun.ac.za

Table of Contents

Frontsmaterial

Chapter 1: Introduction

Chapter 2: The conceptualization of the Bantu noun class system

Stellenbosch University https://scholar.sun.ac.za

Chapter 3: Data collection methods and approaches to noun classification

Chapter 4: Empirical evidence for rethinking the chichewa noun class system

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Table of Tables

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 1
Introduction

1.1 Focus of the study

The present study focuses on the phenomenon of noun class and agreement systems in Chichewa, a Bantu language spoken in parts of Eastern and Southern Africa. Noun classification is regarded as one of the strongest areal distinctive features of Bantu languages (Heine, 1982: 190; Mchombo, 2004: 2; Katamba, 2006: 103). However, it is widely acknowledged that the phenomenon of Bantu noun classification is not yet properly understood (see i.a. Worsley, 1954; Amidu, 1997; Maho, 1999; Mchombo, 2004: 3; Katamba, 2006: 119–120; Braver & Bennett, 2015; Byamugisha, Keet & de Renzi, 2018: 2641). The existing Bantu-wide noun class framework, originally proposed by Bleek (1862, 1869), fails to account for a wide range of morpho-phonological and semantic patterns, which linguists have accordingly been forced to view as a relic of an older, underlying system that has become less regular due to language change (see i.a. Givón, 1971a, 1972; Maho, 1999; Demuth & Weschler, 2012: 70). However, in the present study, I contend that a schema which is characterised on the basis of diachronic considerations and which is not independently characterizable in terms of being accessible to the language acquirer is not a pattern that can credibly be postulated as the 'system' underlying native-speaker knowledge (cf. Amidu, 1997: 3, 15, 66). Acquirers do not have direct access to the history of their language (Lightfoot, 1979, 1991, 1999); the only historical facts that can shape naturally acquired grammars are those encoded via synchronic cues (see i.a. Lightfoot, 1979, 1991, 1999; Yang, 2016; Biberauer, 2017, 2018a). Moreover, on the basis of detailed evidence from Chichewa, I argue that the major problem with the traditional Bantu noun class framework emanates from the type of data that informed the original analyses. With specific reference to a properly sampled nominal dataset from *Chichewa monolingual dictionary*, I will show that the traditional Bantu noun class framework is in fact based on a partial set of data (see also Worsley, 1954: 286). Accordingly, one of the objectives of this study is to rectify this bias and address the empirical gap that undermines the current noun classification system. A second aim is to propose a novel account of the facts.

Chichewa belongs to the Niger-Kordofanian language family, specifically to the Benue-Congo phylum. On Guthrie's (1967, 1971) classification system of Bantu languages,

Stellenbosch University https://scholar.sun.ac.za

Chichewa is classified as belonging to Zone N, Unit N31. According to Mchombo (2004: 1, 2006: 142, 2007: 204), Chichewa is a variety of Nyanja, a language that is spoken in Malawi, Mozambique (in the Tete and Niassa provinces), Zambia (in the Eastern Province) and Zimbabwe. In Malawi, Chichewa is a lingua franca throughout the country (Kayambazinthu, 1998: 371; Baldauf Jr & Kaplan, 2004; Msaka, 2015: 1–2). The variety under investigation in the present study is that which is, in Malawi, generally considered to be "standard Chichewa", being associated with formal domains such as education, publishing, news media, etc.

Although the Bantu noun class system in general has been the focus of much scholarly attention, detailed analyses focusing on Chichewa noun classes are limited. The most detailed studies in this regard are Watkins (1937); Harding (1966); Corbett & Mtenje (1987) and Matiki (2001). Many other studies only make broad references to the Chichewa noun class system, Bresnan & Mchombo (1995) and Mchombo (2004) being two much-cited examples.

1.2 The research problem and research objectives

1.2.1 The phenomenon

Noun classification/gender is a grammatical feature of some languages in terms of which nouns are grouped according to a set of criteria. The classification is always morphosyntactically reflected on other words that are in construction with the noun, such as adjectives, verbs, etc. in a mode that is usually known as *grammatical agreement* (Hockett, 1958: 231; Steele, 1978: 610; Corbett, 1991: 1; Dimitriadis, 1997: 1; Alcock & Ngorosho, 2004: 4; Bateman & Polinsky, 2010: 41). By way of illustration, consider the simple Russian examples in (1).

(1) Gender and grammatical agreement in Russian (Corbett, 2005: 126, 2013)

(a) Žurnal ležal na stole.
magazine lay.M on table
"The magazine lay on the table."

(b) Kniga ležal-a na stole.
book lay.F on table
"The book lay on the table."

(c) Pis´mo ležal-o na stole.
letter lay-N on table
"The letter lay on the table."

Stellenbosch University https://scholar.sun.ac.za

In (1), the three nouns; *žurnal*, *kniga* and *pis´mo* are taken to belong to three different genders/classes, namely masculine, feminine and neuter, respectively. In turn, this classification is reflected in the morphological forms of the verb by means of null or overt suffixes. The three different forms of the verb, *ležal*, *ležal-a* and *ležal-o* are therefore due to the type of noun they are in construction with (Corbett, 2005: 126, 2013). In many languages, the systematic covariance illustrated in (1) between the noun and the verb spreads to many other word categories, such as adjectives, determiners, numerals, etc. (cf. Corbett, 1991: 105, 2006: 4; Pollard & Sag, 1994: 60; Preminger, 2014: 5). The relationship between noun classes/genders and the agreement on associated words is the hallmark of a noun classification system.

Noun classification is a widespread phenomenon in natural languages, estimated to occur in about 40% of the world's languages (Corbett, 2005: 127, 2013).[1] According to Maho's (1999: 2, 53) sample of 333 Bantu languages, with the exception of a language called Komo, the rest are observed to employ the noun class system (cf. Guthrie, 1971: 42). Although the noun class system is one of the major characteristics of Bantu languages, the system is, crucially, not observed to be homogenous across the various languages: there are significant variations in terms of the number of noun classes and the nature of the agreement systems (Maho, 1999: 127–142; Katamba, 2006: 108–109).

Crosslinguistically, the phenomenon of nominal classification is also manifested in various ways and has been known by a variety of names, such as *gender, noun class, classifier* and *categorisation systems*. In some cases, these terms are used almost interchangeably (cf. Corbett, 1991: 10; Bresnan & Mchombo, 1995: 183; Maho, 1999: 142–146; Garbo, 2013: 116; Kilarski, 2013, chap. 2). However, a clear distinction is usually drawn between classifier systems and gender/noun class systems. Classifier systems are taken to be a different phenomenon because (i) they are less grammaticalized compared to the gender/noun class systems and (ii) the classification is not obligatory as some nouns may remain unclassified (Contini-Morava & Kilarski, 2013: 266).[2] Although the terms *gender* and *noun class* can refer

[1] The percentage is based on a sample of 257 languages where 112 were found to present gender/noun class systems as reported in the *World Atlas of Language Structure* (Corbett, 2005: 127, 2013). Kilarski (2013: 11) also states that gender/noun class systems are more typical of fusional and agglutinating languages.

[2] Examples of classifier systems are found in diverse families of languages such as the Malayo-Polynesian, the Austro-Asiatic, the Sino-Tibetan, the Altaic, the Dravidian, and the Indo-Aryan (Senft, 2007: 680).

Stellenbosch University https://scholar.sun.ac.za

to the same phenomenon, they are also usually associated with different linguistic traditions and language typologies (Corbett, 1991: 10; Maho, 1999: 142–146; Kilarski, 2013: 1). In the present study, I will use the term *noun class* except where it is necessary to highlight a linguistic tradition-related distinction. Next, I introduce the Bantu noun class problem.

1.2.2 The Bantu noun class problem

Bantu languages are known to exhibit one of the most complex noun class (NC) systems, with the NCs ranging from 2 to over 23 and a pervasive agreement system that affects words associated with the noun within and outside the noun phrase (NP) (Watkins, 1937; Corbett & Mtenje, 1987; Maho, 1999; Mchombo, 2004; Carstens, 2008). In trying to understand the underlying criterion for the various NCs, many studies appeal to the proposals largely associated with Bleek (1862, 1869) and Meinhof (1899, 1906; 1932), which have come to be known as the *Bleek-Meinhof system* (see i.a. Maho, 1999: 3–4; Katamba, 2006: 105). In terms of this system, nouns are classified primarily based on morphological markers on the noun themselves, popularly known as *noun class prefixes* (*NCP)* (see i.a. Mufwene, 1980; Mugane, 1997: 33; Maho, 1999; Katamba, 2006: 105). There are claims that the Bleek-Meinhof system is semantically motivated (see i.a. Burton & Kirk, 1976; Batibo, 1987; Orr, 1987; Spitulnik, 1989; Kgukutli, 1994; Selvik, 1996, 2001; Contini-Morava, 1997; Demuth, 2000; Palmer & Woodman, 2000). Although the number of classes varies from one language to another, the classes are identified by a standardized Bantu-wide numbering system that ranges from 1 to 23 (Maho, 1999: 50–54; Katamba, 2006: 104). These prefixes were largely obtained through linguistic reconstruction as they are assumed to have existed in Proto-Bantu (PB), the hypothesised parent Bantu language (Maho, 1999, chap. 5; Katamba, 2006: 5). The Bleek-Meinhof NC system is represented in Table 1.1, with Chichewa giving an indication of the extent to which this modern Bantu language approximates the putative proto-system:

Stellenbosch University https://scholar.sun.ac.za

Table 1.1: The set of noun classes, NCPs and their associated semantic features in Proto-Bantu

NC	Proto Bantu NCP	Chichewa NCP	Assumed semantics of the NCs
1	*mù-	*m(u)-*	humans
1a	*ø-	*ø-*	kinship terms, personified animals
2	*va-	*a-*	honorific, plural to NC1
2a	*va-	*a-*	honorific, plural to NC1a
3	*mu-	*m(u)-*	trees, plants, inanimates
4	*mi-	*mi-*	plural NC3
5	*li-	**li-*	miscellaneous, paired things, augmentatives
6	*ma-	*ma-*	liquids, collectives, plural to NCs 5,9,11,14 and 15
7	*ki-	*chi-*	inanimates, manner/style, diminutives, augmentatives
8	*bi-	*zi-*	plural to NC7
9	*n-	**N*	animals
10	*n-	**N*	plural to NCs 9 and 11
11	*du-		long thin things, abstracts
12	*ka-	*ka-*	diminutives
13	*tu-	*ti-*	plural to NC12
14	*bu-	*u-*	abstract nouns, mass nouns, plural to NC12
15	*ku-	*ku-*	infinitives
16	*pa-	*pa-*	locatives, ‘near’ or ‘explicit’
17	*ku-	*ku-*	locatives, ‘remote’ or ‘general’
18	*mu-	*m(u)-*	locatives, ‘inside’
19	*pi-		diminutives
20	*ɤo		augmentatives, diminutives
21	*ɤi-		augmentatives, pejoratives
22	*ɤa-		plural to NC20
23	*ɤe-/*i-		locative, unspecified

Table 1.1 is based on Maho (1999: 51), Mchombo (2004: 6) and Katamba (2006: 109).

As shown in Table 1.1, each NC is associated with one prefix, with the asterisk signifying linguistic reconstruction. As stated above, these NCs are also assumed to be based on semantic patterns which are no longer consistent today, but are believed to have held in PB (see i.a. Richardson, 1967: 378; Maho, 1999: 63; Katamba, 2006: 116).

Stellenbosch University https://scholar.sun.ac.za

There are at least two serious challenges that arise from the characterisation of the Bantu NC system represented in Table 1.1. First, the NCP is not a consistent criterion in many classes. For example, the prefixes illustrated in Table 1.1 are not found on every noun in the respective NCs in Chichewa (cf. Matiki, 2001: 66–67). For instance, NC1 and NC5 alone have nouns with various morphological characteristics such as the following: (i) nouns without any prefix, (ii) nouns derived by suffixes only, (iii) nouns derived by conversion and (iv) nouns derived by other prefixes from the ideal represented in Table 1.1. Consider the nouns in Table 1.2 (see Chapters 3 and 4 for more examples and discussion).

Table 1.2: Nouns with various morphological characteristics in NC1 and NC5

	NC1	**NC5**
a.	(i) **m**-sodzi PREFIX-N.STEM “fisherman”	(ii) **li**-tali PREFIX-ADJ.STEM “length”
b.	(i) ø-galu NULL-dog “dog”	(ii) ø-gulu NULL-group “group”
c.	(i) gonth-**i** from gonth-a V.STEM-SUFFIX V.STEM-SUFFIX “deaf person” “become deaf”	(ii) funs-**o** from funs-a V.STEM-SUFFIX V.STEM-SUFFIX “question” “ask”
d.	(i) nyénga from nyenga “deceiver/cheater” “cheat/deceive”	(ii) lémba from lemba “script/letter” “write”
e.	(i) **na**-ku-bala PREFIX-INFINITIVE-give birth “parent”	(ii) **d**-ombol-**o** from ombolol-a PREFIX-V.STEM V.STEM-SUFFIX “rescue/ransom” “rescue”
f.	(i) **ka**-soze PREFIX-look out “spy, intelligence agent, type of baboon”	
g.	(i) **ma**-nthu PREFIX-N.STEM “mother ant”	

As shown by the data in (b-g) in Table 1.2, there are more noun-types in NC1 and NC5 than the *m*- and *li*-prefixed nouns on which the Bleek-Meinhof schema is based, respectively. In Chapters 3 and 4, I will show that many of these noun types and derivation strategies are systematic and statistically significant.

The second challenge relates to the assumed semantic attributes, which also do not form explicable patterns if one operates with this noun class schema (see also Chapter 2, Section 2.4.3). For example, although NC1 is analysed as the human class, it also contains several

Stellenbosch University https://scholar.sun.ac.za

non-human denoting nouns, including some prefix *m*-derived nouns. Consider examples in (2) (see also Chapter 4, Section 4.3).

(2) a. m-neni
PREFIX-say
"verb"

b. m-lumikizi
PREFIX-join
"conjunction"

c. ka-dzadza
PREFIX-be full
"funnel"

d. ka-gwinthi
PREFIX-stunted
"small unripe pumpkin"

e. na-m-tibule
PREFIX-PREFIX-break
"pounded green maize"

f. naliti Afrikaans loan word (Afr.LW)
needle
"needle"

Take note that the counter-examples in Table 1.2 and those in (2) do not independently make much sense without referring to a statistically representative sample, a problem that Worsley (1954: 286) describes as "casual exampling". In other words, the question that needs to be considered is to what extent the various noun-types ascribed to a given class conform to the expectations of the traditional Bleek-Meinhoff-based criteria, and to what extent they contain "exceptions" to these criteria? I deal with this problem in Chapters 3 and 4.

The just cited inconsistencies with the Bleek-Meinhof schema and many others are largely assumed to have arisen due to language change (see i.a. Doke, 1954; Cole, 1967; Givón, 1971a, 1972; Okhotina, 1975; Mchombo, 1978: 117; Maho, 1999, chap. 5). In this regard, the Bantu noun class system as represented in the Bleek-Meinhof system is taken to be a relic of the old regular system of PB. Strikingly, though, the inconsistencies are such that some studies have described the noun classification in Bantu as having "fallen apart", and being "random" and "in inexplicable chaos" (see i.a. Richardson, 1967: 378; Givón, 1971a: 33–34; Herbert, 1991: 105; Kishindo, 1998: 45; Maho, 1999: 67).

The problem with these traditional views about the Bantu noun classes is that they do not *explain* the synchronic grammar of the noun classification system (cf. Amidu, 1997: 3, 15, 66). Crucially, particularly from a generative perspective, the question is how does a child acquirer of Chichewa postulate the noun class system today? Amazingly, all Chichewa speakers seem to converge on the same noun class system. Studies focusing on the acquisition of noun classes in other Bantu languages show that these systems are typically in place by around the age of 3 (see Demuth, 2000; Demuth & Weschler, 2012: 71). If we

Stellenbosch University https://scholar.sun.ac.za

assume that the system has undergone changes, the question is what the *patterns* of that changed system are that enable a child acquirer's parser to work out the system? According to what is currently known about language acquisition, a highly irregular distribution of linguistic evidence cannot be the basis of successful language acquisition (see i.a. Crain, 1991; Smith & Tsimpli, 1995; Tettamanti, Alkadhi, Moro, Perani, Kollias, *et al.*, 2002; Singleton & Newport, 2004; Hudson-Kam & Newport, 2005; Yang, 2016; Biberauer, 2017, 2018a; Yang, Crain, Berwick, Chomsky & Bolhuis, 2017).

What appears to be the case with the Bleek-Meinhof system is that diachronically informed linguists, but not child acquirers, have access to the evidence that is assumed to hold the key to the grammatical structure of the Chichewa noun class system. More generally, though, the picture is usually the inverse: as Gagliardi & Lidz (2014: 58) observe, "[w]hile all children acquire the structure of their native language in a mere five (or so) years, with little apparent effort or confusion, language scientists fare considerably worse in identifying that structure" (cf. Yang, 2016: 4).[3] In this regard, it is clear that genuine insight into how the child successfully converges on the grammar of their native-language can only be obtained if linguists seek to operate with the data that acquirers can reasonably be assumed to have at their disposal. Children learning a particular language do not have access to all the information-types that are potentially at the disposal of linguists, such as grammaticality judgments, diachronic evidence, corpus data, etc. (Yang, 2016: 4). Given the foregoing, the question that arises is therefore as follows: what is the primary evidence that children work with in the course of acquiring the Chichewa noun class system?

As introduced above, in all typical noun class systems, there are two types of information that cue the noun class system of a particular language, namely *noun-external distributional information* and *noun-internal distributional information* (Gagliardi, 2012: 14; Gagliardi & Lidz, 2014: 59). Noun-external distributional information relates to evidence reflected in the agreement system whereas noun-internal distributional information relates to the shared semantic or morpho-phonological properties of the nouns that make up a particular noun class (Gagliardi & Lidz, 2014: 59). Crucially, the former can be taken to be highly reliable as agreement markers (AMs) are very regular while the latter is taken to be probabilistic as it

[3] However, Gagliardi (2012: 16) notes that in some cases linguists may easily identify the noun class system while children may struggle with the patterns.

Stellenbosch University https://scholar.sun.ac.za

varies significantly compared to the noun-external distribution information. Gagliardi (2012: 17) argues that these two types of information are obviously available to the child acquirer (cf. Gagliardi & Lidz, 2014: 59). These two types of information have also been postulated to be the basis of many noun class systems and have led to different characterisations of noun classes. For example, classes based on noun-internal and noun-external information have been variously identified as 'Controller genders' and 'Target genders' (Corbett, 1991: 150–160) or 'Morphological noun class' and 'Syntactic noun class' (Bokamba, 1993: 210–218) or 'Head classes' and 'agreement classes' (Evans, 1997: 106), respectively (cf. Corbett, 1991: 7–8; Dahl & Fraurud, 1996: 54; Aikhenvald, 2004: 1033).

As shown in Table 1.1, the Bleek-Meinhof system is predominantly based on noun-internal information as there are some distinct noun classes sharing the same AM (cf. Corbett & Mtenje, 1987: 8). However, what the evidence emerging in the present study suggests, in line with what Corbett & Mtenje (1987) have previously proposed for Chichewa, is that it would be more insightful to derive the noun class system primarily on the basis of noun-external distributional information (cf. Gagliardi, 2012: 27) (see Chapters 3 & 4). In summary therefore, two positions are taken in the present work: first, I take a child language acquisition perspective and second, following Corbett & Mtenje (1987), I organise the noun classes primarily on the basis of agreement information before considering the distribution of noun-internal characteristics.

As clearly shown by the Chichewa dataset to be introduced in Chapter 3, the problems with the Bantu noun class framework relate to the schema itself and not necessarily to either the language data or to its historical trajectory, as is typically claimed in the existing literature. The schema is observed to only account for a fraction of Chichewa nouns, which shows that it is not based on a properly sampled dataset. Similar observations have also been made by Worsley (1954: 286), who argues that the "casual exampling" approach common among Bantu noun class studies is responsible for obscuring the real nature of the system. Worsley (1954: 286) proposes that a proper quantitative analysis of Bantu noun classes would reveal semantic correlations in animate nouns and to a lesser degree among inanimates, as well as phonological regularities in the assignment of loanwords (cf. Greenberg, 1954; Kilarski, 2013: 166–167). Corbett (1991: 59, 63, 315) also insightfully argues that the randomness that linguists associate with various noun class systems is deceptive because phonological and morphological regularities are found in systems that appear relatively opaque (cf. Alcock

Stellenbosch University https://scholar.sun.ac.za

& Ngorosho, 2004: 4; Braver & Bennett, 2015). Both the methodological concerns and the observations regarding the phonological patterns in the noun class system are found to be relevant in the way I analyse the Chichewa noun class system (this will be illustrated in detail from Chapter 3 onwards).

Given the background presented in this section, in the present research, I intend to contribute to the existing discussion by seeking to answer the following questions.

(3) a. Are the Bleek-Meinhof system and the views around it compatible with existing language acquisition models?
b. Why is the Bantu noun class system so elusive to linguists but not to the child acquirer?
c. What are the simple cues available to the child learner in the process of acquiring the Chichewa noun class system?
d. What kind of data would reveal the cues in (3c) above?
e. What is the relationship between the observed semantic and morpho-phonological patterns in the Chichewa noun class system?

Given the research questions in (3), the aims of the present study are two-fold: (i) to present a carefully sampled dataset and (ii) to propose a novel account of the empirical facts, one that is compatible with theories of language acquisition. To achieve these aims, I will be guided by the following specific objectives.

(4) a. To present a systematically sampled set of data from the *Chichewa monolingual dictionary*.
b. To re-examine the relationship between the noun-external information (AMs) and the noun-internal information (NCP, semantic and phonological).
c. To account for the relationship between semantic and word-initial phonological features within the noun class system.
d. To propose a synchronically plausible account of the Chichewa agreement-based classification system.
e. To account for the underlying principles of the proposed Chichewa agreement-based classification system.

In the next section, I introduce the theoretical assumptions that will inform the tasks outlined in (4).

1.3 Theoretical assumptions

The noun classification phenomenon draws on many aspects of linguistic structure, including lexical, phonological, morphological, syntactic and semantic aspects (see i.a.

Stellenbosch University https://scholar.sun.ac.za

Levelt, 1989; Corbett, 1991; Kilarski, 2013). Noun classification is also observed to strongly involve language-external aspects, such as social-cultural categorisation (Fortune, 1970: 95–96; Dixon, 1972: 306–311; Corbett, 1991: 33; Kilarski, 2013: 213–214). Accordingly, there are many facets to the research objectives stated (4) that require proper empirical and theoretical grounding. Two aspects are key, however. First, any noun class framework must be amenable to explication via plausible mechanisms and processes of child language acquisition, i.e. we need to understand how the noun class system is acquired. Second, the noun class system is morpho-syntactically manifested in language structure, and this structure must also be explicable in terms of more generally applicable theories of language structure. To satisfy these two key benchmarks, I appeal to a specific set of well-defined linguistic assumptions. I introduce these theoretical assumptions next.

This research will be cast within the generative approach to linguistic analysis, especially the current Minimalist research programme outlined in Chomsky (1995) and subsequent works. Like the generative approach more generally, the Minimalist Program is, in broad terms, concerned with generating empirically and theoretically informed hypotheses regarding the knowledge of language, its structure and how it is acquired (Chomsky 1986). Within the generative tradition, this human language capacity is understood to be embodied in the Language Faculty (Chomsky & Lasnik, 1993; Chomsky, 1995, 2005, 2017a; Hauser, Chomsky & Fitch, 2002). This language faculty can be viewed as comprising two components, namely 'the faculty of language in the narrow sense' (FLN) and 'the faculty of language in the broad sense' (FLB) (Hauser *et al.*, 2002: 1570). On the one hand, the FLN only includes the core grammatical computations which generate grammatical structures of language, currently understood to be the generative computational operation of *Merge* (Chomsky, 2015: 233, 2017b: 297ff), and, for most researchers (*pace* Hornstein, 2008), *Agree*.[4] The FLB, on the other hand, includes the FLN itself and other organism-internal components, namely the sensory-motor and conceptual-intentional subsystems. While the sensory-motor and conceptual-intentional systems are observed to be available in other species, the FLN is assumed to be specifically endowed in humans (Hauser *et al.*, 2002:

[4] (External) Merge and Move (internal merge) are core cognitive structure-building operations that are thought to underlie the recursive hierarchical organisation of human language and thought (Chomsky, 2017b: 297ff, a: 202ff; Hornstein, 2018: 55ff). Specifically, Merge combines two constituents to derive another complex constituent, whereas Move is a syntactic operation that results in a constituent being displaced (moved) from one position to another in the same structure, once again to generate a "larger" structure than that which was present pre-Move (Chomsky, 2008, 2015: 223ff).

Stellenbosch University https://scholar.sun.ac.za

1571). This 'innate disposition' in humans has been assumed to be invariant in all human languages, hence designated *Universal Grammar* (UG) since Chomsky (1965) (Hauser *et al.*, 2002: 1577; Chomsky, 2017a: 200). There are also organism-external factors such as the social-cultural, ecological and physical environment that may have a bearing on language, although they are not the preoccupation of generative linguistics (Hauser *et al.*, 2002: 1570). These theoretical views form the backbone of the specific assumptions about language acquisition and the structure of nominal expressions that I introduce next.

1.3.1 Assumptions about acquisition

To evaluate whether a particular account of the noun class system is plausibly acquirable, I assume the minimalist *Three Factors Model* (Chomsky, 2005; Biberauer, 2017, 2018a). Specifically, the study adopts the version outlined in Biberauer (2017, 2018a), which holds that language acquisition is the product of three interacting factors, namely, Factor 1: a highly constrained UG; Factor 2: Primary Linguistic Data (PLD) – the data that language-acquiring child picks up on in the course of language acquisition, sometimes called the *intake* (see below) and Factor 3: the non-language-specific, general cognitive principles, a core representative of which is argued to be an acquisition bias designated *Maximise Minimal Means* (MMM). This acquisition model is schematised in (5).

(5) UG + PLD + MMM -> steady-state grammar (Biberauer, 2018a: 99).

Each of the three factors in (5) is assumed to play a crucial role in the process of language acquisition. According to Biberauer (2018a: 100), UG has the two key functions of leading the acquirer to 'expect' that the input will have (i) recursive structure which is the output of binary Merge and (ii) a structure which 'reflect[s] the UG-given formal feature template' ([F]). The claim, then, is that UG predisposes the language acquirer to execute a "specific type of formal analysis of the input" (Biberauer, 2018a: 100). It is partly due to this innate ability that acquirers are able to postulate appropriate rules for their languages (Gagliardi, 2012: 7; Biberauer, 2018a: 100).

The specific language rules are generally assumed to be postulated based on the nature of Factor 2, i.e. the PLD. Therefore, the role of the PLD is largely to provide systematic linguistic content which the acquirer can parse. In this regard, the ambient language data is taken to act as 'input' to the child. However, it is not all 'input' that is relevant in language acquisition, as has been argued that there are differences between 'input' and 'intake' (Gagliardi, 2012;

Stellenbosch University https://scholar.sun.ac.za

Lidz & Gagliardi, 2015; Fodor & Sakas, 2017; Biberauer, 2018a: 98). Therefore, from the acquirer's perspective, the role of PLD should be understood in terms of the 'intake' and not necessarily the entire linguistic input. This, of course, means that the linguist needs to think carefully about the aspects of the input that are likely to qualify as 'intake' from the child's perspective (see i.a. Gagliardi, 2012).

To answer the question of how the child approaches the PLD to postulate a grammar, Biberauer (2017, 2018a) proposes a range of both overt and covert structures from which the acquirer's parser may draw generalizations. Biberauer's proposal rests on a core linguistic assumption that predates generativism, namely Saussure's view, which can be traced back to late 19th century, that speakers memorise the 'arbitrary pairing between form and meaning' (Saussure, 1966). Following Chomsky (1995), we can take this arbitrary pairing to involve form - in some sense, phonological - or [P]-features, and meaning and, thus, semantic or [S]-features. Second, again following Chomsky (1995), the [P]- and [S]-features are taken to also interact with a different set of what are known as formal or [F]-features. It is, therefore, further assumed that part of language learning involves acquirers noticing and memorising these arbitrary pairing of [P]-, [S]- and [F]- features. Specifically, Biberauer (2017:43) proposes that there are degrees of arbitrariness in human language as characterised in (6).

(6) a. lexically stored, idiosyncratic conventionalized sound-meaning mappings involving [P]- and [S]-features, and
b. grammatically regulated and thus more systematic conventionalised sound-meaning mappings, involving [P]-, [S]- and/or [F]- features (all of the feature-types need not be simultaneously present, as (7) shows).

The major difference between (6a) and (6b) is that, in the former, acquirers simply lexicalise the [P] and [S] pairing, while in (6b), acquirers notice the pairing and postulate rules which they generalise to all relevant cases. The one-to-one pairing in (6a), Biberauer (2017: 42) designates *Saussurean arbitrariness*; the more systematic, but yet still arbitrary pairings in (6b), she terms *systematic departures from Saussurean arbitrariness*, with the designation being intended to highlight the basis on which acquirers will differentiate these aspects of the intake. (7) gives an indication of the forms that these 'systematic departures from Saussurean arbitrariness' may take (see Biberauer, 2017, 2018a, b for more detailed discussion):

Stellenbosch University https://scholar.sun.ac.za

(7) a. Cases where there are two/multiple forms that convey only one meaning, known as *doubling/agreement,* or one form with no meaning, e.g. expletive elements. In these cases, the regularity determining the realisation of the agreement/doubling and the expletive element is captured by the postulation of appropriate [F]-features.

b. Cases where there is meaning without form, what is denoted as *systematic silence*: e.g. null exponence, null arguments, null complementisers, ellipsis, etc. Again, the regularity underlying the distribution of the null elements is captured by the postulation of appropriate [F]-features.

c. Cases where one form contributes multiple meanings, depending on placement and distribution, known as *multifunctionality*. In such cases the acquirer postulates underspecified 'homophones', which are devoid of [F]-features, alongside phonologically null functional heads, which bear the [F]-features and, as merge-sites for the underspecified forms, determine their distribution.

d. Movement, specifically where such movement conveys 'extra' meaning, for example, in topicalization, wh-movement, etc. In such cases, the acquirer postulates the relevant [F]- features that regulate the observed word-order patterns (see also (9) below).

According to Biberauer (2017: 43–45), the patterns shown in (7) represent *systematic departures from Saussurean arbitrariness*. The implication is that [F]s are postulated if they are observed to control some systematic contrast, which cannot be accounted for by appealing to Saussurean arbitrariness only. In this regard, [F]-features must be postulated where one-to-one pairing patterns break down.

What is key for the present study is that phonological and semantic features are present anyway (as the basis of the linguistic signs that 'words' are), and that the *Saussurean arbitrariness* assumption leads us to expect that the anyway-required [P]- and [S]-features can serve as the basis for formal (i.e. grammatical) features. Crucially, from this input the acquirer identifies the various [F]s that are grammaticalized, shaping the grammar of the language in question (Biberauer, 2018a: 103). Finding an aspect of grammar where BOTH phonological and semantic features more and less transparently feed into the structuring of grammar, as is the case for Chichewa noun class system that I propose in the present study,

Stellenbosch University https://scholar.sun.ac.za

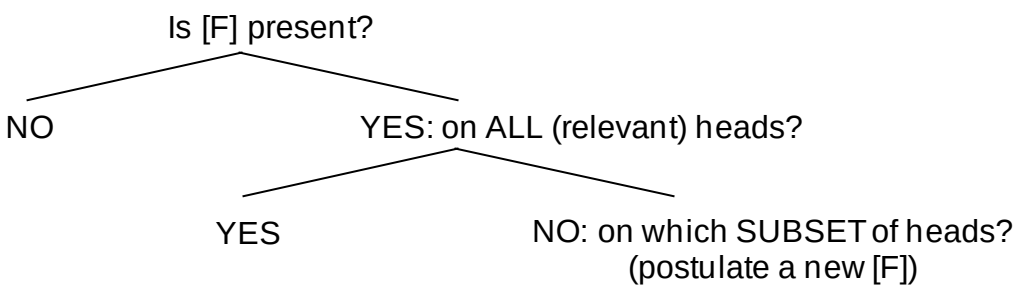
is therefore very much what one would expect for **early-acquired/basic** grammatical properties (see Tsimpli, 2014 for overview discussison and see also Chapters 5 and 7).

The third factor that Biberauer (2018: 103) identifies is MMM, which is taken to be a characterisation of one of the "general cognitive constraints regulating the way humans construct, and interpret the world around them." These are also attested in other systems, e.g. human writing systems. Crucially, MMM, when it applies in the domain of language, appears to give rise to the following properties.

(8) Linguistic manifestations of MMM (Biberauer, 2018a: 103)
   a. Feature Economy (FE): Postulate as few features as possible to account for the systematic regularities in the input.
   b. Input Generalisation (IG): Maximize the use that is made of the features postulated.

The properties in (8) entail that the acquirer parsing the input will home in on systematic regularities and "will postulate grammars that recycle [F]s as much as possible" (Biberauer, 2018a: 103). By way of illustration, this language acquisition process can be assumed to take the path schematised in (9), where [F] is a particular formal feature:

(9)



In (9), the PLD either gives the acquirer a reason to postulate [F] or not. If, for example, the input contains a systematic contrast between overtly realised and null subjects or between wh-words which move to Spec-CP as opposed to their non-interrogative counterparts that stay in situ, this needs to be captured in some way by an [F]. Crucially, [F] will only be postulated if there is a systematic regularity in the input that triggers the postulation of [F]. The first *NO* above thus represents - for the *linguist's* convenience - the case where there is no triggering input, i.e. it is not the consequence of an acquirer having considered whether [F] plays a role in their grammar or not. In *wh in situ* languages, for example, acquirers will have no evidence of *wh*-movement and so they will not postulate the presence of this [F] (often designated [wh] in the literature) on the basis of the discrepancy between wh-marked and non-wh-marked elements (although they may have other cues in the intake indicating

Stellenbosch University https://scholar.sun.ac.za

the need to postulate this [F], e.g. a shared morpheme, agreement-triggering properties, or similar). The non-postulation of [F] is therefore the default; triggering input is required for [F] to be postulated.

If there is positive evidence supporting the postulation of [F] - as in the first *YES* above - MMM is assumed to guide the acquirer to make as much use of this [F] as possible. Unless there is evidence to the contrary, each new [F] will thus apply to the whole class of elements with which it was initially associated (all pronouns are initially assumed to be omissible where an acquirer has picked up on systematic null-argument realisation, all *wh*-elements are assumed to bear the [wh]-feature and an appropriate movement trigger where systematic *wh*-movement has been detected, etc.). If the child receives input which constitute exceptions to the [F]-based generalisation being entertained at a given stage, a further [F] will be postulated to delimit the class of elements exhibiting the observed behaviour in a suitable way, with this process being repeated until the formal properties of the target grammar in the relevant domain are correctly characterised (Biberauer, 2018a: 104). According to this model, the language acquisition process is thus partly enabled by the innate predisposition of the acquirer's 'mind', general cognitive abilities and constraints, and by the nature of what is accessible in the PLD.

The foregoing theoretical assumptions are relevant to the present study as follows: First, I will only assume patterns that are in the PLD to serve as the basis for the rules that acquirers can postulate. Second, I will assume that the acquisition of language only makes use of 'plausible cognitive resources and psychological mechanisms' (Yang, 2016: 41; Biberauer, 2018a: 103). In other words, the way I interpret the noun class data and the proposals I make about them must thus be plausible in terms of these constraints. Significantly, as is the case in the generative frameworks, no system or subsystem of the noun class can be primarily understood in terms of diachronic evidence as acquirers do not have direct access to such information. The most important new considerations that the Three Factors model introduces are a suggestion as to what kind of input might be the most significant, and the expectation that both phonological and semantic properties could play a role in helping the acquirer to identify categories.

Two further questions arise at this point, though: (i) How much evidence is required to postulate a regular rule? (ii) How does the acquirer deal with exceptions? To answer these questions, I further adopt Yang's (2016: 171) Tolerance Principle, a rule-based evaluation

Stellenbosch University https://scholar.sun.ac.za

metric that can be used as a theoretical device for (i) guiding "linguists to choose among competing analyses" and (ii) enabling "children to select the correct grammar on the basis of the [PLD]". As shown in the preceding discussion, child acquirers notice linguistic patterns that they postulate as rules. It is also widely accepted that children make generalisations of the postulated rules as demonstrated in their ability to go beyond input structures (see i.a. Chomsky, 1980; Culicover, 1999; Bauer, 2001; Yang, 2016). However, considering that there are often exceptions to rules (see i.a. Chomsky & Halle, 1968: 172), generalisations require sufficient evidence to be deemed productive (Yang 2016:171). In this regard, the inferred generalisations must be statistically computed to determine productivity or lack thereof (Yang 2016:171). This productivity can thus be evaluated in terms of the Tolerance Principle, stated as follows:

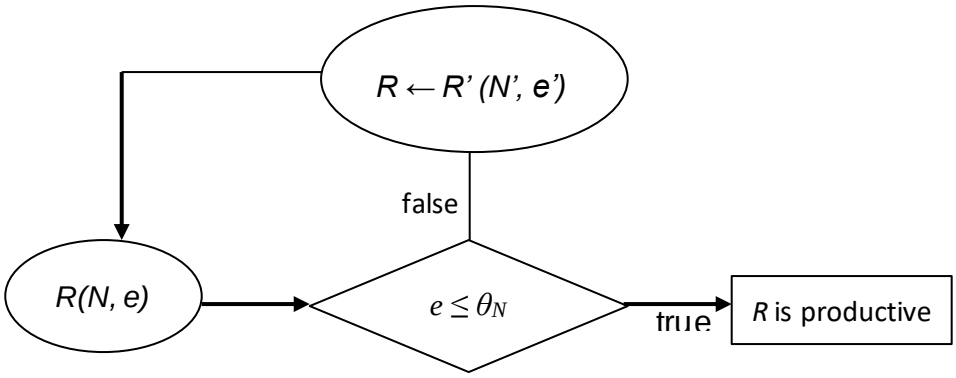
(10) If R is a productive rule applicable to N candidates, then the following relation holds between N and *e*, the number of exceptions that could but do not follow R:

$e \leq \theta_N$ where $\theta_N := \frac{N}{lnN}$ (Yang, 2016: 8–9), where *ln* stands for natural logarithm.

According to Yang (2016: 8–10), the metric in (10) simply means that the rule is productive if the exceptions (*e*) are 'sufficiently' low. The procedure is envisaged as follows: first, a rule R, which captures a generalisation in the input, is postulated based on available evidence in the PLD. The productivity of rule R is then evaluated according to the associated numerical values: N and *e*. If the number of exceptions is not 'below a critical threshold', the child learner postulates a revised rule R' hence obtains a new set of numerical values (N' and e'). The Tolerance Principle is then applied recursively as illustrated in (11) (Yang, 2016: 9).

Stellenbosch University https://scholar.sun.ac.za

(11) The Tolerance Principle as an evaluative metric in language acquisition (Yang, 2016: 10)



When the number of exceptions is below the threshold value, the rule becomes productive. Table 1.2 below shows sample number of cases (N) and their approximate threshold percentages. For justification of the tolerance threshold formula, see Yang (2016, chap. 3).

Table 1.3 The tolerance threshold for rules of varying sizes (Yang, 2016: 67).

N	θ_N	%
10	4	40.0
20	7	35.0
50	13	26.0
100	23	23.0
200	38	19.0
500	80	16.0
1,000	145	14.5
5,000	587	11.7

As shown in Table 1.2, given a small set of 10 cases associated with a particular rule R, 4 is the computed critical threshold value (θ_N). Therefore, if the number of exceptions is less than 40%, the rule is productive; however if, the number of exceptions is above 40% then the rule cannot be productive. Similarly, given a large set of 5,000 cases, the computed critical threshold is 587, implying that this scenario will only accept exceptions not exceeding 11.7% for the rule to be productive. An important property of the Tolerance Principle is that it is easier to postulate a productive rule with a smaller set of data than with a larger set; (Yang, 2016: 67). The implication for a real language learning situation is that a child with a limited set of evidence is likely to postulate rules easier than one with an adult-like set of evidence (Yang, 2016: 67).

Stellenbosch University https://scholar.sun.ac.za

Another important aspect of the Tolerance Principle regards its account of exceptions/irregular forms. According to Yang (2016:38), acquirers are also observed to make a categorical distinction between productive and unproductive processes, which implies that learners are aware of both productive and non-productive forms.[5] The implication is that the exceptions and idiosyncratic forms are part of the grammar, but they are only learned by rote memorisation (cf. Chomsky & Halle 1967:172, Yang 2016:41, 216). This explains why the exceptions to the rules are always finite in number while the items following productive rules are infinite. In this regard, Yang (2016:60) claims that exceptions and rules are necessary as they "optimise/minimize the time complexity of language use". More important though is the fact that the Tolerance Principle guides learners to avoid resorting to excessive memorisation and to prefer productive rules of grammar (Yang 2016:106). In effect, therefore, the assumed language learning strategy guides acquirers to follow rules that maximize productivity (Yang, 2016: 72), as also independently argued by Biberauer (2017, 2018a).

While it is common practice in linguistics for theorists to draw generalisations from data without following a particular statistical model (Yang, 2016: 75), I conjecture that if children draw generalisations based on statistical weight of evidence, it would be equally illuminating to draw statistically computed generalisations from the corpus data that I use in the present study. Therefore, in addition to using the Tolerance Principle as a rule-based tool to understand how acquirers draw generalisations from available data, I also employ it in drawing generalisations from the Chichewa corpus. In this regard, the Tolerance Principle will help to determine whether a particular distribution in my corpus supports the productivity of a particular rule or not.

1.3.2 Assumptions about noun structure

While the foregoing assumptions will guide the way we understand the properties of an acquirable noun class system, they do not help to explain the formal structure of the nouns themselves. To address this specific aspect of the phenomenon, I adopt the (Minimalist) approach which has become known as *Distributed Morphology* or *DM* (Halle & Marantz, 1993; Embick & Noyer, 2007). Significantly, there is already a DM analysis of nominal

[5] The phenomenon of regular and irregular rules has been a long-studied puzzle in linguistics (see Yang, 2016, chap. 1 and references cited there)

Stellenbosch University https://scholar.sun.ac.za

expressions in a Bantu language called Luganda (Ferrari, 2005). I therefore follow Ferrari's (2005) instantiation of DM assumptions.[6] In DM terms, word structure and phrase structure are generated by the same rules of syntax (Ferrari, 2005: 14; Embick & Noyer, 2007: 290), which entails that there is no distinction between what is traditionally known as morphology and syntax. In this regard, words are complex elements obtained from syntactic derivations that merge *roots* and *abstract morphemes* (Embick & Noyer, 2007: 295).

Roots are small unanalysable meaningful units that are devoid of grammatical features, which are taken to be sequences of complex phonological features producing phonotactically well-formed strings such as $\sqrt{\text{CAT}}$, $\sqrt{\text{OX}}$ and $\sqrt{\text{SIT}}$ (Embick & Noyer, 2007: 295). Ferrari (2005: 15) defines roots as "lexical forms endowed with meaning, but lacking functional content." It is also assumed that roots do not, in the syntax, appear 'bare'; they must be categorised by one of the categorial functional heads (c), namely noun (*n*), verb (*v*), and adjective (*a*) (Embick & Noyer2007:295, Marantz 1995). Following Ferrari's (2005: 15) assumptions, the association of a categorial feature and a root derives what is designated a *stem*, as schematised in (12).

(12) $\sqrt{root}$ + (c) = stem(c)

Although the notion of stem is not common in mainstream DM, the stem is a key notion in Bantu morpho-syntax (see i.a. Kishindo, 1985: 3; Carstens, 1991: 15; Mugane, 1997: 50–51; Mchombo, 2004: 22; Downing & Mtenje, 2017: 14). According to the schema in (12), stems are always specified for categorial features (Ferrari, 2005: 15). Therefore, a stem is a complex structure minimally made up of a root and a category defining functional head.

The abstract morphemes, on the other hand, are taken to lack phonological properties and to express such functions as Tense, Aspect, Mood, Number, Person, etc. (Ferrari, 2005: 19; Embick & Noyer, 2007: 295). Ferrari (2005: 16) argues that inflectional features only attach to stems. Following the analysis being introduced here, a Bantu plural noun can be schematised as in (13), with the morpheme order not being relevant at this point.

(13) PL + *n*+ $\sqrt{root}$ = plural noun

[6] Although I adopt Ferrari's (2005) instantiation of the DM framework in analysing Bantu noun structure, from Chapter 3 onwards, I depart from her specific analysis of the Bantu noun class system.

Stellenbosch University https://scholar.sun.ac.za

Functional categories are then assumed to be supplied with phonological features by the mechanism known as "Vocabulary Insertion" (Embick & Noyer, 2007: 297). The Vocabulary is understood to be a list of the phonological exponents which are paired with conditions on insertion onto the featurally specified functional projection. Each phonological exponent that has been paired with grammatical context is known as a 'vocabulary item'. For example, in English the regular plural is expressed by the vocabulary item /-z/, as illustrated in (14) (Embick & Noyer, 2007: 298).

(14) Z ↔ [PL]

In (14) there is a one-to-one pairing of a set of phonological features and plural functional features in English. The Vocabulary Insertion assumption is also crucial in accounting for patterns where one form is used to perform two distinct functions – cf. the discussion of multifunctionality in (7c) above. For example, as I will show in Chapter 6, it is commonly attested in many languages (including Bantu languages) that augmentative (AUG) and diminutive (DIM) affixes are also used to mark speaker-hearer perspectives such as pejoration, endearing, etc. (cf. Embick & Noyer, 2007: 299–300; Déchaine, Girard, Mudzingwa & Wiltschko, 2014: 26). In this regard, one AUG affix may give literal, denotative meanings of size but also convey connotative pejorative meanings. Systematic syncretism of this type is assumed to be due to the fact that phonological exponents are underspecified in terms of the sites where they are to be inserted (Embick & Noyer, 2007: 299–300). By means of illustration, consider the way Hupa, an Athabascan language, marks Person/Number on object (Obj) and subject (Subj) nouns (Embick & Noyer, 2007: 300).

(16) a. [+1prs+PL + Subj] ↔ dɨ
b. [+2prs+PL+Subj] ↔ oh
c. [+PL+Obj] ↔ noh

Take note that while the first and second person Subj plural markers are expressed by two different forms in (16a,b), the first and second person Obj plural marker is expressed by only one phonological form in (16c). However, the notion of 1st person and 2nd person plural is still conveyed in this language albeit by one phonological form. Therefore, the phonological form *noh* is underspecified as it can serve two different functions; put differently, it can be inserted into two different functional contexts.

Thus far, what we have in place, then, is that the basic ingredients for deriving words are (i) the generative capacity, (ii) the functional elements and (iii) the language-specific roots.

Stellenbosch University https://scholar.sun.ac.za

However, Ferrari (2005: 12–13) observes that there are some limitations to the DM view assumed here. For example, she (2005: 12–13) states that all languages also contain a significant number of unproductive lexicalised forms (cf. Yang, 2016; Biberauer, 2017: 43), which cannot be assumed to be systematically derived by the syntactic processes espoused by DM. Ferrari (2005: 13) argues that the non-productive word forms can only be assumed to have been acquired as irregular forms and represented by rote memorization. As discussed above, there are always irregular and idiosyncratic forms in various linguistic structures (see i.a. Fanselow & Féry, 2002; Baerman, Corbett & Brown, 2010; Yang, 2016). Therefore, like Ferrari, I do not assume that all nouns are derived in a predictable manner.

1.3.3 Interim summary

To summarise this section, I assume the generative approach to language analysis. Three specific theoretical assumptions are adopted, each focused on specific concerns of the study. First, the Three Factors Model (Chomsky, 2005; Biberauer, 2017, 2018a) is adopted for evaluating the noun class system to be proposed so that child language acquisition is suitably taken into account. Second, the Tolerance Principle is employed as an analytical tool for assessing rule productivity in relation to the set of data considered. Third, DM assumptions serve as analytical tools for analysing the morphosyntactic structure of the data. I will highlight each of these theoretical assumptions in the relevant sections of the thesis.

In the next section, I provide a brief outline of each chapter of the thesis.

1.4 The structure of the thesis

The rest of the thesis is structured as follows. In Chapter 2, I provide a general overview of how Bantu noun classes have been conceptualized. I begin by outlining some basic facts of the Bantu noun class system and then show how ideas regarding the Bantu noun class have developed. Then, I present the various competing analyses of the noun class system before outlining the perennial challenges and the issues that have remained unresolved to date.

In Chapter 3, I introduce the empirical evidence that this study is based on. I begin the chapter by discussing some methodological concerns for collecting data of the type to be considered in the present study. Then, I introduce the corpus dataset excerpted from the *Chichewa monolingual dictionary*. The various nouns are then categorised according to the

Stellenbosch University https://scholar.sun.ac.za

AM that they control, which gives rise to 12 agreement classes. I also introduce the various morphosyntactic properties of the nouns that form the dataset.

In Chapter 4, I discuss properties of the expressions that make up each agreement class introduced in Chapter 3. I show that each agreement class includes expressions of diverse morpho-phonological and semantic characteristics. However, despite the diverse features, each agreement class contains expressions that share sets of semantic and/or phonological features. In this chapter, I also draw a distinction between two types of agreement classes: one containing less predictable nominal expressions which I designate the *agreement-based noun class* (ANC) subsystem and the other containing highly predictable derived expressions, which include both nominal and non-nominal expressions, such as evaluative NPs, locatives, non-locative PPs, CPs, and underspecified agreement triggers, which I term the *general agreement class* (GAC) subsystem. Therefore, Chapters 3 and 4 achieve a descriptive typology that is prerequisite for the more analytical discussion in the following chapters, specifically chapters 5, 6 and 7.

In Chapter 5, I propose a new ANC system for Chichewa based on the noun-internal information of the agreement classes introduced in Chapter 4. I show that primarily, the nouns are classified on a semantic basis which gives rise to two super-ANCs, one animate/agentive and the other non- inanimate/agentive. The inanimate/non-agentive super-ANC is observed to further classify its nouns on the basis of word-initial phonological characteristics such as [±coronal], [±back], [±round], [±labial] and [±anterior]. I also show that the features underlying the proposed Chichewa noun class system are hierarchically organised and that the proposed system is also attested in other languages.

In Chapter 6, I discuss properties of the highly predictable derived expressions such as diminutives, augmentatives, honorifics, locatives, CPs, etc. I show that evaluative nominal expressions are structurally different from the less predictable nouns that were considered in Chapter 5. I also argue that locative expressions in Chichewa do not belong to the nominal category but are simply locative PPs and that their ability to trigger agreement takes the same form as other non-nominal expressions such as CPs and non-locative PPs. I argue that the AMs triggered by these expressions do not encode grammatical functions of subject and object but serve discourse and pragmatic functions of Topic and Focus. In the chapter, I also introduce the second aspect of the agreement system which is not considered in the Bleek-Meinhof system, namely that involving the dummy functions of some AMs. All in all, I

Stellenbosch University https://scholar.sun.ac.za

argue that agreement in Chichewa is not only triggered by nominal expressions but also by non-nominal elements such as CPs, PPs and underspecified or null elements, justifying the need to broaden our view of the phenomenon from the narrow *noun class system* to more general *agreement class system*, where the ANC is only a subsystem.

In Chapter 7, I provide an analytical discussion of the underlying principles of the Chichewa ANC system. I argue that the Chichewa ANC system is organised along three key principles, namely (i) the semantic primacy principle, (ii) the phonological primacy principle and (iii) the constituent-edge principle. I further show that all three of the principles are manifested in many of the world's languages and that they are pervasive in a wide range of grammatical structures. The principles are also found to be core to language acquisition process crosslinguistically. I argue that the three principles conspire to make the Chichewa noun and agreement classes a well organised system that is plausibly acquirable.

Finally, I summarise and conclude the thesis in Chapter 8. Following the summary, I highlight some aspects of the study that require further investigation.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 2
The conceptualization of the Bantu noun class system

2.1 Introduction

In the present Chapter, I outline the facts and interpretations that have influenced the conceptualization of the Bantu noun class system in the literature to date. I will show how analyses of the Bantu noun class system have developed and highlight the most significant views on the subject. The literature on the Bantu noun class system is abundant; therefore, it is not possible to conduct a complete review of the various works. I will therefore only attempt to pay specific attention to the core recurrent issues in linguistic inquiry on the phenomenon. I will broadly do the following (i) present the empirical facts that have informed the analyses of Bantu noun classes generally, (ii) outline the major competing accounts of the relevant facts and (iii) highlight the major issues that have remained elusive in the analyses of the phenomenon.

The rest of the Chapter is organised as follows, in section 2.2, I present a simplified illustration of the facts that have led to the standard analysis of Bantu noun classes. This is followed by section 2.3, where I present a brief history of scholarly attention to the phenomenon. I will also introduce what has come to be the standard analysis of the Bantu noun class system. In Section 2.4, I outline the various perspectives concerning the principles that encode the Bantu noun class system. This is followed by section 2.5, where I present an overview of two non-standard analyses of Bantu languages, one proposed by Amidu (1997) and the other by myself (2003). In Section 2.6, I highlight the major outstanding issues that demand further attention. Finally, in section 2.7, I summarise and conclude the present Chapter.

2.2 Some basic facts about noun classification in Bantu

The traditional understanding of the phenomenon at hand is largely based on the following simplified empirical facts and observations. As stated in Chapter 1, my own choice of examples here is thus highly biased so as to illustrate the traditional views on the Bantu noun class system. Later, in Chapter 3 and 4 I demonstrate that many previous studies have to a large extent been misled by this practice of selective exemplification, notably also in Chichewa, where I have access to a much more complete dataset.

Stellenbosch University https://scholar.sun.ac.za

As introduced in Chapter 1, the noun classification system is manifested in two ways, namely via noun-internal and noun-external grammatical features. First, as regards the former, noun class features are observed to be manifested on nouns themselves in the form of noun class prefixes (NCPs) (see i.a. Greenberg, 1977; Corbett & Mtenje, 1987; Maho, 1999: 1–2; Mchombo, 2004; Katamba, 2006). Consider the following examples.

(1)		**Singular**	**Plural**	
	a.	mu-nthu	a-nthu	
		NCP-being	NCP-being[1]	
		"person"	"people"	
	b.	m-konz-i	a-konz-i	
		NCP–make-NFV	NCP–make-NFV	
		"maker/editor/sculptor"	"makers/editors/sculptors"	
	c.	m-khirisitu	a-khirisitu	English loan word (Eng.LW)
		NCP-Christ	NCP-Christ	
		"Christian"	"Christians"	

As shown by the glosses in (1), nouns are a combination of the NCP and the stem. For example, in Chichewa the singular noun *munthu* is taken to be formed by combining the NCP *mu-* with the monomorphemic noun stem *-nthu*, whereas the plural form is derived by combining the NCP *a-* with the same stem (*-nthu*) forming *anthu*. In some cases, especially with verb stems, the verb's final vowel (FV), *a-* is replaced by a *nominalizing final vowel* (NFV), *i-* as shown in (1b) (see i.a. Kishindo, 1985: 3–6; Mchombo, 2004: 113–114). Sometimes, words of foreign origin may also undergo similar derivational processes, as shown in (1c).

In a similar manner, nouns considered to belong to another class are assumed to have a different NCP, as shown in (2).

(2)		**Singular**	**Plural**
	a.	li-tchowa	ma-li-tchowa
		NCP-hair on a tail of an animal	NCP-NCP-hair on a tail of an animal
		"fly-whisk"	"fly-whisks"

[1] Due to the theoretical implications glossing conventions have, the glossing in the present chapter may vary from one source to another and from one point in the text to another. However, the most preferred glossing for this study will be introduced in Chapter 3.

Stellenbosch University https://scholar.sun.ac.za

b.	li-tali	ma-li-tali
	NCP-long	NCP-NCP-long
	“length”	“lengths”

In (2), the singular noun *litchowa* is made up of the NCP *li-* and the nominal stem *-tchowa*. The plural form is a combination of the plural prefix *ma-* and the singular form, *litchowa* deriving *malitchowa* as shown in (2a). This is then observed to be the regular pattern as also illustrated by example (2b). Take note that plural marking in this class is additive, that is, the plural form is added onto the form that derived the singular noun, which is different from the case in (1) where plural marking is substitutive (cf. Maho, 1999: 58). This indicates that these prefixes have different morphological status.

Second, as regards noun-external features, there is a system of obligatory agreement markers (AM) that appear on words associated with the noun, both within and outside the noun phrase (Watkins, 1937; Corbett & Mtenje, 1987; Mchombo, 2004). For example, the noun *litchowa* (“flywhisk”) triggers agreement form *li-* on both the demonstrative and the verb as shown in (3) whereas the plural form *malitchowa* controls AM *a-* on both targets in (4).

(3)	**Li**tchowa	**li**-ja	**li**-na-gwa.
	fly-whisk	AM-dem	AM-T/A-fall
	“That fly-whisk fell down.”		

(4)	Malitchowa	**a-**ja	**a**-na-gwa.
	fly-whisk	AM-dem	AM-T/A-fall
	“Those fly-whisks fell down.”		

As shown in (3), in some cases, there is identity (concordance) in morphological form between the NCP and the AM on the associated words, a phenomenon that is known as *alliterative concordial agreement* (Corbett & Mtenje, 1987: 4; Corbett, 1991: 117; Katamba, 2006: 111; Kilarski, 2013: 17; Déchaine *et al.*, 2014: 39). Despite the fact that not all nouns trigger alliterative concordial agreement (Katamba, 2006: 111), Bantu linguists have generally identified it as a major feature of Bantu languages (see i.a. Corbett & Mtenje, 1987: 4; Corbett, 1991: 117; Kilarski, 2013: 17; Déchaine *et al.*, 2014: 39).

As was stated in Chapter 1, the grammatical relationship displayed by the NCPs and the AMs is a strong piece of evidence that the nouns are actually in classes of some kind (Steele, 1978: 610; Corbett, 1991: 105ff; Bokamba, 1993: 220). This is irrespective of the fact that the NCP and the AM may not be identical in form; the systematic covariance suffices to support such a conclusion. Since the plural and singular nouns bear different NCPs and

Stellenbosch University https://scholar.sun.ac.za

trigger different types of AMs on the associated words, they have also been identified as independent classes. However, some studies have taken the singular/plural pairs as belonging to the same broader class, often referred to as *genders* (Watkins, 1937; Hyman, 1980: 11; Corbett & Mtenje, 1987; Carstens, 1991, 2008; Katamba, 2006: 103). In this case, examples (4 & 5) can be identified as forming one gender. However, as I will show from Chapter, the singular-plural pairing is not always straightforward since some plural nouns triggered the same AM as singular nouns from other classes. In addition to that I show in Chapter 5 that plural-singular features do not play a significant role in structuring the Chichewa noun class system.

In the next section, I present a brief overview of how these observations have been formalised by earlier studies in Bantu linguistics.

2.3 The emergence of the Bantu noun classification schema

The illustrated observations, in Section 2.2, form the basis of a long tradition of Bantu nominal classification studies, which spans a period of about four centuries. The first study is that of Brusciotto around 1659 who in his grammar of Kikongo, classified nouns on the basis of their concords, specifically those appearing in possessive constructions (Doke, 1935: 116; Maho, 1999: 13; Katamba, 2006: 103–104).[2] Nevertheless, most studies point to Bleek's (1862) *A comparative grammar of South African Languages* as the most notable landmark work in the study of Bantu noun class systems (see i.a. Maho, 1999: 14; Katamba, 2006: 104). Bleek reconstructed what he assumed to be the ancient Bantu noun class system and devised a numbering schema that identifies each class by means of Arabic numerals. These classes were identified on the basis of the so-called NCP.[3] The framework as it is today has undergone a number of revisions, starting with Bleek's (1869) own second

[2] The Catholic priest, Fr. Hyacinth Brusciotto de Vetralla is reported to have been the first to write a grammar of a Bantu language. According to Doke (1935: 87), this "work was the first to record that hall-mark of Bantu, the noun-class system." Brusciotto is reported to have coined the term 'principiations' to indicate the categories into which the nouns fell, according to their first element. Of special relevance to the present study is the fact that "his division into principiations is made according to concord … and not noun prefix" (Doke, 1935: 99). Kikongo is spoken in an area that is currently known as the Democratic Republic of Congo (DRC) (Lewis *et al.*, 2015).

[3] Bleek is reported to have "hypothesized that noun class prefixes in Bantu originally were nouns in their own right which could appear in compounds such as 'person-x', 'tree-x'. As a result of grammaticalization and semantic bleaching the relevant nouns lost their ability to appear independently and always had to be appended to another noun" (cited in Katamba (2006: 106)). For different but related views about the possible origins of the NCP, see i.a. Givón (1975, 1979); Heine (1976); Voorhoeve (1980) and Katamba (2006: 106).

Stellenbosch University https://scholar.sun.ac.za

volume, followed by many others such as Meinhof (1899, 1906; 1932); Guthrie (1948, 1967), Meeussen (1973) (see Hendrikse & Poulos, 1992; Maho, 1999; Hendrikse, 2001; Katamba, 2006; Bostoen & Bastin, 2015). Bleek and Meinhof are regarded as the major architects of the classification system; as such, it is sometimes referred to as the *Bleek-Meinhof system* (Maho, 1999: 3–4; Katamba, 2006: 105). The reconstructed noun class system is generally assumed to have held in the parent Bantu language, variously referred to as *Ur-Bantu* (UB) or *Proto-Bantu* (PB). Table 2.1 below summarises the various revisions of the number and prefix system.

Table 2.1: Versions of the Bantu noun class system (Maho, 1999: 247; Katamba, 2006: 104)

	Bleek's Ancient Bantu (Bleek, 1869: 282ff.)	**Meinhof's UB** (Meinhof & Van Warmelo, 1932: 39ff.)	**Meeussen's PB** (Meeussen, 1967: 97)	**Guthrie's PB** (Guthrie, 1967: 9)	**Welmers's PB** (Welmers, 1973)
1	**mŭ*	**mu-*	**mu-*	**mo-*	**mo-; 1a ø*
2	**ba-*	**ua-*	**ba-*	**ba-*	**va-; 2a * va*
3	**mŭ*	**mu-*	**mu-*	**mo-*	**mo-*
4	**mi-*	**mi-*	**mi-*	**me-*	**me-*
5	**di -~ *li-*	**li-*	**i*	**yi-*	**le-*
6	**ma-*	**ma-*	**ma-*	**ma-*	**ma-; *ma-*
7	**ki-*	**ki-*	**ki-*	**ke-*	**ke-*
8	**pi-*	**vi-*	**bi-*	**bi-*	**vi-; 8x *li-*
9	**n*	**ni-*	**n-*	**ny-*	**ne-*
10	**thin-*	**li-ni-*	**n-*	**ny-*	**li-ne*
11	**lu-*	**lu-*	**du-*	**do-*	**lo-*
12	**ka- (13)*	**ka- (13)*	**ka-*	**ka-*	**ka-*
13	**tu- (12)*	**tu- (12)*	**tu-*	**to-*	**to-*
14	**bu-*	**vu-*	**bu-*	**bo-*	**vo-*
15	**ku-*	**ku-*	**ku-*	**ko-*	**ko-*
16	**pa-*	**pa-*	**pa-*	**pa-*	**pa-*
17	-	**ku-*	**ku-*	**ko-*	**ko-*
18	-	**mu-*	**mu-*	**mo-*	**mo-*
19	-	**pî-*	**pi-*	**pi-*	**pi-*
20	-	**ɤu-*	-	-	**ɤo*
21	-	** ɤî-*	-	-	**ɤi-*
(22)	-	-	-	-	**ɤa-*
23	-	-	*-I (24)*	-	**ɤe-*

Stellenbosch University https://scholar.sun.ac.za

Notwithstanding the foregoing, in some elementary grammars and textbooks, for example in Chichewa, the Bleek-Meinhof numbering system is often not adopted wholesale; instead sets of singular and plural prefixes are used to identify the various noun classes. For the noun classes that have no overt prefixes, the AM serves to provide the mnemonic labelling. Table 2.2 depicts this kind of noun class system in Chichewa as in elementary grammar books such as Nankwenya (1992: 38) and Hullquist (1988), see also Maho (1999: 8, 319) for similar characterisation in other Bantu languages.

Table 2.2: Mnemonic table of Chichewa noun classes as in elementary grammars

Mnemonic class name	**Equivalent Bleek-Meinhof numbering**	**Examples**
Mu-A	1/2	**mu**-nthu (“person”) / **a**-nthu (“people”)
Mu-Mi	3/4	**mu**-tu (“head”) / **mi**-tu (“heads”)
Li-MA	5/6	phiri (“hill”) / **ma**-piri (“hills”)
Chi-ZI	7/8	**chi**-sa (“nest”) / **zi**-sa (“nests”)
I-Zi	9/10	nyanja (“lake”) / nyanja (“lakes”)
Ka-Ti-	12/13	**ka**-mutu (“small head”) / **ti**-miti (“small heads”)
U-Ma	14/6	**u**ta (“bow”) / **m-**auta (“bows”)
Ku-	15	**ku**-dya (“to eat”)
Ku-Pa-Mu	17/16/18	**ku**-mudzi / **pa**-mudzi / **m**-mudzi (“to/on/in the village”)

Although the Bleek-Meinhof numbering is not directly referred to in the Chichewa elementary grammars, the analyses have all the hallmarks of the traditional Bleek-Meinhof system, especially as outlined in Watkins (1937: 24–25). The idea of combining the singular/plural pairs is also adopted in many studies (see i.a. Doke, 1935: 64; Corbett & Mtenje, 1987; Carstens, 1991). Therefore, these variations do not imply any differences in the classification systems. Unless otherwise stated, in the present Chapter I will mainly refer to the traditional noun classes by their Bleek-Meinhof numbering system as NC1, NC2, NC3, etc. rather than by the Mnemonic class nomenclature. However, in Chapter 4 I introduce a different nomenclature that takes the same essence as the mnemonic approach.

Returning to the Bleek-Meinhof schema presented in Table 2.1 above, although the table depicts a maximum of 24 reconstructed noun classes, no modern-day Bantu language has been found to have all of them. Most Bantu scholars claim that the noun class system in the daughter languages of PB have lost some classes over time (see i.a. Givón, 1971a,b; Maho, 1999: 50–54; Katamba, 2006: 106). The highest number of classes in one language appears

Stellenbosch University https://scholar.sun.ac.za

to be 21, such as in Ganda and Ici Ndali as reported in Katamba (2006: 108) and Kishindo (1998: 46), respectively. However, Fortune (1955: 52–53) reports that there are 22 noun classes in Shona, considering there to be a subclass NC1a/2a within NC1/2 pair. Many studies do not treat the subclasses NC1a and NC2a as distinct noun classes, though. Languages with such large numbers of classes are considered to instantiate a "traditional" Bantu noun class system (Maho, 1999: 50–54; Katamba, 2006: 108). At the other end of the continuum, there are also some Bantu languages which are reported to have fewer noun classes, such as three or two. This is observed to be common in the northern Bantu area, for example the language called Mbati has only two classes (Richardson, 1957: 39) and another one called Kako has 3 (Guthrie, 1971: 34). These have come to be referred to as *reduced* or *simplified noun class systems* (Maho, 1999: 3, 50–54; Katamba, 2006: 108). The extreme case is that of the language called Komo, which is said to have no noun classes at all (Guthrie, 1971: 42; Maho, 1999: 53).

Considering these typological variations, it is an oversimplification to talk of the Bantu noun class system as one homogenous phenomenon or identifying one subgroup as representing a typical pattern. In this regard, it is, therefore not a straightforward task to provide an all-encompassing definition of Bantu noun class systems. However, the putative definition is captured in Fortune (1955: 51), who views the Bantu noun class as "a group of nouns which do not differ in prefix and which govern the same concords" (cf. Bokamba, 1993; Maho, 1999; Katamba, 2006; van der Spuy, 2009). However, Fortune's (1955) definition is not representative of the facts, as there are distinct noun classes that trigger the same AMs, for example the Chichewa NC3 and NC14, NC4 and NC9 (see Chapter 4 for specific details). Although it is not explicitly stated in many works, it appears that much of the literature related to the Bantu noun class system has focused on the so-called *traditional systems*, which should rather be considered as characteristic of **some** Bantu languages. This practice has to a large extent misled the direction of the linguistic enterprise in Bantu noun classification systems. Maho (1999: 10) refers to this skewed focus as the *factor of standardization* which he aptly states "has the tendency to create pan-dialectal grammars that can obscure true linguistic variation."(cf. Janda, 1982)[4] However, even within the putative *traditional types*, there are still disagreements regarding the proper characterisation of the noun class system.

[4] The present study will also illustrate that the standardization factor has led to wrong generalisations in Chichewa.

Stellenbosch University https://scholar.sun.ac.za

Some studies put strong emphasis on the NCP while others, such as Guthrie (1967: 13) and Corbett & Mtenje (1987), define the noun class as a "distinct type of agreement". Though the NCP is taken as the key determinant of the classification system as illustrated in the foregoing discussion, some studies have analysed the NCP as also having additional functions, namely derivational and inflectional functions (Mufwene, 1980). The prefixes are said to have derivational functions in the sense that they are used to form nouns from verbal, adjectival or other nominal stems, as illustrated in (5). They are, however, also treated as inflectional morphemes because the prefixes are also seen to convey grammatical information, for example number as in (6) (Mufwene, 1980; Schadeberg, 2001; Ferrari, 2005).[5]

(5) Kiswahili (Mufwene 1980:248)

a.	-refu	u-refu
	"long"	"length"
b.	-levi	u-levi
	"drunk"	"drunkenness"

(6) Chichewa

a.	galu	a-galu
	NC1.SG.dog	NC2.PL-dog
	"dog"	"dogs"
b.	phiri	ma-piri
	NC5.SG.mountain	NC6.PL-mountain
	"mountain"	"mountains"

In examples such as (5), the adjectival roots are assumed to be nominalised by virtue of the prefix *u-* (Harding, 1966: 25–35; Bresnan & Mchombo, 1995: 239). In (6), the Chichewa example shows that by adding prefix *a-* and *ma-* to the nouns *galu* and *phiri,* the nouns express plural number and at the same time change noun class. It should therefore be borne in mind that the so-called *NCP* can appear to constitute a bundle of many functions.

Studies have also shown that the NCP is not the only criterion in determining noun class membership. For example, semantic considerations have also been shown to play a role in the classification criteria of some noun classes (see Section 2.4.3). To provide a proper

[5] Parallels are also drawn with some Indo-European languages, for example, gender assignment in German may depend on the noun's suffix 'where the diminutive suffixes *-chen*, *-lein* override a semantic criterion in the neuter nouns M*ä*dchen "girl", Fr*ä*ulein "miss"'(Kilarski, 2013: 18).

Stellenbosch University https://scholar.sun.ac.za

account of the system that is encoded by the Bantu noun class system is a topic of ongoing debate, as I will demonstrate in the next section.

2.4 The principles behind the Bantu noun classification system

The preceding discussion has already shown that the Bantu noun class system is not as clear-cut as it is conceptualised in the Bleek-Meinhof system. Although the system is based on a set of NCPs, both the literature on this subject and the empirical evidence emerging in this study (see Chapters 3 & 4) show that a very significant number of nouns do not have overt prefixes (cf. Matiki, 2001: 66–67). For example, in Chichewa the prefixless noun *gule* ("dance") triggers AMs similar to nouns of the *m(u)-* prefix group (i.e. NC1 on the Bleek-Meinhof schema), whereas the other prefixless noun *gulu* ("group") belongs to the *li-* prefix group (NC7). The Bleek-Meinhof approach claims that some of these nouns have lost their prefixes in the course of language evolution;[6] hence, they are referred to as *null prefix nouns* (see i.a. Doke, 1954; Givón, 1971a; Herbert, 1985; Bokamba, 1993; Ferrari, 2005; Batibo & Kgolo, 2016).[7] The unfortunate implication of such accounts, however, is that we begin to see language change taking place in Bantu languages as unsystematic to the extent that the relevant phenomenon eludes explanation. Givón's (1971a: 33) widely quoted statement is not far from this observation; he writes, "[t]he great bulk of non-human, non-derived concrete nouns in the Bantu lexicon appear to be distributed all over the noun-class system *in inexplicable chaos* [emphasis added - PM]" (cf. Herbert, 1991: 105; Kishindo, 1998: 45; Maho, 1999: 67). If we look at the Bantu noun class system through the prism of the Bleek-Meinhof system, Givón's statement seems justified.

The pursuit of the principles underlying the Bantu noun class system is perhaps the most contentious topic in Bantu language studies, as I will continue to show in the next subsections. Three criteria are generally debated, namely (i) the NCP, (ii) grammatical agreement (concord) and (iii) the semantic criterion. Most studies employ at least two or even all three of these criteria at the same time, most commonly within the Bleek-Meinhof schema. I argue that by adopting the Bleek-Meinhof schema, one has already unwittingly

---

[6] The changes in question are taken to manifest in three ways, (i) merging of some noun classes (Richardson, 1957: 39; Givón, 1971a: 34; Herbert, 1991; Maho, 1999: 6), loss of prefixes (Doke, 1954; Cole, 1967; Givón, 1972), and semantic changes (Givón, 1971a; Denny & Creider, 1986). For related discussion on the general evolution of Bantu languages, see also Okhotina (1975) and Maho (1999, chap. 5).

[7] Givón (1971a: 33,35) also refers to them as *non-derived* or *original* nouns.

Stellenbosch University https://scholar.sun.ac.za

accepted the prefix criterion as the primary basis for nominal classification (see section 2.5.2 and Chapters 4 & 5 for a detailed discussion supporting this position). In the next three sections, I summarize the debate for the three proposed criteria.

2.4.1 The NCP criterion vis-à-vis the locus of noun class features

Although the Bleek-Meinhof schema has been widely accepted and there seems to be a high level of consensus about the defining feature of Bantu noun classes, the status of the NCP is an issue at the centre of ongoing controversy in the literature. One source of controversy concerns the exact locus of the noun class features, that is, whether it is in the prefix, the stem, both or neither.[8]

2.4.1.1 NCP as the locus of noun class features

The first view takes the noun prefix to be the host of noun class features (see i.a. Doke, 1954: 51; Welmers, 1971, 1973; Kishindo, 1985: 2; Sproat, 1985; Myers, 1987; Bresnan & Mchombo, 1995; Mchombo, 2004: 3; van der Spuy, 2009: 3; Hlungwani, 2012). For instance, Welmers (1971: 2) claims that it is more accurate to say that a stem "appears in" classes rather than saying that a noun "belongs to" a given class or pair of classes. One of the standard Bantu interlinear morpheme-by-morpheme glossing conventions has also been modelled on this view, where the prefix is glossed as carrying the noun class feature. Consider the following examples.

(7) **Theoretical implications of the standard interlinear glossing in Bantu studies**

a. Chichewa (Mchombo, 2004: 4)
m-lenje
C1-hunter
"hunter"

b. Zulu (Poulos & Bosch, 1997: 9)
umu-ntu
C1-person
"person"

c. Makhuwa (Kisseberth, 2006: 560)
Ø-kharamu
Class1a-lion
"lion"

In (7a) the Chichewa noun *mlenje* is broken down into two morphemes, the NCP *mu-* (marking NC1 membership) and the nominal stem *-lenje*. A similar glossing is also shown

[8] This locus issue is well captured in the title of Greenberg's (1977) article, "Niger-Congo noun class markers: prefixes, suffixes, both or neither".

Stellenbosch University https://scholar.sun.ac.za

with the Zulu example in (7b). The Makhuwa noun *kharamu* in (7c) does not have a prefix but in this case, it is assumed to have a covert prefix (Ø), which serves as the class marker indicating that the noun also belongs to NC1a.

Empirical support for the view that noun class features are found on the prefix are plentiful in Bantu languages. For example, in Chichewa, by varying the prefixes, *mu-, u-, chi-, zi-, ka-* and *ti-*, attached to the nominal stem *–nthu* ("being"), the derived noun also changes its class membership from NC1, NC2, NC7, NC8, NC12, to NC13, correspondingly as illustrated in (8) (cf. Bresnan & Mchombo, 1995, n. 35; Carstens, 2008: 137–138).[9]

(8) **Change in class membership of nouns formed from stem *-nthu* in Chichewa**

Prefix	**mu**	**a-**	**chi-**	**zi-**	**ka-**	**ti-**	**u-**
Noun	mu-nthu	a-nthu	chi-nthu	zi-nthu	ka-nthu	ti-nthu	u-mu-nthu
Class	1-being	2-being	7-being	8-being	12-being	13-being	14-person
Gloss	"human being"	"human beings"	"thing/big thing"	"things"	"small thing"	"small things"	"personhood"

The pattern in (8) is precisely what Welmers (1971: 2) conceived as a 'stem appearing in classes'.

Although the view that noun class features are carried by the prefix gets strong support from facts such as those illustrated in (8), contrary examples are also abundant. For example, as shown in Table 2.1, in Chichewa and many other languages, the prefixes for NC1 and NC3 are homophonous (cf. Msaka (2003: 10) [Chichewa], Maho (1999: 309) [Kikongo], Contini-Morava & Kilarski (2013: 270) [Kiswahili], van der Spuy (2009: 206–207) [Zulu]). It is hard to claim that the noun prefix plays any major role in the noun classification of the nouns in (9) below.

(9) Chichewa nouns derived by noun prefix *m-* belonging to NC1 and NC3

	NC1		**NC3**
a.	(i) **m**-lim-**i**	(ii)	**m**-tanth-**o**
	NCP-cultivate-NFV		NCP-climb-NFV
	"farmer"		"bridge"

[9] There are two other nouns that can be formed from this root in Chichewa. These are *ma-nthu* ("mother/ queen of ants"), NC1 and *thu-nthu* ("trunk/ body structure") for NC5. I have omitted these examples here because they are not recognised in the previous studies and I discuss them in Chapter 4, Sections 4.3.6 and 4.4.3.

Stellenbosch University https://scholar.sun.ac.za

b. (i) **m**-sodz-**i**
NCP-fish-NFV
"fisherman"

(ii) **m**-sok-**o**
NCP-sew-NFV
"line of stitching/seam"

c. (i) **m**-leng-**i**
NCP-create-NFV
"creator"

(ii) **m**-tay-**o**
NCP-lose-NFV
"miscarriage/abortion"

All the nouns in this example are derived from verb stems by attaching prefix ***m-*** and changing the verb's final vowel. By the look of things, the classification of the nouns in (9) can be a factor of either suffixes (-***i*** and -***o***) or the stems. However, the evidence is apparent that the classification in (9) is also dependent on the NFV and not the prefix (see also Chapter 3, Section 3.3.2).[10] For example, in Chichewa, when the same stems used for NC3 in (9d, e, f) are suffixed with the NFV *-i* the generated nouns potentially trigger NC1 AMs (cf. Givón, 1971a: 35; Kishindo, 1985: 3–6; Katamba, 2006: 103; van der Spuy, 2009: 198, 206).[11] It is clear that the NCP is not the sole morphological determinant of noun classification despite it being the key criterion in the Bleek-Meinhof system. Next, I consider views that take the stem as the host of noun class features.

2.4.1.2 The stem as the locus of noun class features

Carstens (1991, 1993, 1997: 370, 2008: 137) among others, argue that "much-noted alternations" of the kind illustrated in (8) have obscured the simplicity of the Bantu noun class system. Carstens (1997: 370) observes that arguments for prefixes as the locus of noun class features are based on what she describes as *non-core phenomena* such as diminutive formation where prefixes are regular and predictable as they attach to nouns from any other noun class. She argues that class/gender features are actually located in the stem (see also Kinyalolo, 1991: 231; van der Spuy, 2009: 203). Carstens proposes an analysis of Bantu noun classes as a gender system, like that manifested in Romance languages (see also Carstens, 2008 for more explicit comparisons). Under this view, gender features are inherent in the stem and the so-called NCPs become "gender-specific number morphology", which are assumed to be added to nouns by means of morpho-syntactic rules, specifically

[10] Givón (1971a: 35) also observes that the suffix *-i* is a common morpheme for nominalizing subject-agentive verbs in many Bantu languages. He, however, does not associate it with the noun classification process. For other forms of suffixations in Bantu languages see also Maho (1999: 91,97), and Ashton, Mulira, Ndawula & Tucker (1954: 371–379).

[11] For contrary analyses of the suffixes in Luganda, see Ferrari (2005: 353–358) who denies the role of suffixes in nominal derivation and classification of Luganda.

Stellenbosch University https://scholar.sun.ac.za

what are referred to as *Spell-Out rules* (Carstens, 1991: 18–19, 2008: 135).[12] This proposal takes singular and plural pairs as belonging to one gender, with the consequence that the first 11 classes in Kiswahili reduce to six, as shown in Table 2.3.

Table 2.3: ***Genders in Kiswahili*** (Carstens, 1991: 18, 2008: 136)

Gender	Traditional noun classes
Gender A:	stems of NC1/ NC2
Gender B:	stems of NC3/ NC4
Gender C:	stems of NC5/ NC6
Gender D:	stems of NC7/ NC8
Gender E:	stems of NC9/ NC10
Gender F:	stems of NC11/ NC10

According to this approach, the information related to noun class and number can be schematised as in (10). Contrary to the traditional glossing convention shown in (7) above, here the noun prefixes only carry number features whereas the noun class features are assumed to be located in the stem.

(10) **A schematic glossing of noun class features in Kiswahili based on Carstens (2008: 140)**.

a. m -tu
[singular] [Gender A]
[person]
"person"

b. wa -tu
[plural] [Gender A]
[person]
"person"

In more specific terms, Carstens (2008: 136) states that gender is an inherent lexical feature of stems and number is a syntactic feature, which appears in the head position of the number phrase, into which the stem must move to receive its prefix. She argues that the Spell-Out rules for each gender specify the type of number prefix, one for singular and the other for plural. For example, in (11) below, Gender A in Kiswahili specifies that its singular form should be the prefix *m-* while the plural form should be *wa-*.

[12] A similar noun class system has also been proposed for Chichewa by Watkins (1937:24), see also Corbett & Mtenje (1987:6). However, these other works continue to consider the NCP as the locus of noun class features.

Stellenbosch University https://scholar.sun.ac.za

(11) Some Spell-Out rules yielding Kiswahili noun prefixes (Carstens, 2008: 136).

[Singular]	<-> /m-/	/_N Gender A
[Plural]	<-> /wa-/	/_N Gender A
[Plural]	<-> /vi-/	/_N Gender D
[Singular]	<-> /ki-/	/_N Gender D

The advantage of Carstens' (1991) approach is that it resolves the inconsistencies encountered by the NCP approach assumed in the Bleek-Meinhof system, especially in accounting for null prefixed nouns and noun classes with multiple prefixes. However, the first challenge that Carstens' approach faces is that it cannot account for examples such as (8) above, where the stem is seen to combine with several prefixes and in each case the noun class changes. In such instances, the prefix is seen to express functions beyond just number. The second challenge is that not all noun classes have clear-cut singular-plural patterns. For example, in Chichewa there are plural noun classes that take their singular forms from different classes (see Chapter 4 for specific details).

Carstens' analysis has also been criticised on several grounds by Bresnan & Mchombo (1995) and Amidu (1997). On the one hand, Amidu (1997) challenges the general Bleek-Meinhof approach of classifying nouns based on prefixes and specifically the Kiswahili analysis as proposed by Carstens (1991).[13] Bresnan & Mchombo (1995: 180), on the other hand, are of the traditional view that treating prefixes as syntactic units violates the lexical integrity principle, "which states that words are built out of different structural elements and by different principles of composition than syntactic phrases".[14] So far I have shown what appears to be conflicting evidence, one supporting the noun prefix and the other supporting noun stem as the locus of noun class features. This has motivated the third view, which I turn to next.

[13] The essence of Amidu's arguments is summarised in Section 2.5.1 below.

[14] This argument is no longer compatible with the most recent generative morphosyntactic frameworks such as Minimalist syntax (Chomsky, 1995), Cartographic syntax (Cinque & Rizzi, 2008), Distributed Morphology (Halle & Marantz, 1993; Harley, 2014) and Nanosyntax (Taraldsen, 2010) where the distinction between morphology and syntax is no longer recognised. I therefore do not pursue their arguments any further

Stellenbosch University https://scholar.sun.ac.za

2.4.1.3 Both NCP and stem as the locus of noun class features

A third view on the locus of noun class features is based on the derivational characteristics of the Bantu noun class system (Mufwene, 1980; Myers, 1987; Ferrari, 2005; Ferrari-Bridgers, 2008). Under this view, both the stem and the prefix are taken to provide the noun class/gender features for nominal classification. As shown in the two contrasting stances discussed above, in some cases noun class assignment seems to be dependent on the NCP and in other cases on the nominal stem or NFV. Arguments that consider both the NCP and the stem as sources of the noun class features are supported in detail in Ferrari's (2005) work. However, in the present study, I will not review Ferrari's (2005) proposals because they do not illuminate the phenomenon any better. I argue that the problems regarding the locus of noun class features are due to the Bleek-Meinhof characterisation of the noun class system, which Ferrari's work does not challenge.

In summary, as the preceding discussion has shown, the locus of noun class features remains an unresolved issue, with different types of empirical evidence pointing in different directions. However, a broader perspective shows that there is something incorrect about the manner in which the examples are selected and given weight. For example, why should nouns with prefixes (derived/inflected nouns) be given more weight in determining the nature of the nominal system than non-derived (prefixless) nouns? I (2003) argue that non-derived nouns are equally substantive and proposes a nominal classification analysis that focuses on the non-derived nouns before considering the derived nouns (see section 2.5.2 for overview discussion).

First, however, I turn our attention to grammatical agreement, the second, oft-cited basis for identifying noun classes.

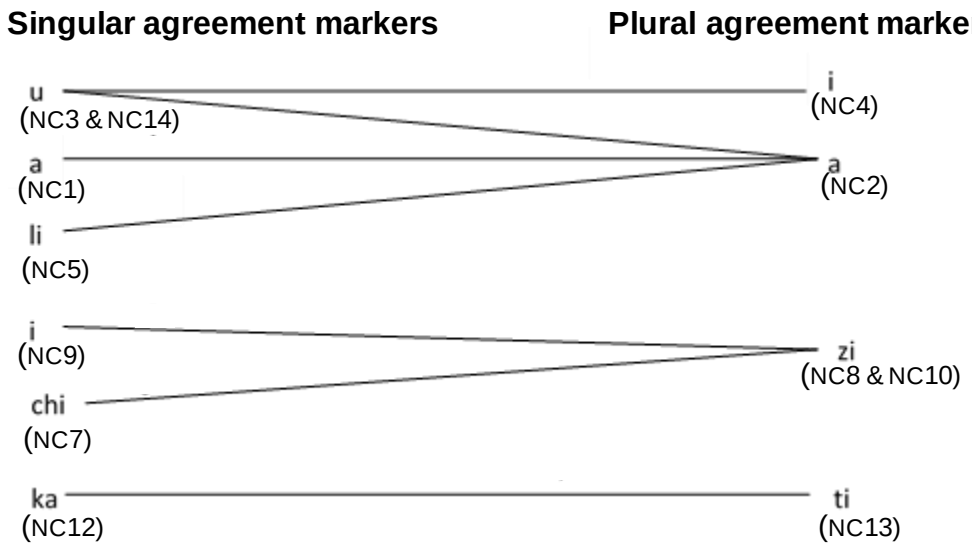
2.4.2 The grammatical agreement (concord) criterion

Although NCPs have played an important role in the development of the Bleek-Meinhof system, it should be pointed out that, as introduced in Chapter 1, a noun class is broadly considered as a set of nouns triggering the same concords and sharing some noun-internal features (e.g. the NCP) (see i.a. Fortune, 1955: 51; Hockett, 1958: 231; Bokamba, 1993: 211–237; Dimitriadis, 1997: 1). However, it is not explicitly stated in many studies how the NCP criterion and the agreement evidence work together to realise the noun classes; and this is of key importance, especially given that there are many distinct noun classes

Stellenbosch University https://scholar.sun.ac.za

triggering the same AMs. As I have stated above, by adopting the Bleek-Meinhof system, one has already confined the analysis to the NCP criterion. Taking the AM-based approach is to challenge the Bleek-Meinhof system as this gives rise to a much smaller noun class system. With the exception of Corbett & Mtenje (1987), not many studies have taken this route in studies of the Chichewa noun class system. Corbett & Mtenje (1987) proposes an agreement-based (*target gender*) noun class system for Chichewa as summarised in Table 2.4.[15]

Table 2.4 The AM-based gender system of Chichewa (Corbett & Mtenje, 1987: 8)[16]



There are at least two important points about Table 2.4 that are instructive. First, the Table shows that there are 6 singular AMs and 4 plural AMs in Chichewa which give us 10 distinct agreement-based noun classes. Here, we first see the disparity between the *NCP-based noun classification approach* and *agreement-based noun classification approach* in Chichewa. Since agreement is in fact the only way to detect the noun class system - as I have shown above, we cannot straightforwardly rely on nominal prefix markings. Corbett & Mtenje's (1987) analysis is a pointer to something that is amiss with the traditional Chichewa noun class system. Take note, though, that NC15, NC16, NC17 and NC18 are excluded from this noun class system because Corbett & Mtenje (1987: 6) are sceptical about the status of these expressions since they lack normal singular-plural pairing. Second, some

[15] This work builds on Watkins' (1937) study of Chichewa grammar. Corbett & Mtenje(1987: 6) state that they grouped singular and plural class pairs while also referring to the Meinhof numbering following Doke's (1935:64) recommendation.

[16] Table 2.4 has been adapted slightly, for example the Bleek-Meinhof numbering in brackets has been added and the previous numbers counting the genders have been removed for expository purposes.

Stellenbosch University https://scholar.sun.ac.za

noun classes that control 'homophonous' AMs have been combined into one. These are: the singular NC3 and NC14 and the plural NC8 and NC10. At the same time, however, plural and singular noun classes that control homophonous AMs have been left as distinct noun classes, for example NC4 and NC9. Justification for this treatment of singular and plural classes has not been provided. I resolve these issues in chapters 3, 4, and 5.

As noted above, the second challenge with the noun class system represented in Table 2.4 is the exclusion of NCs 15, 16, 17 and 18. An agreement-based approach ought to explain how these expressions come to control agreement in a manner that seems to parallel that of the nouns that distinguish between singular and plural. This is even more serious considering that in these classes, we also find nominal expressions, such as nominal infinitives. By excluding these agreement classes, then, we risk missing the general principles behind the Bantu noun and agreement system.

The third challenge with Corbett & Mtenje's (1987) analysis relates to how the underlying principles behind this revised noun class system in Chichewa are to be accounted for. Corbett & Mtenje continue to follow the traditional accounts, claiming that nouns are assigned to their gender on the basis of the criteria discussed for the broader Bantu noun systems: these are morphological form (the NCPs) and the semantic basis, for example that nouns denoting humans belong to Gender 1 (Corbett & Mtenje, 1987: 9–10). This account of the principles behind the revised Chichewa noun class system is rather simplistic and does not offer any new solutions to the problems discussed so far. For example, the combined NC 3 and 14 are traditionally identified by two different criteria (see Table 2.5 below), which have not been reconciled in the revised system. Therefore, apart from showing that the Bleek-Meinhof system is redundant, Corbett & Mtenje (1987) do not provide an illuminating account of the underlying principles for their proposed system. This is why, about two decades later, Mchombo (2004: 3) writes "[o]f interest is the question of the basis for this classification of nouns. This is an issue that still awaits a definitive response."

In the next section, I summarise proposals for the third basis of Bantu noun class system, the semantic criterion.

2.4.3 The semantic criterion

The manner in which NCPs (a basis for controller genders) and concordial agreement (a basis for target genders) have been analysed in the reviewed literature does not reveal a

Stellenbosch University https://scholar.sun.ac.za

unified underlying principle behind the Bantu noun class system. The picture we have is that of divergent accounts applicable to some specific type of data, each of which is correct but not compatible with the other (see also Palmer & Woodman, 2000). In this regard, a third criterion has been explored: the possibility of a semantic basis to the classification system (cf. Burton & Kirk, 1976; Batibo, 1987; Orr, 1987; Spitulnik, 1989; Kgukutli, 1994; Selvik, 1996, 2001; Contini-Morava, 1997; Demuth, 2000; Palmer & Woodman, 2000). These studies have experimented with a wide range of diagnostics and approaches, such as cognitive/psycholinguistic classification (cf. Contini-Morava, 1996; Selvik, 1996, 2001; Moxley, 1998), cultural classification (Palmer & Woodman, 2000), type frequency (Matiki, 2001) and language acquisition evidence (Kunene, 1979; Suzman, 1980; Connelly, 1984; Tsonope, 1987; Demuth, 2000).[17]

Just as in the case of the literature on the NCP criterion, studies advocating for a semantic basis for the Bleek-Meinhof system are also characterised by extreme lack of consensus, both in relation to particular languages and in relation to the Bantu family more generally (cf. Givón, 1971a; Batibo, 1987: 5–6; Herbert, 1991: 105; Maho, 1999: 63). On the one hand, there are works, such as that of Richardson (1967) and Amidu (1997) that take the most sceptical stance on the possibility of any semantic generalisation where the Bantu noun class system is concerned. Richardson (1967: 378), for instance, is widely cited as having argued for the position that the Bantu noun classification system is arbitrary. He (1967: 378) claims that "it is impossible to prove conclusively by any reputable methodology that nominal classification in Proto-Bantu was indeed widely based on conceptual implication." However, to some extent he (1967: 378) does acknowledge that there is something in the Bantu noun class system suggestive of a semantic organization, especially NC1/NC2 as denoting human beings and NC12/NC13 as denoting diminutive meanings. By contrast, Amidu (1997: Chapter 1 & 7) strongly opposes any appeal to a semantic criterion in the classification of nouns in Kiswahili (see Section 2.5.1).

Contrary to Richardson (1967) and Amidu (1997), on the other hand, there are some confident claims about the semantic basis for the Bantu noun class system. For example, Denny & Creider (1986: 217), among others, in a direct response to Richardson's (1967)

[17] For a comparative review of Richardson (1967), Selvik (2001) and Palmer & Woodman (2000), see Dingemanse (2006).

Stellenbosch University https://scholar.sun.ac.za

statement, claim to have a "reputable methodology" to show that noun prefixes are associated with configurational or shape meanings and they show that this type of formal meaning encoding is common among the world's languages (see also Cole, 1955; Denny & Creider, 1986). Amidst these disagreements, some highly speculative semantic generalisations, believed to have held in PB, have been proposed. For example, it has been suggested that NC1 is composed of nouns that refer to humans or animates in general. It has also been consistently stated that the derived nouns making up the diminutive NC12, the infinitive NC15, and the locatives NC16, NC17 & NC18 have a regular semantic basis. However, take note that, with the exception of NC12, the three classes (NC16, NC17 & NC18) have been treated with scepticism and have been excluded from the core noun class system in some studies (see Givón, 1972: 12–13; Corbett & Mtenje, 1987; Carstens, 1997: 370, 2008). The proposed semantic bases associated with the NCPs were shown in Table 1.1 and are repeated here as Table 2.5 (Maho, 1999: 51; Demuth, 2000: 6; Katamba, 2006: 114–116).

Table 2.5: Bantu noun classes and their commonly associated semantics

Noun class	NCP	Common semantics of the NC
NC1	*mù-	humans
NC1a	*ø-	kins, personified animals
NC2	*va-	honorific, plural to NCs1 and 1a
NC2a	*va-	honorific, plural to NC1a
NC3	*mu-	trees, plants, inanimates
NC4	*mi-	plural to NC3
NC5	*li-	miscellaneous, paired things, augmentatives
NC6	*ma-	liquids, collectives, plural to NC 5,9,11,14 and 15
NC7	*ki-	inanimates, manner/style, diminutives, augmentatives
NC8	*bi-	plural to NC7
NC9	*n-	animals
NC10	*n-	plural to NCs9 and 11
NC11	*du-	Long thin things, abstracts
NC12	*ka-	diminutives
NC13	*tu-	plural to NC12
NC14	*bu-	abstracts, mass nouns
NC15	*ku-	infinitives
NC16	*pa-	locatives, 'near' or 'explicit'
NC17	*ku-	locatives, 'remote' or 'general'

Stellenbosch University https://scholar.sun.ac.za

NC18	*mu-	locatives, ‘inside’
NC19	*pi-	diminutives
NC20	*ɾo	augmentatives, diminutives
NC21	*ɾi-	augmentatives, pejoratives
NC22	*ɾa-	plural to NC20
NC23	*ɾe-/*i-	locative, unspecified

The semantic generalisations in Table 2.5 above are not consistent across studies or languages, with the exception of NC1, which has consistently been identified as referring to humans or animates. Nevertheless, the consistencies observed about the semantics of NC1 are not incontrovertible: several examples of non-animate nouns are also found. For example, in Chichewa there are nouns such as those derived from the traditionally neglected prefixes *ka-* and *na-* which are either animate or inanimate, see example (11).[18] In Chapters 4 and 5, I will show that the distribution of inanimate nouns shown in (11) is in fact systematic.

(11) a. ka-lozera
NCP-point for
“contour ridge”

b. ka-yera
NCP-be white
“white beans”

c. na-nyongo
NCP-gall bladder
“clitoris”

d. na-khoma
NCP-wall/hit
“large tray coated with castor oil and charcoal”

In addition to the counterexamples in (11), the semantic regularities for NC1 are also overridden by the spread of animate nouns across the entire Bantu noun class schema, for instance, NC7, NC9, NC14, etc. Maho (1999: 63), for example, correctly observes that, “virtually all noun classes contain what grammarians label as *miscellaneous nouns*.” A commonly held view to account for this state of affairs is that the contemporary noun class is an eroded version of an earlier system in PB (cf. Givón, 1971a; Batibo, 1987; Bokamba, 1993; Batibo & Kgolo, 2016). This position, of course, is not accepted by others, for example, Spitulnik (1987: 7–8) argues that the “‘random’, ‘non-systematic’ state of present day Bantu languages” is not a “relic of a more homogenous proto-system”. She (1987: 7–8) claims that the problem is with the tradition of accounting for the semantic organisation of the noun class systems by "giving an inventory of the typical referents of the *morphologically distinct classes* [emphasis added - PKM]" (cf. Hendrikse & Poulos, 1992: 202; Zawanda & Ngcobo,

[18] See section 2.5.2 and Chapter 3 for a detailed discussion of these prefixes.

Stellenbosch University https://scholar.sun.ac.za

2008: 316). Spitulnik's (1987) observation points out one of the major weaknesses of endeavours to establish a universal semantic basis in the Bleek-Meinhof system – that is, the schema is largely a formal system, devised on the basis of morphological characteristics and not on a semantic basis. Contrary to her own observation, Spitulnik (1987) goes ahead to propose a highly abstracted cognitive-semantic analysis of the Bleek-Meinhof noun class system.

Although the attempts to establish a semantic basis for the Bantu noun class system have clarified some aspects of Bantu noun class systems, most of the studies have been observed to be speculative, vague and based on non-representative data samples (see i.a. Matiki, 2001: 71; Katamba, 2006: 111–119; van der Spuy, 2009: 201). Katamba (2006: 119) concludes that despite such studies offering some insights into the problems concerning the Bantu noun class system, the "[…] real problems still remain unresolved."

2.4.4 Interim summary

As I have shown in the preceding discussion, Bantu scholars have generally adopted the diachronically based Bleek-Meinhof system and to a large extent the studies have worked to fit their analyses into this framework. These analyses share four key foci. First, the view that the prefix is the major criterion for noun classification. Second, the assumption that the NCP-determined noun classes show traces of semantic grounding. The semantic properties are described either on the basis of concrete meanings of real-world objects or of sets of abstract cognitive notions. Third, the assumption that the two criteria are not consistently manifested in modern day Bantu languages due to language evolution. For this reason, the Bantu noun class system eludes a proper analysis. Fourth, the assumption that all AMs are triggered by nominal expressions; hence the phenomenon is designated as strictly a noun class system to the extent that expressions that do not show singular-prefix contrasts are excluded in some analysis of the noun class system. However, there are two important contrary analyses to the putative Bleek-Meinhof system. First is Carstens (1991, 2008) who argues that NCPs do not carry the noun class features but noun class specific number features. Carstens' work, however, continues to endorse the Bleek-Meinhof system by organising the noun classes based on the putative singular-plural noun class pairs. The second work with divergent views is that of Corbett and Mtenje (1987) where the noun classes are identified by the agreement-based approach which yields a relatively smaller

Stellenbosch University https://scholar.sun.ac.za

noun class system. However, this work also assumes the putative Bleek-Meinhof criteria as determining the underlying principles to the classification system.

By now, we can see that most works on the Bantu noun class system have set out from the assumption that the Bleek-Meinhof system is correct. This is one of the potential problems that has led to the limited level of success in this domain of inquiry. In the next section, I review works that have radically rejected the Bleek-Meinhof system. The first is the work of Amidu (1997) followed by my own (2003) work.

2.5 Synchronic approaches to the study of Bantu noun classes

In the next two subsections, I present analyses that do not take the Bleek-Meinhof assumptions as the point of departure. These works are Amidu (1997) and my (2003) undergraduate research paper. Both works question the diachronic approach in analysing the synchronic state of Bantu languages. I summarise Amidu's (1997) core arguments and proposals in the next section.

2.5.1 Amidu's (1997) *class projection principle analysis*

Amidu (1997: 5) observes that traditional analyses' failure to provide a proper account of the Bantu nominal classes could be due to the following two factors: [19]

(12) a. "an inherent weakness in Bantu language structure itself," or,
b. "a misunderstanding by linguists of the principles underlying Bantu noun class."

According to (Amidu, 1997: 5) the problems are due to the latter. He argues that "the failure of the noun classes theory derives from a basic fallacy of Bantu grammatical conception of classes." He claims that the generalisations made by the traditional analysis are not compatible with empirical facts in Kiswahili. In this regard, Amidu (1997: 2–3, 28-49) rejects the Bleek-Meinhof account of the Bantu noun class system in its entirety. Amidu's (1997) views are, however, presented in his own intricate theoretical framework, which he calls *Linguistic Empirical Grammar* (LEG). To avoid diverting from the core objectives of the study, this theoretical framework will not be introduced; instead I provide the essence of his arguments in non-LEG terms.

[19] Amidu (1997) also refers to a series of his own works where these ideas have been initially presented. This (1997) work, however, happens to be the only work that I have access to, and it is the latest in his series.

Stellenbosch University https://scholar.sun.ac.za

Amidu (1997) challenges the traditional analysis on several grounds. First, he (1997: 2) questions the use of diachronic criteria for the synchronic analysis of what he calls modern classes in Kiswahili and Bantu in general. Amidu (1997: 200) argues as follows:

> Diachronic and comparative rules may help to throw light on areas of difficulties in some l.u.d [language under discussion] but they are not the synchronic rules of the l.u.d. The rules of a grammar are, therefore, those which are in operation here and now.[20]

In this regard, Amidu (1997: 15) claims that the only justification for advancing the traditional analysis "appears to stem from a desire to impose a uniform general diachronic Bantu pattern on the language [Kiswahili] in opposition to its synchronic evidence."[21] He therefore motivates a synchronic analysis of the phenomenon in Kiswahili.

Second, Amidu (1997) challenges the traditional appeal to semantic criteria in accounting for the grammatical patterns in the synchronic grammar of Kiswahili. As was demonstrated in Section 2.4.3 above, the issue of semantics in structuring Bantu noun classes is muddled, with conflicting views. Amidu (1997: 191–233) raises numerous further counterexamples to previous semantic analyses of Kiswahili classes. His main contention however, is that "class affixes are formal phonological constructs which are only accidentally conceptual in meaning" (Amidu, 1997: 1).

Third, Amidu (1997: 2, 28–49) rejects the idea that noun classes control concordial affixes on other word categories such as adjectives and verbs. Instead, he (1997: 5) claims that the phenomenon has been wrongly conceived as involving *noun classes*, when it should rather be conceptualised as centring on ***genetic classes*** as reflected in the title of his book, *Classes in Kiswahili*. The underlying assumption here is that the presence of what have been referred to as NCPs and AMs is not controlled by one individual head, the noun, but by a mechanism more general than the individual heads, which he refers to as the *class projection principle* (CPP). In this regard, the CPP is assumed to generate a non-hierarchical string which only accidently become hierarchical as a result of post-classification syntactic rules. Therefore, given a Kiswahili sentence comprising a noun, a demonstrative, an

---

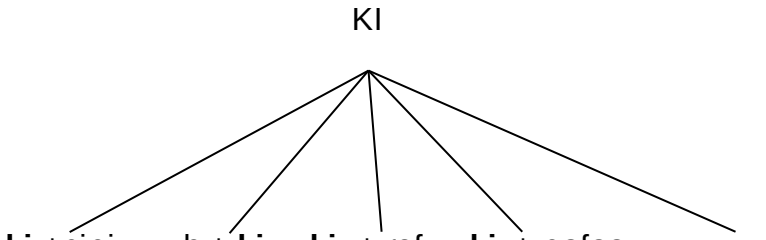
[20] Take note that Amidu's perspective here is similar to that of Lightfoot (1979, 1991, 1999) and others as introduced in Chapter 1, Section 1.1.

[21] This observation is related to what Maho (1999: 10) calls *the factor of standardization*; see also Janda's (1982: 161) critique of what he calls *pan-dialectical grammars*.

Stellenbosch University https://scholar.sun.ac.za

adjective and a verb, the affix *ki-* will be assigned to these various elements as schematised in (13).

(13) **The CPP that derives the Kiswahili sentence** (Amidu, 1997: 32)



[(handle) (this) (long) (it is suitable)]
"This long handle is suitable."

In (13), the affix *ki-* is represented by the underlying form *KI.* As illustrated in (13), the CPP is argued to be responsible for affixing *ki-* to each of the parts of the sentence. The subscript numbers in (13) are simply representing a count of the instances where the affix *ki-* is employed. To account for the covariance observed between the change of the noun and the AMs in the rest of the sentence, Amidu (1997: 37–38) claims that there is a theoretical device, called *tandem limitation*, that sees to it that the leftmost projection dominates all the rightmost projections. However, this may not account for agreement involving postposed controllers, such as object agreement.

Although Amidu (1997) seems to radically depart from the traditional analysis, his proposal makes the same mistake as the traditional approach, namely assuming that all nouns are derived by means of affixation. Due to this, Amidu's approach fails to account for forms of class agreement involving null prefix nouns and nouns that are derived by other prefixes different from the AM. In Chichewa, just as in many other Bantu languages, not all nouns are derived by affixation. It is not clear how the CPP can assign a class marker to a noun in one case, but not in another; consider (14) in this connection:

(14) a. Li-tchowa li-modzi li-ku-sowa.
PREFIX-flywhisk 5.one 5.SM-T/A -miss
"One flywhisk is missing."

b. Gulu li-modzi li-ku-sowa.
ø-group 5.one 5.SM-T/A -miss
"One group is missing."

c. Gulu *modzi li-ku-sowa.
ø-group 5.one 5.SM-T/A -miss

Stellenbosch University https://scholar.sun.ac.za

d. Na-kubala m-modzi a-ku-yankhula.
PREFIX-giving birth 1.AM-one 1.SM-T/A-speak
"One parent is speaking."

In Amidu's (1997) terms, the noun *li-tchowa* ("flywhisk") in (14a) gets assigned the class marker *li-* but under the same circumstances the noun *gulu* ("group") (14b) is not assigned a class marker. Equally in (14d), the noun is derived by a different prefix, which is not homophonous with the AM. It is not clear why the CPP applies differently in this instance. In addition to this, the putative null assignment or variant prefix only happens on nouns and not on modifiers and verbs; for example, it is not possible in Chichewa for the adjective to have a null AM, as shown in (14c).

As radical as Amidu's (1997) arguments may sound, on different grounds I agree with him that the problem at hand cannot be viewed strictly as a noun class phenomenon. As I will show in chapters 4, 5 and 6, there are also non-nominal elements that control agreement in Chichewa, such as CPs and PPs, which suggest that the phenomenon should also be viewed in more general terms as involving agreement classes. However, I disagree with him regarding the fact that the agreement system does not involve a noun class subsystem. There are clear cases of semantic and phonologically-determined agreement in Chichewa (see also Mchombo, 1978: 104; Matiki, 2001: 76–77), which cannot be accounted for without appealing to the notion of noun classification. Besides, this phenomenon is not only unique to Bantu languages, but observed crosslinguistically. Therefore, Amidu's (1997) proposal cannot isolate Bantu languages, and Kiswahili specifically, as regards the phenomenon of agreement and nominal classification.

In summary, Amidu (1997) only succeeds in pointing out the major weaknesses of the traditional Bantu noun class analysis. For example, his criticism of diachronic and comparative approaches to account for the synchronic facts of Bantu grammar is indeed well founded. As pointed out in Chapter 1, the diachronic perspective assumed in the Bleek-Meinhof system is not in keeping with modern generative perspectives on how grammars are constructed: the acquirer is key, and the acquirer obviously has no access to the history of their language.

However, Amidu's complete rejection of the presence of a noun class system in Bantu and also of some regular aspects of the semantic criterion in the noun class system is difficult to

Stellenbosch University https://scholar.sun.ac.za

defend in languages like Chichewa.[22] As I will show here, the apparently random traits of semantic and morpho-phonological regularities within the Bleek-Meinhof system are quite illuminating of a system that has indeed been misconceived. For this reason, Amidu's (1997) analysis misses some very important insights from the descriptive work that the traditional analysis has accumulated.[23] In the next section, I consider my (2003) proposal, which specifically attempts to construct the noun class system of non-derived (prefixless) nouns.

2.5.2 The non-derived noun class system (Msaka 2003)

This section provides a brief outline of the analysis proposed in my (2003) undergraduate research paper. With reference to Chichewa, I (2003) argue that the Bleek-Meinhof schema is primarily a morphologically-based account of the facts of noun classification. This view is motivated by the observation that in Chichewa and many other Bantu languages, the noun prefixes largely have derivational and inflectional functions (cf. Givón, 1972; Mufwene, 1980; Ferrari, 2005). I (2003: 10) claim that taking the prefix as the basis for nominal classification implies that Bantu languages have a lexicon that is purely derived, i.e. without non-derived or basic nouns. He argues that this cannot be the case considering that there is also a bulk of non-derived nouns in Chichewa which are unfortunately typically described as having lost their prefixes over time and become null-prefix forms.[24] He points out a number of challenges for the NCP approach. Firstly, he observes that even though NC1 is associated with the prefix *m(u)*- there are other productive derivational affixes that derive nouns for this class, namely *ka*- and *na*-. Therefore, NC1 alone is composed of nouns with four different morphological characteristics i.e. those made up of prefixes *m(u)*-, *ka*-, *na*- and additionally also the so-called *null prefix no*uns, as summarised in (15).

(15) **Morphological characteristics of NC1 in Chichewa** (Msaka 2003:12)

(a)	m(u)-	mu-nthu (PREFIX-being -> "person")
(b)	ka-	ka-soze (PREFIX-scout -> "scout, spy")
(c)	na-	na-ma-lira (PREFIX-PREFIX-cry -> "bereaved person")
(d)	Ø	bakha ("duck")

[22] My evaluation of Amidu's (1997) framework is based on empirical evidence from Chichewa since he claims that his analysis can be applied to other Bantu languages. The discussion offered here should not be read as disputing Amidu's proposed analysis for Kiswahili in its entirety.

[23] Also see Dingemanse's (2008) review of Amidu's ideas in another related work.

[24] See also, for example, Ferrari (2005: 12–13); Lieber (2004: 9) and Givón (1971:33), regarding cross-linguistic observations of unproductive lexical forms.

Stellenbosch University https://scholar.sun.ac.za

The prefixes *ka-* and *na-* shown in (15) are also as productive and systematic as the prefix *m(u)-* (see Chapter 4, Section 4.3).

Secondly, as discussed in Section 2.4.1 (see examples (8) and (9)), I (2003), observe that although NC1 is standardly identified by prefix *m(u)-*, this morpheme is also homophonous with NC3. The traditionally accepted distinguishing feature is that the *m(u)-* for NC1 refers to humans; hence NC1 is described as a human class while NC3 refers to inanimates, trees, plants, etc. Thirdly, I also observe that NC1 contains a lot of other nouns that do not refer to humans, implying that this description could be narrow and misleading.

To deal with this and many other problems, I (2003) propose a noun class system that is primarily based on non-derived nouns (the so-called *null prefix nouns*). I observe that by focussing on non-derived nouns only, a different pattern of nominal classification system emerges. I (2003: 21–22) claim that NC1 and NC5 in Chichewa are actually distinguished on a semantic basis, where NC1 is animate and NC5 is inanimate. This claim is supported by the observation that homophonous nouns oscillate between the two classes, as illustrated in (16). For instance, in (16a), when the noun *bowo* refers to an animal ("roan"), it belongs to NC1 and when it refers to an inanimate object ("hole") it belongs to NC5. This also applies to all the other examples (16b-f).

(16) Homophonous nouns for NC1 and NC5 (Msaka, 2003: 21–22).

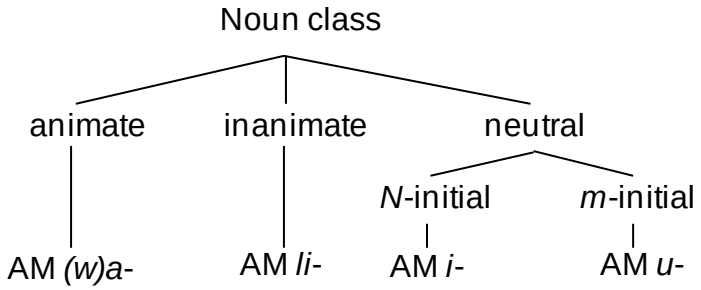
	Noun	**NC1 meaning**	**NC5 meaning**
a.	bowo	"roan"	"hole"
b.	tambala	"rooster/cock"	"coin"
c.	gompho	"an underground insect"	"very short pair of trousers"
d.	fulu	"tortoise"	"a big tall handless gourd for storing water"
e.	godi	"porcupine"	"holes in a board game called *bawo*"
f.	gaka	"pangolin"	"scaly skin found on crocodile skin"

In this regard, I claim that this could be a clue to the fact that NC1 and NC5 are at the opposite poles of animacy. This yields two classes on the basis of semantic features. The other nouns are, however, seen to belong to their classes on a formal basis, mainly their *syllable-initial characteristics* rather than morphologically derived prefixes. These are labelled *neutral* in (17) below. For example, NC9 includes nouns with the following characteristics (i) syllabic nasal **N**, (ii) homorganic nasal cluster initial or (iii) nasalised vowel initial, whereas NC3 takes nouns that start with syllabic nasal ***m***. This yields what I (2003)

Stellenbosch University https://scholar.sun.ac.za

describe as a cognitively-based noun class system of strictly non-derived nouns in Chichewa, which is summarised in (17).

(17) **The non-derived noun class system in Chichewa**



I (2003) argue that having identified a cognitively-based noun class system comprising only the non-derived nouns, it becomes possible to deal with the derived lexical items. This is thus the inverse of the traditional approach. The derived nouns are ultimately also allotted to the noun class system in (17), either on a semantic or a formal basis. For example, nouns derived by the prefixes *m(u)-, ka-, na-*, are assigned to the so-called animate class, whereas the inanimate nouns formed by *m-* go to the *m*-initial class on formal basis.

Not all derived nouns fit into this noun class system, however. For example, nouns derived by the prefix ***chi-*** do not fit into any of the outlined classes in (17). I (2003) state that in the traditional schema there are sets of nouns that are consistently regular in triggering agreement forms, for example those which have been known as NC12/NC13, NC15, NC16, NC17, NC18 and all the plural classes. Adopting Corbett's (1991, chap. 6) distinction between *noun* and *agreement classes*, I claim that these classes are not *noun classes* per se but *agreement classes*. The full set of the proposed agreement classes is as given in (18).

(18) **The agreement classes (Msaka, 2003: 31)**

a. Diminutive class – AMs *ka-*, *ti-*
b. Augmentative class – AM *ch(i)-*
c. Gerundive class – AM *ku-*
d. Locative class – AMs, *pa-; ku-;* and *mu-.*
e. Plural NC2, NC4, NC6, NC12 – AMs, *a-; mi-* and *a-*

One traditional noun class has not yet been addressed: the traditional NC7. According to my (2003: 34) analysis, the non-derived nouns with initial syllable *chi-* could have been assigned to the inanimate class, but are attracted to the augmentative agreement class simply because their AM is homophonous with the augmentative agreement class.

Stellenbosch University https://scholar.sun.ac.za

Although I (2003) do not properly account for the relationships between semantic and phonological regularities, between derived and non-derived nouns, there does seem to be adequate empirical evidence for the broad conceptualization of the noun class system in Chichewa as proposed in that work. The distinction between inherent gender features (involving highly idiosyncratic nouns) and other grammatically acquired features (involving morphologically productive expressions) is key here. Some studies have noted the distinction between the less productive nouns and the highly productive ones (see i.a. Givón, 1972; Carstens, 1991, 2008), with others omitting the productive type from the nominal class system (see i.a. Corbett & Mtenje, 1987). However, this distinction has not been considered seriously in the existing literature, and lack of care about this distinction has led to some serious misconceptions about the Bantu noun class system, as I (2003) have shown. The evidence clearly shows that the Bantu noun class system indeed incorporates two distinct subsystems, that is, noun classes that are based on inherent properties of the lexical items and another group of both nominal and non-nominal expressions that control agreement on verbs and modifiers. The consequence of this has been overemphasis on the overtly grammatical properties rather than the inherent noun class (gender) facts (cf. Carstens, 1997: 370).

There are two challenge to my (2003) analysis, which the present work would like to address systematically, namely (i) it does not explain in any principled way how the semantic and word-initial phonological criteria interact, and (ii) like the Bleek-Meinhof-oriented studies, it was not based on a properly sampled dataset, for example the agreement classes proposed exclude several other expression types e.g. manner nominals, CPs, prepositions, etc. These issues are dealt with in more detail in Chapters 3, 4 and 5.

What follows immediately is a summary of the major issues raised in the present Chapter.

2.6 A summary of the main problematic issues facing the analysis of Bantu noun classes

In the preceding discussion, I have outlined key facts and views regarding the phenomenon of nominal classification in Bantu. I have shown that the Bleek-Meinhof framework has been assumed as the standard analysis of the relevant facts. However, the literature on the Bantu noun class system is characterised by extreme lack of consensus on various points. This controversy reflects the fact that this phenomenon is still not properly understood (cf.

Stellenbosch University https://scholar.sun.ac.za

Mchombo, 2004: 3; Katamba, 2006: 119). Below is a summary of the main problems that have been pointed out in the course of the discussion.

The first problem relates to the methodological underpinnings of the traditional linguistic inquiry into Bantu noun classification, the problem described as *casual exampling* in Worsley (1954: 286). Although this issue has not come out strongly in the reviewed literature except insofar as it was mentioned in passing by Worsley (1954: 286); Amidu (1997: 66–74); Matiki (2001: 70), and Katamba (2006: 118–120), this problem pervades the Bleek-Meinhof system. I argue that some generalisations that are characteristic of the traditional Bantu noun classification system are based on partial data. For example, the claims that (i) the prefix is the major criterion in all the noun classes, (ii) NC1 contains nouns referring to humans[25] with *m(u)*- being the sole prefix, (iii) NC3 is for plants, (iv) NC14 is for abstract nouns, among others seem not to be supported by either qualitative or quantitative empirical evidence emerging from Chichewa. Following Yang (2016), I argue that contentious generalisations such as those proposed to underlie the Chichewa noun class system can only be supported or challenged by a proper statistical metric. I therefore apply appropriate statistical metrics to various generalisations in Chapters 3, 4 and 5.

In addition to the problem of casual exampling, the studies have been compromised by the practice of employing diachronic approaches to account for the synchronic grammar of modern-day Bantu languages. As has been strongly argued by Amidu (1997: 15, 200), diachronic evidence plays its own specific role in linguistic inquiry but it cannot be the preoccupation of a grammarian who is attempting to understand how the synchronic grammar of a particular language could be described so that we can also understand how it can be acquired by its speakers. The appeal to diachronic traces of the role of the NCP does not account for the current grammar of the various Bantu languages.[26] The fact that children are able to acquire the noun class system generation after generation is evidence that the system is not in the kind of chaotic state that is presented in the reviewed literature. It appears there is a mismatch between what linguists perceive and what children make out of the same data: whereas linguists see chaos and disorder, children perceive systematic

[25] A view stated in many works (see i.a. Corbett & Mtenje, 1987: 4–5, footnote 1; Maho, 1999: 15; Katamba, 2006: 113, 115; Kramer, 2015: 252)

[26] Further to this, the role of suffixes is also ignored even though they appear to serve a derivational role in some languages such as Chichewa.

Stellenbosch University https://scholar.sun.ac.za

regularity. In the present study, I attempt to describe the type of evidence that children might pay attention to in the process of acquiring the Bantu noun class systems. In Chapters 3, 4 & 5, I demonstrate how a synchronic approach reveals a different type of noun classification in Chichewa and possibly in other related languages also.

The third issue relates to what Maho (1999: 10) refers to as the *factor of standardisation*. Considering the observed diversity of Bantu noun class systems, it is puzzling that the views about the Bantu noun class system are expressed as if it is one homogenous phenomenon.[27] While the comparative analyses provide important insights into the understanding of Bantu languages, the studies in Bantu noun classes have also overemphasised the role of this approach. I argue that some important language-specific properties are sometimes glossed over in the interest of creating a coherent comparative picture (see also Amidu, 1997: 15). For example, the assumption that NC1 is marked by a single prefix, *m(u)*- and that it is a human class is not supported by the empirical evidence in Chichewa; perhaps however, this observation could hold in other languages. In Chapter 4, I present a detailed analysis of the various classes in Chichewa that seem to counter the comparative analysis.

The fourth problem specifically relates to Amidu's approach which rejects the empirical and theoretical evidence generated by the traditional approach in its entirety. It is indisputable that the traditional analyses have amassed descriptive evidence that any proper account of the Bantu noun class system cannot ignore. For example, the consistencies relating to agreement marking by the traditional NC1, NC3, NC5, NC7 and NC9 do suggest the existence of noun groups. Such traces of regularity in the traditional analysis are characteristic of a system that is not yet fully understood. As Palmer & Woodman (2000: 226) observes, several proposals found in the previous literature provide at least a partial account of the phenomenon and the question is whether these can be subsumed by a more principled account. Therefore, Amidu's (1997) complete rejection of such evidence is not persuasive.

In the next section, I summarise and conclude the present chapter.

---

[27] See also Bokamba (1993: 228) on some important differences between the nature of the noun class systems in certain Bantu languages, for example Kiswahili and Lingala.

Stellenbosch University https://scholar.sun.ac.za

2.7 Summary and conclusion

I began the present chapter by illustrating a set of simplified but commonly attested facts in the discussion of Chichewa noun class studies and Bantu languages in general. These selected facts show that in these languages, some noun groups are associated with a particular type of prefix. The prefix and the noun were also shown to have a grammatical effect when in construction with adjectives, numerals, demonstratives, verbs, etc. This apparent relationship with affected words is realised by means of identical or non-identical affixes on the related words. These grammatical relationships were shown to be the basis for treating the phenomenon as involving a noun class system. Following a long-established tradition within Bantu language studies, these sets of nouns were identified with serial numbers, which then gave rise to the traditional analysis, namely the *Bleek-Meinhof system*. However, even with only a simplified relationship between the noun and its dependents in play, the classification system was shown to be based on some intricate morpho-phonological, syntactic and semantic principles. All these principles have only proven to apply to subsets of data. Consequently, there is no consensus in the literature about the precise principles underlying the Bantu noun class system.

Despite all the challenges that the traditional analysis faces, I have only encountered two analyses that significantly depart from this framework namely, Amidu (1980, 1997) and my own (2003) work. On the one hand, Amidu's work rejects the traditional analysis in its entirety and instead proposes an analysis that I can only comprehend as "syntactically-based" where no single word controls the agreement, with a broader principle at a syntactic level applying instead. This approach was also found to fail in relation to matters such as the most basic facts of how a change in the noun alone triggers change in the agreement of the associated words but not vice versa. My (2003) analysis, on the other hand, reimagines the Chichewa noun class system by focusing exclusively on non-derived nouns. However, my (2003) analysis is only sketchy and is also not based on a properly sampled dataset. The issues explored in this chapter set the scene for the next chapter where I discuss both qualitative and quantitative evidence in favour of a novel conceptualisation, one building on my (2003) initial proposals of the noun class system in Chichewa. This evidence will show the need for a new approach to the phenomenon.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 3
Data collection methods and approaches to noun classification

3.1 Introduction

The goals of the present chapter are twofold: first is to outline the data collection methods and the classification approach adopted in the present study, and second is to introduce the dataset and the morphological characteristics of the nouns that make up this dataset. To achieve these objectives, I have divided the chapter into three major sections. The first is Section 3.2, where I discuss the methodological underpinnings and the dataset in the following order: in Subsection 3.2.1, I provide the background to methodological approaches to collecting data for noun classification, followed by Subsection 3.2.2, where I introduce the Chichewa nominal dataset.; finally, in Subsection 3.2.3, I introduce the *agreement-based noun classification approach* that I have adopted in the present study. This approach achieves twelve *agreement classes* (AC) in contrast to the seventeen noun classes identified by studies taking the Bleek-Meinhof NCP-based approach. Section 3.2 concludes with Subsection 3.2.4, where I discuss the sample size and the statistical distribution of the Chichewa nominal dataset in each proposed AC. The second part of the Chapter is Section 3.3, where I discuss the morphological characteristics of the nouns in the dataset. Here, I argue that, contrary to the traditional assumption, the nouns that comprise the Chichewa nominal dataset are morphologically diverse. I show that Chichewa nouns come in at least four different morphological types. Crucially, some of them are derived without prefixes, while other ACs are derived by more than one type of noun prefix. In the third part of chapter, Section 3.4, I introduce how the various morphological elements of the nominal expressions will be understood in the present study. I conclude the chapter in Section 3.4.

3.2 Methodological considerations and the Chichewa nominal dataset

3.2.1 Background

In the previous chapter, I raised several problems relating to the methodological underpinnings of previous analyses of Bantu noun class systems. Most important among these is the observation that many previous analyses of the Bantu noun class system were speculative, vague and based on partial data samples (see also Worsley, 1954: 286–287; Matiki, 2001: 71; Katamba, 2006: 111–119). This has led to several conflicting views

Stellenbosch University https://scholar.sun.ac.za

regarding certain empirical facts and their subsequent conceptualisation. The prevalence of prefixes *ka-* and *na-* is a case in point. Givón (1971a: 35) and Maho (1999: 75, 252–253) treat these forms as traces, a relic of the old system whereas I (2003: 12) report that the forms are as productive and prevalent as the other traditionally recognised prefixes in Chichewa. Another conflicting observation relates to the status of derived and non-derived nouns in at least some Bantu languages. For example, Givón (1971a:35) claims that the number of non-derived or original NC1/2 [human] nouns is extremely small in most Bantu languages, "seldom exceeding ten or twenty" whereas I (2003: 19–20) argue that non-derived word forms are prevalent in Chichewa, especially in NC1 and NC5. As Worsley (1954: 286) observed, conflicting views of this kind can be resolved with a properly designed quantitative analysis of empirical evidence and a proper quantitative evaluation of proposed linguistic generalisations (cf. Yang, 2016: 75, 106). Further to that, I would like to argue that a classification task requires a representative sample from the lexical population whose characteristics can be generalised to the entire population. Such a task cannot be informed by a dataset that is generated by introspection or by casual exampling. The major objective of this chapter is therefore to set out how the present research project satisfies this empirical requirement.

However, it should be noted that the present study will not be the first to carry out a quantitative analysis of the Bantu noun class system. Such analyses of Bantu noun classes have been carried out before. For example, Worsley's (1954) own study incorporates a quantitative analysis of nominal expressions in a language called Hehe, Zone G, Group 60. Unfortunately, however, the challenge with this study also relates to the data sampling procedure. Worsley (1954: 286–287) acknowledges that his dataset of 500 words, taken from his personal field notes, was also not carefully sampled. This study does not, therefore, resolve the problems identified by the present study. For example, Worsley fits his data in the traditionally identified prefix system and does not consider the other prefixes that I have argued are active in Chichewa. A summary of Worsley's quantitative representation of the NCs is given in Table 3.1, below. Take note that Worsley (1954) uses the singular-plural NCP mnemonic pairs to identify the various NCs in Hehe. Since the corresponding traditional Bleek-Meinhof NCs were not indicated in Worsley's study, those in Table 3.1 have been suggested based on the descriptive properties mentioned in Worsley's (1954: 286–287) discussion.

Stellenbosch University https://scholar.sun.ac.za

Table 3.1: A quantitative analysis of noun classes in Hehe (Worsley, 1954: 286)

Associated NCs	Worsley's class	Criteria	Criteria correlation percentage	Total attested nouns per NC
NC1/2	*umu-/ava-*	prefixes *umu-/ava-* & personal ref	100% (94/94)	94
NC3/4	*umu-/imi-*	no definite content (names of months, the moon, time)	Not given	Not given
NC5/6	*ili-/ama-*	body parts other	30% (21/70)	70
NC7/8	*ugu-/ama-*	augmentatives	66% (4/14)	6
NC9/10	*iyi-/itsi-*	N-class, animals	50% (67/133)	133
NC12/13	*aka-/utu-*	diminutive	100% (14/14)	14
NC14	*uwu-/*	abstract nouns body parts other	71.5% (31/41) 2% (1/41)	41
?	*ulu-*	concords *ulu-/itsi-* man made objects body parts abstract nouns	 33.5% (22/66) 10.5% (7/66) 19.5% (13/66)	 66
?	*iki-/ifi-*	body parts	10	Not given
	Other categories e.g. natural phenomena, buildings, foods, nouns relating to horticultural activities etc. are scattered all over the noun class system			

As shown in Table 3.1, in addition to an unsystematic sampling procedure, Worsley's (1954) study does not provide a representative statistical distribution of the various criteria in their respective NCs, where some figures are not given. Besides, Worsley adopts the Bleek-Meinhof system which does not resolve the very problem that he rightfully points out.

The second work that has also employed a quantitative approach is Contini-Morava (1997). This study was based on a computerized database of 4650 nouns from Kiswahili which were subcategorized according to a combination of morphological and semantic criteria (Contini-Morava, 1997: 604–605). The objective of the study was to identify the semantic attributes of each noun class in the Bleek-Meinhof schema. Since the data was strictly organised in terms of the Bleek-Meinhof system, Contini-Morava's study does not, however, help to resolve the empirical problems this classification system gives rise to. For example, while maintaining the NCP-based noun class system of the Bleek-Meinhof framework, the study concludes that the Kiswahili NC system is semantically motivated. Contini-Morava proposes an intricate semantic network for each NC, see for example Appendix (1A) and (1B) showing

Stellenbosch University https://scholar.sun.ac.za

semantic network for NC3 and NC7, respectively (see also Maho, 1999: 65). However, I will show that the quantitative approach adopted in the present study does not confirm similar findings for Chichewa.

The third work known to employ a quantitative approach is Palmer & Woodman (2000). This study was based on a corpus of 941 nouns from the *Standard Shona dictionary* (Palmer & Woodman, 2000: 233). One problem with this study was that it only focused on NC3 nouns, which is one of the most homogenous NCs in Bantu, since this class comprises mainly syllabic nasal *m*-initial nouns. However, Palmer & Woodman undermine the role of the phonological regularity and instead claim that Shona NC3 grammaticizes and lexicalizes five central cultural scenarios as shown in Table 3.2.

Table 3.2: Palmer & Woodman's (2000: 229) proposed grammaticized semantic scenarios for Shona NC3

a. The spirits of ancestral chiefs bring rain, thunder, and lightning
b. People pray to the ancestors
c. Grain is pounded daily with a mortar and pestle
d. Doctors cure with herbal medicines that are ground in a mortar and pestle
e. Trees, shrubs, and herbs are associated with coolness, moisture, and medicine

However, Palmer & Woodman do not show any quantitative distribution of these categories to determine their productivity. They only acknowledge that there are several nouns that satisfy multiple constraints and some that fail to be explained both culturally and linguistically (Palmer & Woodman, 2000: 243). They (2000:242) conclude that "[o]ne of the reasons that Bantu noun classes resist elegant analyses is that many terms satisfy multiple semantic constraints." In this regard, Palmer & Woodman's quantitatively-oriented study does not also offer solutions to the problems facing the Bleek-Meinhof system (cf. Dingemanse (2006: 9).

The fourth study that takes a quantitative approach is that of Ferrari (2005: 171–175), who strongly motivates her quantitative and qualitative analysis of all Luganda noun classes. The study was based on two electronic dictionary databases, one of approximately 6251 tokens and the other of 1080 tokens. As with the other works, the nouns were simply grouped according to the Bleek-Meinhof NCP-based system. The quantitative aspect of Ferrari's study partly wanted to achieve the following objectives:

Stellenbosch University https://scholar.sun.ac.za

(1) a. to verify whether the Luganda's NC system is semantically determined or arbitrary
b. to determine the spread of nouns across the noun classes
c. to consider which noun classes are productive
d. to identify NCP's role in number marking
e. to identify the number of loan words across the 19 NCs

Ferrari's quantitative analysis is summarised in Table 3.3.[1]

Table 3.3: A quantitative representation of nouns in the Luganda NCs

NC	NCP	NUMBER	%	LOAN	SEMANTICS
NC1	mu-	SG	10	loans ≤ 2.5%	+human (100%)
NC1a	ø	SG	10.7	loans	+human, deity, title, miscellaneous (50%)
NC2	ba-	PL	0	-	+human
NC3	mu-	SG	8.9	loans ≈ 8%	plants, long-round objects, loans, miscellaneous (20%)
NC4	mi-	PL	0.4	-	plural of NC3
NC5	li-/0	SG	7.5	loans ≈ 6.3%	miscellaneous and loans
NC6	ma-	PL	0	-	plural of NC5
NC6a	ma-	invariable	2.8	no loans ≥1%	mass, abstract, liquids, miscellaneous (25%)
NC7	ki-	SG	15	no loans ≥2.5%	concrete objects/instruments, miscellaneous (25%)
NC8	bi-	PL	1.1	-	plural of NC7
NC9	N/0	invariable	20	loans ≈9.1%	animals, miscellaneous and loans
NC10	ø	invariable	1.6	-	plural of NC9 and NC11
NC11	lu-	SG	8.6	no loans	miscellaneous
NC12	ka-	SG	8	no loans	miscellaneous & diminutive
NC13	ga-	PL	0	-	plural of NC12
NC14	bu-	invariable	4.5	no loans	abstract (75%) and miscellaneous
NC15	ku-	invariable	0	-	infinitive nominalizer
NC18	tu-	SG	0	-	count classifiers
NC20	gu-	SG	0	-	augmentative
NC22	ga-	PL	0	-	plural of NC22

Two of the objectives in (1) are of central relevance to the present study, namely (1a) and (1b). With respect to objective (1a), Ferrari (2005: 191) argues that some classes have a semantic core which indicates that the system is semantically organised, despite a

[1] In Table 3.3, some details have been changed from the original table because there were inconsistencies with the discussion. For example, the ≥ (equal or more than) symbol in the NC1 row has been changed to ≤ (equal or less than) based on the discussion offered on page 176.

Stellenbosch University https://scholar.sun.ac.za

widespread occurrence of what she designates miscellaneous nouns in the various classes as shown in Table 3.3. As regards objective (1b), she (2005:192) concludes that there are 8 singular (NCs 1, 1a, 3, 5, 7, 9, 11, 12) and 6 plural (NC2, 4, 6, 8, 10, and 13) productive NCs in Luganda. However, productivity here is based on the quantity of nouns per NC, which undermines NCs that are morphologically productive: such nouns are not lemmatised in a dictionary, meaning that the size of NCs associated with productive processes will necessarily seem (much) smaller than they actually are. The other challenge is that the study does not offer a clear distribution patterns that would explain how the miscellaneous nouns interact with those that are semantically determined. Again, this study does not resolve the problems with the traditional Bleek-Meinhof framework, and the findings are not compatible with the facts in Chichewa.

Despite employing quantitative analyses, the four studies mentioned in the preceding discussion did not resolve most of the problems to be addressed in the present study. For example, none of these works addresses the issue of noun prefixes that have been ignored or deemed non-productive in the traditional literature. One reason for this state of affairs may be that the earlier studies were pursuing different goals, in relation to which it was not necessary to address the problems laid out in the present study. In the present study, I am particularly interested in determining patterns of noun distribution with regard to their morphosyntactic agreement characteristics. Therefore, instead of examining the nature of nouns in the Bleek-Meinhof schema, I group all the nouns according to the type of agreement markers (AMs) they control, irrespective of the morpho-semantic properties of the nouns. As discussed in chapter 2, this type of approach was also followed in Corbett & Mtenje's (1987) work. However, there are two limitations to Corbett & Mtenje's study: (i) it assumed the Bleek-Meinhof system to account for the underlying criteria in the identified agreement classes and (ii) it did not consider the wide-range of data that this study is based on. Therefore, before subjecting the data to a qualitative analysis, I will first conduct a carefully designed quantitative data analysis. The tenets of this design require that I provide a meticulous procedure for data gathering, processing and analysis so that the methods can be replicated for verification elsewhere (Kumar, 2011: 104; Greener, 2012). To ensure this, I will now describe the manner in which the data that form the basis of my proposals have been sampled, collected and processed.

Stellenbosch University https://scholar.sun.ac.za

3.2.2 The Chichewa nominal dataset

The data for the present study was extracted from the electronic version of the *Chichewa monolingual dictionary*. This dictionary is a culmination of an extensive lexicography project carried out from 1972 by the Chichewa Board, and then, from 1996, by the Centre for Language Studies (CLS) (Kamwendo, 1999: 47; Kishindo, 2001: 266). The first printed copy of the dictionary was published in 2000.[2] Subsequently, the dictionary was computerised and uploaded online in 2007 under the auspices of the Norwegian Programme for Development, Research and Education (NUFU).[3]

How does this, then, fit in with the traditional sampling techniques? The kind of data involved in the phenomenon under investigation here is what is referred to as *individual attribute data* as opposed to the so-called *cultural data* (Bernard, 2011: 113). The term *individual attribute data* implies that the investigation is interested in attributes of each element in the population. In my case, I am interested to uncover the attributes of each noun that qualifies it to control a particular AM so that I can draw generalisations regarding these attributes across the entire Chichewa nominal lexicon. According to Bernard (2011: 113), if the goal is to estimate some population parameters from these individual attributes, "then a scientifically drawn, unbiased sample is a must." Social sciences research mainly employs two broad types of sampling methods, namely *probability sampling* and *non-probability sampling* (Plowright, 2012: 38; Picardi & Masick, 2014: 155). On the one hand, probability sampling involves a random selection of cases in the research population. Randomisation aims at getting a representative sample (so that each case has the chance of being selected): such a dataset enables one to generalise the findings to the entire population from which the sample is drawn (Bernard, 2011: 113; Plowright, 2012: 38; Picardi & Masick, 2014: 154–155). With nonprobability sampling, on the other hand, selection of elements is not randomised. For example, a researcher may conveniently gather individuals for a sample on feasibility grounds e.g. location, time, etc. (Picardi & Masick, 2014: 156). Non-probability sampling is appropriate with cultural data, where expert informants are

---

[2] The bibliographic details for this publication are as follows: Centre for Language Studies (CLS). 2000. *Mtanthauzira mawu wa Chinyanja (Chinyanja/Chichewa monolingual dictionary).* Blantyre: Dzuka Publishing Company Ltd.

[3] The electronic version of the dictionary can be accessed via the following link: http://www.unima-cls.org/printLexicon/. See also the summary of the Malawi Lexicon (MaLex) project at http://www.unima-cls.org/Docs/nufu-2009-plakat.pdf and http://unima-cls.org/pmwiki/pmwiki.php/Main/Malex.

Stellenbosch University https://scholar.sun.ac.za

conveniently selected; with such a non-representative sample, the results need not to be generalised to the entire population (Bernard, 2011: 113, 143). Most previous studies focussing on Chichewa noun classes either do not state how the data was collected or base their analyses on evidence gathered by introspection. While the introspection approach works for certain linguistic phenomena, classification (and rule generalisation) requires a well sampled dataset so that attributes of individual nouns or sets of nouns can be generalised to the entire nominal inventory of Chichewa (cf. Yang, 2016: 75); hence the need for the probability sampling approach.

Although the sampling technique adopted in this study does not squarely fall into the probability sampling category, the sample is still taken to be representative. The sampling technique is better understood in *corpus linguistic* terms. According to McEnery & Hardie (2012: 6), in corpus linguistics, there are two broad approaches to data gathering, namely **monitor corpus** and **balanced** or **sample corpus** approaches (cf. Gries, 2009: 10–11). In the monitor corpus approach, the dataset grows over time, whereas in a balanced corpus, data is carefully sampled "as it exists at a given point in time [and] is constructed according to a specific sampling frame" (McEnery & Hardie, 2012: 6). Since the corpus collection for the *Chichewa monolingual dictionary* started in the early 1970s and has been growing over time (see Kamwendo, 1999: 47; Kishindo, 2001: 266), it can be considered to reflect the monitor corpus approach. According to McEnery & Hardie (2012: 6–7), such a corpus eventually becomes highly representative.

In preparing data for the present study, nouns were extracted from the MySQL database using the phpMyAdmin database management toolkit.[4] The MySQL search query was specified to select all entries associated with the following attributes: part of speech as noun, noun class, definition, plural form and variant spellings. This search query excludes all non-nominal entries such as verbs, adjectives, ideophones, etc. The extracted nouns were then exported to an Ms Excel spread sheet, where each of the attributes were represented in separate columns. The spread sheet was then exported to the data cleaning software, OpenRefine. The data cleaning processes that were performed included removing duplicated entries, correcting spellings, reclassifying incorrect entries, etc. Finally, the nouns were grouped according to classes based on the AMs that they control. This gave rise to

[4] For the relevant phpMyAdmin documentation see https://www.phpmyadmin.net/.

Stellenbosch University https://scholar.sun.ac.za

what are designated as *agreement classes* or *target genders* (Corbett & Mtenje, 1987; Corbett, 1991: 147–150). As shown in Chapter 2, Section 2.4.2, the difference between the traditional NCP-based noun class and the proposed agreement classes is that the latter is wide-ranging, in some cases including more than one traditional NC; see Section 3.2.2 below for more details about the agreement class approach (cf. Corbett & Mtenje, 1987: 8). In addition to that, the proposed agreement classes also include expressions that have not been previously associated with any noun class on the Bleek-Meinhof schema. Then basic statistical operations were performed on the data, which included generating frequencies of nouns in various relevant categories, such as agreement class, traditional NC, prefix and suffix types, deverbatives, etc.

There were three main problems that arose in using the Chichewa dictionary database in this study. The first one is that nouns in this dictionary were classified based on the traditional assumptions of noun classification in Chichewa. There are two traditional assumptions that have caused classification problems in the dictionary. Firstly, the assumption that the prefix determines the noun class has led to incorrect classification where the noun with or without a particular prefix triggers AMs of other classes not associated prototypically with that class. For example, as I will discuss in Chapter 4, Section 4.6.4, there are several nouns that were allocated to NC14 instead of NC1, possibly due to lack of the NC1 prefix, or accidental homophony of AMs in some domains or confusion with the plural forms. Consider some of these nouns in (2).

(2) a. kondaine (“love potion”)
b. kholowa (“sweet potato leaf”)
c. singano (“needle”)

The nouns in (2) control the same AM as the other nouns in NC1; however, these were allocated to NC14 in the dictionary.

Secondly, the assumption that singular noun classes are closely connected with their corresponding plural classes (see i.a. Corbett & Mtenje, 1987; Carstens, 1991, 2008; Taraldsen, 2010: 1524) has been overgeneralized. Specifically, in the Chichewa dictionary, noun classes are identified by the traditional mnemonic singular-plural pairings as was shown in Table 2.2 of Chapter 2. However, the singular plural pairing is not as straightforward as is assumed in the classification system used in the *Chichewa monolingual dictionary* and also in some previous studies (cf. Mugane, 1997: 23; Maho, 1999: 151–190;

Stellenbosch University https://scholar.sun.ac.za

Taraldsen, 2010: 1524). For this reason, many nouns which were not supposed to be in some singular classes were assigned to such classes on the basis of their assumed plural counterparts. For example, many *ma*-initial nouns were classified as belong to NC5/6 (*Li-Ma*) because the prefix *ma-* is associated with plural nouns for NC5, even though the nouns do not control AMs for NC5. Consider some examples in (3), see also Chapter 4, Sections 4.5 and 4.6.4.

(3) a. mafuta (“fat, oil, fuel”)
b. madzulo (“evening”)
c. machira (“stretcher, hammock”)

The nouns in (3) do not have singular counterparts as they are non-count or mass nouns. However, these nouns were classified as belonging to the noun class pair 5/6 (*Li-Ma*), but in terms of agreement evidence, they should have been treated simply as NC6 nouns. It is worth noting that in some studies, nouns of this type are treated as forming a subclass, NC6a (cf. Fortune, 1970: 97; Welmers, 1971: 13; Hyman, 1980: 180; Maho, 1999: 249; Ferrari, 2005: 174).

Besides, as I will show in Chapter 4, there are several nouns in Chichewa that do not have singular-plural distinction. Therefore, it is not necessarily correct to associate every noun with singular-plural pairs. A similar observation has been made in seven other Bantu languages of chiShona, isiXhosa, isiZulu, Kikuyu, Kinyarwanda, Luganda and Runyankore (see Byamugisha *et al.*, 2018). These were some of the problems that necessitated another level of careful revision of various entries in the dataset. Due to this, many nouns have been reclassified. These discrepancies will be dealt with in detail in their rightful sections in Chapter 4. The reclassified nouns have also been included in the appendices section of this dissertation.

The third problem is that the dictionary only captures lexicalised forms; however, there are many predictable and productive word formation processes whose output forms are not lemmatised in the dictionary. It is therefore difficult to provide statistics for these words; suffice to say they are productive in the sense that they are derived by regular rules. In the present study, I treat both the lexicalised and the non-lexicalised, productive forms with equal importance. I will also highlight the various productive grammatical processes in the relevant sections of the present chapter and subsequent chapters.

Stellenbosch University https://scholar.sun.ac.za

In the next section, I introduce the set of agreement classes that will be used to determine the noun class system of Chichewa.

3.2.3 Agreement classes in Chichewa

As shown in Chapter 2, there are discrepancies in terms of the number of noun classes derived by the *NCP-based criterion* and the *agreement class-based criterion* of noun classification in Chichewa (see i.a. Corbett & Mtenje, 1987). It is also crosslinguistically observed that there are usually 'mismatches' between overt class markers on the nouns and the AMs on dependent words and predicates (Corbett, 1991: 150–160; Evans, 1997: 106; Aikhenvald, 2004: 1033; Gagliardi & Lidz, 2014: 59). Besides, as stated in Chapter 1, Section 1.2.2, the noun-external information (i.e. AMs) is more predictable than the noun-internal information (i.e. NCP, semantic, etc.). Therefore, following Gagliardi (2012: 27), I suggest that it would be more insightful to consider all the possible AMs in Chichewa first, before considering the distribution of the noun-internal information. This approach also makes acquisition sense as it is generally accepted that children are very good with (regular) inflection (see Yang, 2016: 31–40 for an overview). With the AM facts in place, I will further identify the common properties of all the nouns that control specific AMs. As already introduced above, I will refer to this approach as an *agreement-based classification* and the classes obtained as *agreement classes* (ACs).[5] In this regard, I adopt Corbett's (1991: 148) broad characterization which states that "nouns are in the same agreement class provided that, given the same conditions, they will take the same agreement form." The agreement class approach to noun classification has non-Bantu precedents in the works of Schenker (1955), Zaliznjak (1964) and many others (see Corbett, 1991: 147–150). It has also been emphasised in some studies on Bantu languages (see i.a. Watkins, 1937: 22–23; Corbett & Mtenje, 1987; Dimitriadis, 1997: 1). However, due to the common adoption of the Bleek-Meinhof system, the discrepancies between the agreement-based and the NCP-based approaches in Bantu noun classification systems have not been resolved yet, prompting the diachronic "distortion" claims discussed in the previous chapter.

As I will show in Chapter 4, the agreement-based classification approach has various advantages over the traditional NCP-based approach. For example, it can account for a

[5] Take note that the original Bantu noun classification by Brusciotto around 1659 was also based on the AMs and not on noun prefixes.

Stellenbosch University https://scholar.sun.ac.za

wider range of nouns in the dataset than the latter. In this regard, my discussion of the data is not based on the traditional assumptions of noun classification even though there are many coincidences between the two approaches. From this point onwards, I will only mention the respective Bleek-Meinhof numbering system in parentheses to enhance the clarity of my exposition. As such, my reference to the traditional noun class numbering is only for expository purposes and should not be ascribed analytical significance. Take note also that some ACs may include more than one traditional NC; this should not cause any confusion as this simply signifies that the relevant traditional NCs control the same AM. In the present study, each AC will be identified by the AM that is associated with it. Later, in Chapter 5, the same nomenclature is also used to identify the noun class system that emerges from the broader ACs. This nomenclature is adopted because it identifies the classes accurately with their relevant AMs unlike the arbitrary numbering system used in the Bleek-Meinhof system.

All the nouns in the dataset and the identified predictable expressions not lemmatised in the dictionary are observed to trigger one of the twelve AM sets shown in Table 3.4 below.

Table 3.4: Agreement-based classes in Chichewa

	AC	Traditional NC	Verb SM	Verb OM	Numeral AM	Relative AM	Possessive AM	Adj AM
1	*m(u)-a-*AC	≈NC1	*a-*	*m(u)-*	*m(u)-a-*	*a-*	*(w)a-*	*(w)a-*
2	*a-*AC	≈NC2 & NC6	*a-*	*a-*	*a*	*a-*	*a-*	*a-*
3	*u-*AC	≈NC3 & NC14	*u-*	*u-*	*u-*	*u-*	*u-*	*u-*
4	*i-*AC	≈NC4 & NC9	*i-*	*i-*	*i-*	*i-*	*i-*	*i-*
5	*li-*AC	≈NC5	*l(i)-*	*li-*	*l(i)-*	*li-*	*l-*	*l-*
6	*chi-*AC	≈NC7	*ch-*	*ch-*	*ch-*	*chi-*	*ch-*	*ch-*
7	*zi-*AC	≈NC8 & NC10	*zi-*	*zi-*	*zi-*	*zi-*	*zi-*	*zi-*
8	*ka-*AC	≈NC12	*ka-*	*ka-*	*ka-*	*ka-*	*ka-*	*ka-*
9	*ti-*AC	≈NC13	*ti-*	*ti-*	*ti-*	*ti-*	*ti-*	*ti-*
10	*ku-*AC	≈NC15 & NC16	*ku-*	*ku-*	*ku-*	*ku-*	*ku-*	*ku-*
11	*pa-*AC	≈NC17	*pa-*	*pa-*	*pa-*	*pa-*	*pa-*	*pa-*
12	*mu-*AC	≈NC18	*mu-*	*mu-*	*mu-*	*mu-*	*mu-*	*mu-*

Take note that, although the traditional classification has 17 distinct NCs in Chichewa (see Table 1.1), there are only 12 ACs. As shown in Table 3.4, each AC is identified by its AM. Although some AMs have consonant-vowel combinations, the associated vowels are prone to a wide range of allophonic variations. For simplicity, I will consistently represent the AMs with what appear to be the default vowels except where it is necessary not to. Take note

Stellenbosch University https://scholar.sun.ac.za

also that Table 3.4 only depicts five agreement domains; namely the verb SM, the verb OM, the numeral, relative pronoun and the possessive pronoun. However, this is not an exhaustive list of agreement domains, but it is only a representative sample of many others not included here, such as, demonstratives, ordinal numbers, etc.

Take note further that the *m(u)-a*-AC in row number 1, has two AM forms, *m(u)*- and *a*-. On the one hand, the AM *m(u)*- is observed to occur on the verb's OM and on the numeral slots (cf. Givón, 1971a: 36). The AM *a-,* on the other hand, is found on possessives, adjectives, relative pronouns and in the verb's SM slots. The AMs *m(u)*- and *a*- are observed to be allomorphic; hence they do not motivate realisation of two different classes. This AC will therefore be identified by the shorthand *m(u)-a-.* The phenomenon where more than one AM is used in one AC is referred to as *split agreement* (Heine, 1982: 195; Aikhenvald, 2004: 1039) while the expression-types that control more than one AM are described as *inconsistent/hybrid controllers* (Corbett, 2006: 11, 143). In Chichewa, only the *m(u)-a*-AC is observed to involve split agreement and hybrid controllers; the rest have only one form of AM.

In the next section, I present the statistical distribution of the sampled nouns across the ACs introduced in Table 3.4.

3.2.4 Sample size and statistical distribution of nouns in the agreement classes

The sample size for the present study is 5,828 nouns. This represents approximately 43.5% (5828/13,411) of the entire *Chichewa monolingual dictionary*. These nouns are distributed across the various ACs in the manner shown in Table 3.5 below. The infinity symbol (∞) is used to indicate that the expressions in the respective ACs, or the subset thereof are productive and predictable such that the elements are not lemmatised in the dictionary; hence their frequencies cannot be computed in finite terms.

Stellenbosch University https://scholar.sun.ac.za

Table 3.5: Frequencies of nouns in the Chichewa ACs and traditional NCs

	AC	Associated traditional NC		Number	Percentage
1	*i*-AC	≈NC4	30 (∞)	1453 (∞)	24.9%
		≈NC9	1423		
2	*m(u)-a*-AC	≈NC1	1222	1222(∞)	20.9%
3	*chi*-AC	≈NC7	1075	1075(∞)	18.4%
4	*li*-AC	≈NC5	843	843	14.5%
5	*u*-AC	≈NC3	627	839(∞)	14.3%
		≈NC14	212		
6	*a*-AC	≈NC2	∞	357(∞)	6.1%
		≈NC6	357(∞)		
7	*zi*-AC	≈NC8/NC10	16 (∞)	16 (∞)	0.3%
8	*pa*-AC	≈NC17	13 (∞)	13 (∞)	0.2%
9	*ku*-AC	≈NC15	∞	6 (∞)	0.1%
		≈NC16	6 (∞)		
10	*mu*-AC	≈NC18	2(∞)	2(∞)	0.03
11	*ti*-AC	≈NC13	2 (∞)	2 (∞)	0.03%
12	*ka*-AC	≈NC12	∞	∞	
	TOTAL			**5828**	**100**

The ACs in Table 3.5 have been arranged according to frequency of entries, ranging from the highest to lowest. Take note that the first six rows contain typical nominal expressions, with the most populous AC containing 1453 entries and the least populous containing 357 entries. The next six rows contain morphosyntactically complex and predictable phrases; with entries ranging from 2 to only 16. This observed statistical distribution is significant in two ways: First, the quantitative difference between the entries in the first six ACs and the next set has already been pointed out in Chapter 2, specifically that some classes contain expressions that are derivationally less predictable and those that are obtained by highly regular inflectional or derivational means, such as diminutives, locatives, etc. These highly regular classes do not have stems of their own, but get their lexical items from the less predictable classes and other word categories; hence they are not lemmatised in the dictionary. Consider examples in Table 3.6 where nouns forming the *ka*-AC are derived, in a regular manner, from idiosyncratic nouns in the *a*-AC.

Stellenbosch University https://scholar.sun.ac.za

Table 3.6: The comparison between morphological properties of a- and ka- ACs

	Entries for the *a*-AC	**Entries for the *ka*-AC**
a	mu-nthu PREFIX-N.STEM (“human being”)	ka-munthu DIM-human being (“a small person”)
b	galu (“dog”)	ka-galu DIM-dog (“a small dog”)
c	na-ma-lira prefix-prefix-cry (“bereaved person”)	ka-namalira dim-bereaved person (“a small bereaved person”)

The distinction illustrated in Table 3.6 has been consistently observed in the previous literature, where it has been designated as a distinction between nouns with *primary* and *secondary prefixes* (Guthrie, 1948: 856; Fortune, 1955: 54, 1970: 87; Maho, 1999: 88) (see Chapter 6 for further details).

Second, the statistical distribution is significant in the way we understand the Chichewa NC system. While the productive expressions can easily be accounted for by some simple morpho-syntactic rules, the same case is not true for the less productive elements. It is therefore not surprising that it is the morphologically less predictable nominal expressions that have eluded explanation in traditional Bantu NC studies, which have clearly overgeneralised the pattern of the highly predictable expressions as being key to the entire NC system. Therefore, the figures represented in Table 3.6 are of great significance to the present study especially for identifying sets of lexicalised forms and for testing rule productivity (see Chapters 4 and 5).

In the next section, I introduce the various morphological categories of noun-types that make up the dataset.

3.3 Morphological composition of the Chichewa noun dataset

Before I provide a detailed description of the nominal properties of each AC, one further general clarification should be made about noun classification, noun derivation (nominalization) and noun inflection. The traditional account of the Bantu noun class system assumes a generalised intertwined relationship between noun class and nominal derivation (see i.a. Givón, 1971a; Mufwene, 1980; Ferrari, 2005; Kramer, 2015: 188). With the exception of a few works such as Hlungwani’s (2012) analysis of deverbal noun formation in Xitsonga, this relationship is not been made clear in many analyses of Bantu languages.

Stellenbosch University https://scholar.sun.ac.za

This is especially true of Chichewa where this matter has not received adequate systematic attention.

The claim that nominalisation and noun classification are treated as one grammatical phenomenon in Bantu can be demonstrated in the following ways. Firstly, according to the traditional analysis, the key characteristic of the Bantu NC system is the NCP. The NCP has been viewed as playing both derivational and inflectional functions (cf. Mufwene, 1980; Myers, 1987; Ferrari, 2005). This implies that a typical noun in Bantu must contain a prefix (derivational or inflectional) and a stem. However, as will be shown below, this assumption is not generally compatible with the dataset informing the present study.

Secondly, as already pointed out, NCs are often viewed in pairs of singular and plural. In the *Chichewa monolingual dictionary* and in Chichewa grammar books (i.a. Nankwenya, 1992: 38), the classes are actually recognised by mnemonic pairs of singular and plural prefixes (see again Chapter 2, Table 2.2). The assumption is that each class has clear singular and plural alternations. This assumption is also not supported by the evidence from my dataset.

As will be shown in this section, these assumptions have misguided our understanding of the Chichewa noun classification, and also of nominal derivation/inflection grammatical processes. Firstly, the NCP-based approach has ignored many other derivational processes that produce prefixless, but nevertheless genuine nouns in Chichewa. Secondly, the approach has overgeneralised the singular and plural relationships in the NC system. Thirdly, the approach has not treated the non-derived nouns in Chichewa in any systematic manner.

The nominal dataset informing this study can be effectively characterised in terms of Lieber's (2004) views regarding the lexicon and lexicogenisis. Lieber (2004: 9) classifies lexical items as belonging to two broad categories, namely the *simplex* and the *complex lexicon*. The simplex lexicon comprises words that are not derived by any morphological means, whereas the complex lexicon comprises words that are derived through word formation processes such as derivation, compounding, conversion, etc. These word formation processes are assumed to only extend the simplex lexicon. The dataset I am using in the present study also exhibits these characteristics. As already pointed out in Chapter 1, I follow Ferrari's (2005: 13) views that non-productive word forms are learned via rote memorization (cf.

Stellenbosch University https://scholar.sun.ac.za

Yang, 2016) and thus represented accordingly, and thus **not** as the derivational output of regular morphosyntactic processes. In this regard, the simplex lexicon represents these non-productive word forms. In addition to these nominal expressions, I have also observed that some AMs are controlled by even more complex non-nominal expressions such as CPs, and apparently "simpler" controllers like prepositions; furthermore, there are also cases where the AMs function as dummy or expletive elements (see Chapters 4 & 6). Since this latter type of controller is not represented in the dataset, I will only discuss it from Chapter 4, where I focus on properties of expressions that control the various AMs.

In the next section, I introduce the simplex and complex nominal lexicons as observed in the Chichewa dataset.

3.3.1 The simplex lexicon in the Chichewa dataset

In Chichewa, there are several non-derived nouns whose existence cannot be ascribed to any known productive derivational processes. The examples in (4) illustrate some relevant cases. In these examples and all subsequent examples of this nature, singular and plural forms will be represented to show the various types and functions of noun prefixes.

(4)

	Singular		**Plural**	
a.	bambo ("father")	NC1	a-bambo ("fathers")	NC2
b.	galu ("dog")	NC1	a-galu ("dogs")	NC2
c.	gule ("dance")	NC1	ma-gule ("dances")	NC6
d.	fupa ("bone")	NC5	ma-fupa ("bones")	NC6
e.	uta ("arrow")	NC14	ma-uta ("arrows")	NC6

In (4), the singular forms are not obtained through any known overt morpho-phonological processes in Chichewa; however, the plural nouns have prefixes that apparently mark number. In the traditional literature nouns like the singulars in (4) are described as *null-prefix nouns*, with the problematic assumption being that the prefix must have been lost over time. The alternative view adopted in the present study is that these nouns simply do not have prefixes. My hypothesis is that the morphological form of these nouns is what it seems to be on the surface and that there are therefore no lost prefixes; in other words, child acquirers do not postulate covert counterparts to the assumed prefixes, but instead learn the prefixless nouns as idiosyncratic forms. I further assume that singular nouns of the type shown in (4) form the simplex lexicon. According to the dataset, these nouns are prevalent in the ***m(u)-a-***AC and ***li-***AC, which have been associated with NC1 and NC5, respectively. I provide the relevant statistical distribution of simplex nouns in the dedicated ACs discussion in Chapter

Stellenbosch University https://scholar.sun.ac.za

4. The nouns illustrated in (4) differ morphologically and phonologically from those that are regarded to form the complex lexicon. The complex lexicon is introduced in the next section.

3.3.2 The complex lexicon in the Chichewa dataset

The dataset shows that derived nouns in Chichewa are also morphologically more diverse than originally assumed. In summary, the nouns are observed to fall into the following broad types.

(5) a. Type 1: Nouns derived by covert conversion, tone or final vowel (FV) modification.
b. Type 2: Prefix attached to Type 1 or simplex or other word categories.
c. Type 3: Verbal nouns (infinitival and manner nominals).

Each of these three noun-types will be discussed in the next sub-sections. As introduced in Chapter 1, it is the complex nouns that will be analysed from a DM perspective where it is assumed that words are derived by morpho-syntactic operations that combine roots and functional categories.

3.3.2.1 Type 1: covert conversion, tone or FV modification

Type 1 derived nouns mainly comprise deverbal nouns, i.e. nouns that have been derived from roots associated with verbal expressions. Consider examples in (6) (See also Appendices 2 and 8). In these examples and all subsequent examples, tone will be marked if phonemically significant.

(6)

	Verb stem	**Singular noun**	**Plural noun**
a.	gonth-a ROOT-VFV “be deaf”	gónth-i (NC1) be deaf-NFV “deaf person”	a-gónthi (NC2) PL-deaf person “deaf people”
b.	zunz-a ROOT-VFV “harass/oppress”	zúnz-o (NC5) harass-NFV “crisis/problem”	ma-zúnzo (NC6) PL-problem “problems”
c.	nol-a ROOT-VFV “sharpen”	nol-o (NC9) sharpen-NFV “whetstone”	nolo (NC10) whetstone “whetstones”

In (6), singular nouns are derived from the verb expression by means of tone or FV change. As stated in Chapter 2, Section 2.2, in Chichewa and related Bantu languages, verbs terminate in what can be referred to as *verbalising final vowels* (VFV) *-a* and *-e*. The VFV *-e* is associated with subjunctive mood while the VFV *-a* is assumed to be the default, which

Stellenbosch University https://scholar.sun.ac.za

is associated with all other moods such as indicative, imperative, etc. (see Hyman & Mtenje, 1999: 111; Maho, 1999: 78; Mtenje, 2002: 7, 2007: 35; Mchombo, 2004: 22, 28, 67; Downing & Mtenje, 2017: 69). In (5), the nouns are derived by replacing the VFVs with what can be designated as *nominalising final vowels* (NVFs) *-i* or *-o* with optional tone marking on the derived nouns. As generally observed in previous studies, the NFVs *-i* and *-o* participate in the derivation of some deverbal nouns in many Bantu languages (see i.a. Mchombo, 1978: 107,295, 2004: 113; Kishindo, 1985; Mugane, 1997; van der Spuy, 2009: 6; Contini-Morava & Kilarski, 2013: 270; Mletshe, 2017: 31). In these studies, the NFV *-i* has been described as the *agentive suffix* or *actor-denoting morpheme* which attaches to action verbs.[6] The NFV *-o*, in turn, has been described as the *non-agentive* or *non-actor morpheme*, which, according to Kishindo (1985: 6), derives nouns generally denoting "object of V-ing" – e.g. "object of writing" (cf. Mchombo, 2004: 113).

My dataset confirms these distinctions between the agent and non-agent denotation of the relevant NFVs. However, what has not been considered as a regular nominalisation process in these studies is the use of the NFVs *-i* and *-o* to derive nouns without the aid of the so-called NCPs; this is a matter I will address in Chapter 4, Sections 4.3.1 and 4.4.1.

Importantly in the present context, there are deverbal nouns which are derived without the aid of the NFVs *-i* or *-o*; in other words, they maintain their VFVs, but they may undergo tone modification or covert conversion. Consider the following examples.

(7)	**Verb stem**	**Singular noun**	**Plural noun**
a.	nyeng-a ROOT-VFV "deceive"	nyénga (NC1) "deceiver/mongoose"	a-nyénga (NC2) PL-deciever "deceivers"
b.	lemb-a ROOT-VFV "write"	lémba (NC5) "letter/script"	ma-lémba (NC6) PL-letter "letters/scripts"
c.	ful-a ROOT-VFV "burrow/dig"	fule (NC5) "mound, heap of soil"	ma-fule (NC6) PL- mounds, heaps of soil "small heaps of soil"

[6] According to Comrie & Thompson (2007: 336), the label *agentive nominalisation* does not always mean that the derived noun is necessarily in an 'agentive' relationship with the verb from which it is derived. For example, in English, the deverbal noun *hearer* is from the non-agentive verb *hear.* The same applies in Chichewa.

Stellenbosch University https://scholar.sun.ac.za

Take note that in (7), the singular nouns are derived by means of tone modification or covert conversion. As in (6), the prefixes are only found on plural deverbal nouns, which implies that the prefixes are only functioning as number affixes and not as derivational or noun class prefixes.

I argue that the previous literature on Chichewa and Bantu languages in general has not considered nouns of this type in any systematic manner. Deverbal nouns of this type are also very regular in the *m(u)-a*-AC and *li*-AC (see appendices 2 and 8). The strategies illustrated in (6) and (7) are very productive processes, as we clearly see if we consider the morphological form of some loan words. For example, the English loan word *goal* has taken the following forms.

(8)	gol-a	gol-i	gol-o
	ROOT-VFV	ROOT-NFV (NC1)	ROOT-NFV (NC5)
	"become a goal"[7]	"goal keeper"	"goal post"

In the next section, I introduce nouns that are derived by attaching prefixes to these Type 1 nouns.

3.3.2.2 Type 2: Nouns derived by adding prefixes to Type 1 or other stems

Although Type 1 nouns are prevalent in the *m(u)-a*-AC and *li*-AC, Type 2 nouns are found in at least five ACs, namely *m(u)-a*-AC*; u*-AC*; i*-AC*; li*-AC and *chi*-AC. The Type 2 nouns are introduced briefly in Table 3.7 below. Note that, unlike in Type 1 where there were only deverbal nouns, in Type 2, the prefixes attach to stems from various word categories, e.g. verbs, nouns, adjectives, ideophones, etc. Each example in Table 3.7 is identified by the AM of the AC it belongs to and the associated traditional NC numbers. Prefixes that are not related to number marking will simply be glossed as *derivational prefix* (DPF).

[7] There is no exact English equivalent of this verb. In Chichewa it describes the act of a ball passing the goal post and the subsequent outcome of becoming a goal.

Stellenbosch University https://scholar.sun.ac.za

Table 3.7: A representative sample of Type 2 nouns in Chichewa

	Stem	Prefix	Singular noun & AC		Plural noun & AC	
a.	lond-a ROOT-VFV "watch over"	m-	m-lónda SG-N.STEM "guard"	***m(u)-a-*** (≈NC1)	a-lónda PL-N.STEM "guards"	***a-*** (NC2)
b.	dzuts-a cause to wake up-VFV "cause to wake up"	ka-	ka-dzutsa DPF-N.STEM "breakfast"	***m(u)-a-*** (≈NC1)	a-kadzutsa PL-breakast "breakfasts"	***a-*** (NC2)
c.	ku-bal-a INFIN-give birth-VFV "to give birth"	na-	na-kubala DPF-N.STEM "parent"	***m(u)-a-*** (≈NC1)	a-nakubala PL-parent "parents"	***a-*** (NC2)
d.	khut-a be full-VFV "be satiated"	sa-	sa-khúta NEG-be satiated "glutton"	***m(u)-a-*** (≈NC1)	a-sakhúta PL-glutton "gluttons"	***a-*** (NC2)
e.		ma- wa-/wo- bwa-/tsa	see Section 4.3.6 see Section 4.3.7 see Section 4.3.8	***m(u)-a-*** (≈NC1)		
f.	pit-a ROOT-VFV "go"	m-	m-píta SG-N.STEM "duct/path"	***u-*** (≈NC3)	mi-píta PL-N.STEM "ducts/paths"	***mi-*** (NC4)
g.	fulu free person	u-	u-fúlu DPF-free person "freedom, right"	***u-*** (≈NC14)	ma-ufúlu PL-freedom "types of rights"	***ma-*** (NC6)
h.	kumb-a ROOT-VFV "dig"	l-	li-kúmba DPF-N.STEM "narrow way"	***li-*** (≈NC5)	ma-likúmba PL-narrow way "narrow ways"	***ma-*** (NC6)
i.	ik-a ROOT-VFV "place"	dz-	dz-ík-o SG-put-NVF "world, country"	***li-*** (≈NC5)	ma-íko PL-place "countries"	***ma-*** (NC6)
j.	yabw-a ROOT-VFV "itch"	ch-	chi-yábwe SG-N.STEM "itchy caterpillar"	***ch-*** (≈N7)	zi-yábwe PL- N.STEM "itchy caterpillars"	***z-*** (NC8/10)
k.	kalamba-a ROOT-VFV "be old"	N-	nkhálámba DPF.N.STEM "elderly person"	***i-*** (≈NC9)	nkhálámba DPF.N.STEM "elderly people"	***z-*** (NC8/10)

Note that some prefixes in Table 3.7 are number affixes while others are mere derivational prefixes. For example, the traditional noun prefix *m-* (of NC1 and NC3), *chi-* (of NC7), etc. are observed to attach to the simplex or already derived nominal stems. The role of such prefixes is generally observed to only mark whether the stem is singular or plural. This confirms Carstens' (1991, 2000, 2008) arguments that these prefixes are only number markers in Bantu. As shown above, the Chichewa case provides strong evidence in support

Stellenbosch University https://scholar.sun.ac.za

of Carstens' arguments. In addition to functioning as number markers, the prefixes also appear to be required by what can be described as *dependent noun stems*. In this regard, some noun stems can stand on their own, as shown in Type 1, while others require the number prefix in their singular form.

Note also that while the other ACs are shown to comprise nouns bearing only one type of prefix, the *m(u)-a*-AC (≈NC1) and *li*-AC (≈NC5) are shown to include nouns with more than one type of prefix. This property has not been discussed systematically in the previous literature where only one prefix is identified per noun class. I consider these in more detail in their respective sections in Chapter 4 below.

As shown in Table 3.7, another interesting pattern regarding the ACs and the inflectional/derivational prefixes is that every AM also participates in the derivation/inflection processes of a subset of the nouns in the AC. This could be the potential reason for the over-emphasis on prefixes that also appear as AM in the Bleek-Meinhof-oriented analyses. However, I will show in Chapter 4, that the prefixes that do not coincide with the AMs are equally productive.

Next, I introduce the third noun-type that comprise the complex lexicon in the dataset.

3.3.2.3 Type 3:Verbal nouns

Chichewa has two types of verbal nouns, namely infinitival and manner nominals. On the one hand, infinitival nouns are identified by the infinitival prefix *ku-* which attaches to verbal expressions as shown in (9).

(9)	a.	-dy-a	ku-dya
		ROOT-VFV	INF-eat
		"eat"	"eating/to eat"
	b.	yend-a	ku-yi-yend-er-a
		ROOT-VFV	INF-OM-walk-APP-VFV
		"walk/move"	"supervising it/ getting around by it"

In Bantu languages, infinitival nouns have been identified as forming their own noun class, NC15. This class triggers the AM *ku-.*

Manner nominals, on the other hand, are identified by two different prefixes, namely *ka-* and *ma-* (see i.a. Nankwenya, 1992: 46, 56; Bresnan & Mchombo, 1995: 190; Mchombo, 1998: 515–516). These nouns are derived by attaching the prefix *ka-* or *ma-* to a passivized verb

Stellenbosch University https://scholar.sun.ac.za

complex in the subjunctive mood, as evidenced by the VFV *-e* (cf. Mchombo, 1998: 170, 2004: 115–117). The passive morphemes in Chichewa appear in two forms *-idw-* or *-edw-*, the variations being subject to vowel harmony.[8] The derived nouns roughly read '*the manner/way of verb-ing*' (cf. Mchombo, 2004: 115; Comrie & Thompson, 2007: 339). These nouns are shown in the following examples.

(10)	**Verb stem**	**Passivized verb**	**Manner nominal**
a.	lemb-a	lemb-edw-a	ka-lembedwe /ma-lembedwe
	ROOT-VFV	ROOT-PASS-VFV	DPF-be written / DPF-be written
	"write"	"be written"	"manner of writing"/ "manner of writing"
b.	lim-a	lim-idw-a	ka-limidwe /ma-limidwe
	ROOT-VFV	ROOT-PASS-VFV	DPF- be farmed / DPF-be farmed
	"cultivate"	"be cultivated"	"manner of farming"/ "manner of farming"

Although the manner prefixes (*ka-* and *ma-*) are homophonous with several other affixes, the manner nominal prefixes seem to be distinct from their homophonous counterparts, such as the diminutive *ka-*, and the plural *ma-* introduced in Table 3.7 above (I return to this matter regarding the multiple functions of these prefixes in Chapters 4 and 6). Manner nominals have not been considered systematically in previous studies and as such they do not have their own noun classes in the Bleek-Meinhof system. However, the *ma-* type manner nominals consistently trigger the AM for the *a*-AC, one associated with NC2 or NC6, whereas the *ka*-type consistently triggers the same AM as the *ka*-AC, the one associated with the diminutive NC12. As I will show in Chapters 4, 5 and 6, these prefixes are functionally different although they trigger the same AMs. It is not known why the Bleek-Meinhof system identifies a special class for infinitives (NC15), which already shares its AM with the locative NC17, but not for manner nominals which are equally productive.

3.3.3 Interim summary

The major purpose of the present section, Section 3.3, was to introduce the various noun-types that are found in the Chichewa dataset that informs the present study. I have introduced the three major types of nouns in the dataset, namely (i) nouns that are idiosyncratic/irregular in nature (the simplex lexicon), (ii) nouns whose forms are derived by less regular rules (e.g. some type 1 & 2) and (iii) nouns whose morphological forms are

[8] For more details about vowel harmony in Chichewa see Mtenje (1985) and Mchombo (1993, 1999, 2004: 115).

Stellenbosch University https://scholar.sun.ac.za

highly predictable (e.g. type 3). Contrary to the traditional assumptions about noun classes in Chichewa, I have shown that each AC includes more than one noun-type. The expressions forming the noun class system thus support the modified DM view that I adopted in Chapter 1, Section 1.3.2. In terms of this view, the irregular expressions (simplex lexicon) are a result of cases where no productive rule is observed to account for their pattern and are assumed to be acquired by total lexicalisation (Yang 2016:216). By contrast, the productive types (the complex lexicon) are acquired by rule generalisation.

Having introduced the various morphological types, in the next section, I introduce the morphological status of the various morphemes in the nominal expressions.

3.4 The morphological status of the various components of Chichewa nominal expressions

Given the foregoing, it will be proper to clearly state how the various formal units making up nominal expressions are identified in the present study. First, I will identify simplex nominal expressions as bare nouns which can be turned into complex NPs by various morpho-syntactic processes e.g. prefixation or independent modifiers, as illustrated in (11).

(11)

	Bare Noun	**Singular NP**	**Plural NP**	**Dim NP**	**Modifier NP**
a.	galu	ø-galu	a-galu	ka-galu	galu wa-kuda
	dog	SG-dog	PL-dog	DIM-dog	dog ASC-black
	“dog”	“(one)dog”	“dogs”	“small dog”	“black dog”

In (11), the bare noun is equivalent to the noun stem, implying that the expression is not yet inflected in terms of number, for example. Second, complex nominals introduced in Section 3.3.2 are assumed to minimally involve nominal stems (N.stem) that may be dependent or independent contingent on their morpho-phonological specification. Consider (12).

(12)

	N.Stem	**Singular NP**	**Plural NP**	**Dim NP**	**Modified NP**
a.	-nthu	mu-nthu	a-nthu	ka-munthu	munthu wa-ku-fa
	N.STEM	SG-N.STEM	PL-N.STEM	DIM-person	person ASC-INFIN-die
		“(one) person”	“people”	“small person”	“a dead person”
b.	-tu	mu-tu	mi-tu	ka-mutu	mu-tu wa-ke
	N.STEM	SG-N.STEM	PL-N.STEM	DIM-head	head ASC-POSS
		“(one) head”	“heads”	“small head”	“his/her head”
c.	Malawi	m-malawi	a-malawi	ka-mmalawi	mmalawi m-modzi
	PLACE NAME	SG-PLACE NAME	PL-PLACE NAME	DIM-malawian	Malawian AM-one
	“Malawi”	“Malawian”	“malawians”	“small Malawian”	“one Malawian”

Stellenbosch University https://scholar.sun.ac.za

d.	mbuzi	ka-mbuzi	kambuzi	ka-kambuzi	kambuzi	wa-nga
	goat	DPF-goat	PL-type of chilli	DIM-chilli	type of chilli	ASC-POSS
		"type of chilli"	"kambuzi chillis"	"small kambuzi"	"my kambuzi chilli"[9]	

As shown in (12), derivational processes can operate on dependent nominal stems as in (12a-b) or independent stems as in (12c-d). Therefore, nominal expressions may become complex NPs by bound affixes or free modifiers such as adjectives. I will consider the prefixes as playing two functions: first as number affixes (SG and PL), and second as derivational prefixes (DPF). I assume that number affixes only attach to N.stems while DPF may attach to N.stems or to stems of other word categories such as verb stems, adj stems, etc. I discuss the formal structure of Chichewa nominal expressions in more detail in Chapter 6, Section 6.3.5.

In the next chapter, I consider in more detail the elements comprising each AC. Before turning to this matter, let us consider the main findings of the present chapter.

3.5 Conclusion

In the present chapter, I have achieved two major objectives. First, I introduced the methodological considerations that were followed in obtaining the dataset for the present study. I argued that the nature of the phenomenon at hand requires us to make generalisations regarding rules that underlie the noun and agreement class system. To draw such generalisations, the dataset must be obtained by following some systematic random sampling method. I argued that this sampling requirement has been satisfied by using data from the *Chichewa monolingual dictionary*, which in corpus linguistic terms is assumed to be a statistically representative sample. I also introduced the agreement-based classification approach adopted in the present study, which I argued differs significantly from the NCP-based approach that informs the Bleek-Meinhof noun class system. I showed that the agreement-based classification approach gives rise to twelve ACs in Chichewa compared to the traditional seventeen-noun class system. Following this, I presented the statistical distribution of elements that make up the dataset; specifically, I showed that some ACs contain more lemmatised elements than others. I argued that the quantitative division observed in the dataset is significant in the way we approach the agreement system in

[9] So-named because the chilli has the aroma of goat meat, therefore, the *kambuzi* noun in (12d) should not be confused with the homophonous diminutive reading "small goat".

Stellenbosch University https://scholar.sun.ac.za

Chichewa; specifically, that the morphologically predictable expressions and the less productive forms must both be accounted for properly in the theory of Chichewa noun and agreement classes.

The second objective was to introduce the morphological make-up of the Chichewa nominal dataset that informs the present study. Here, I considered the various noun types that comprise the Chichewa dataset. I showed that, contrary to the traditional noun prefix-centred views about Bantu nominal structure, Chichewa has three broad types of nouns. The first type encompasses the simplex lexicon, which is not derived by any morphological processes, but exhibits idiosyncratic properties. The second type includes less regular nominal expressions that are obtained by various strategies such as covert conversion, tone modification, suffixation or prefixation. The third group includes highly productive expressions whose forms are not lemmatised in the dictionary. I argued that many of these noun-types are not considered in any systematic manner in the previous literature, which mainly focused on prefixed nouns. Given this background, I argued that the so-called noun prefix only manifests either as a number affix or a derivational affix. These observations necessitated a revised view of the status of various elements in the nominal structure. I therefore, argued that null prefix nouns are bare nouns that are independent, whereas nominal stems can be independent or dependent. Dependent nominal stems were shown to require a prefix or suffix to function as independent forms.

To sum it all, in the present chapter, I have introduced the descriptive typology that will be prerequisite for more insightful explanation of the phenomenon in Chapters 5, 6 and 7. In the next chapter, I continue to discuss this descriptive typology by focusing in more detail on properties of expressions that comprise each of the ACs introduced in the present chapter.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 4

Empirical evidence for rethinking the Chichewa noun class system

4.1 Introduction

In this chapter, I focus, in more detail, on the properties of the expressions in each of the twelve agreement classes (ACs) introduced in Chapter 3. These ACs will be discussed from Section 4.3 through Section 4.14. In each of these sections, I show that ACs in Chichewa comprise expressions with varied morpho-semantic characteristics. In addition, some ACs include expressions that are not usually associated with the nominal category, such as CPs and non-locative prepositions. There are also cases where the AMs function as dummy affixes such that they are not controlled by any specific agreement triggers. In Section 4.15, I provide an interim summary of the discussion of the various ACs and their membership. In Section 4.16, I propose that the agreement system in Chichewa comprises two subsystems: one triggered by typical nominal expressions, hence deriving *agreement-based noun classes* (ANCs) and the other containing more general expressions, hence forming *general agreement classes* (GACs). I will then show that Chichewa has seven ANCs and five GACs. I will also introduce the idea that the ANC subsystem of Chichewa primarily consists of two super-ANCs that are semantically based, namely the animate/agentive and inanimate/non-agentive ANCs. Further, the inanimate/non-agentive class is observed to subclassify nouns based on their word-initial characteristics. Although the GACs govern agreement in a morphosyntactically predictable manner, they are also observed to be partly phonologically determined as the expressions making up these classes share certain word-initial phonological characteristics. In Section 4.17, I discuss the relationship between the nominalisation and the noun classification grammatical systems in Chichewa. Finally, I conclude the chapter in Section 4.18.

4.2 The dataset

As introduced in Chapter 3, all the nouns in the dataset and the productive (and thus non-lemmatised) expressions in Chichewa that I am considering trigger at least one of the twelve AMs. These expressions will be discussed in groups on the basis of their associated ACs. The discussion will proceed according to the order represented in Table 4.1.

Stellenbosch University https://scholar.sun.ac.za

Table 4.1: Summary of ACs in Chichewa

	Agreement class	Associated traditional NC	Number of tokens	Percentage
1	*m(u)-a*-AC	≈NC1	1222 (∞)	20.9%
2	*li*-AC	≈NC5	843	14.5%
3	*a*-AC	≈NC2 & NC6	357 (∞)	6.1%
4	*u*-AC	≈NC3 & NC14	839 (∞)	14.3%
5	*i*-AC	≈NC4 & NC9	1453 (∞)	24.9%
6	*chi*-AC	≈NC7	1075 (∞)	18.4%
7	*zi*-AC	≈NC8 & NC10	16 (∞)	0.3%
8	*ka*-AC	≈NC12	∞	
9	*ti*-AC	≈NC13	2 (∞)	0.03%
10	*ku*-AC	≈NC15 & NC16	6 (∞)	0.1%
11	*pa*-AC	≈NC17	13 (∞)	0.2%
12	*mu*-AC	≈NC18	2 (∞)	0.03
	TOTAL		**5828**	**100**

As pointed out in Chapter 3, the quantitative disparity in the number of tokens between the first six ACs and the next six ACs is very striking. Later, in Section 4.16, I will show how this points to the existence of two different types of ACs in Chichewa.

In the next section, I discuss expressions that are found in the *m(u)-a*-AC.

4.3 The *m(u)-a*-AC, which includes NC1

According to the dataset, nominal expressions that trigger AMs *m(u)-a-* (associated with NC1) make up approximately 21% (1222/5828) of the sampled nominal lexicon. However, as indicated by the infinity symbol in Table 4.1, there are also several sets of potentially productive nominal derivation processes whose nouns are not lemmatised. Contrary to what the traditional Bantu noun class analyses suggest, the morphological characteristics of these nouns are quite diverse. As introduced in Chapter 3, Section 3.3, in addition to the traditional prefix *m(u)-*, there are also several nouns that do not involve the traditional prefix *m(u)-*. For example, the dataset includes nouns derived via the prefixes *ka-*, *na-*, *sa-*, *ma-*, *o-/wo-*, *a-/wa-*, *cha-*, *bwa-* and *tsa-*. In addition to these nouns, there are also nouns that are formed by several other strategies such as covert conversion, NFV suffixation, tone modification, compounding, etc. Further still, there are simplex nouns that are not obtained through any known word formation processes. As far as can be ascertained, some of them have not

Stellenbosch University https://scholar.sun.ac.za

been documented in any systematic manner in Chichewa grammars. Table 4.2 shows how these noun-types are distributed in the *m(u)-a*-AC.

Table 4.2: Statistical distribution of noun-types in the m(u)-a-AC

	Noun-types and morphological characteristics	**Number of tokens**	**Percentage**
a	Type 1, NFV, tone modification, conversion	38	3.1%
b	Type 2, *m(u)*-initial	150	12.2%
c	Type 2, *ka*-initial	365	29.9%
d	Type 2, *na*-initial	125	10.2%
e	Type 2, *sa*-initial	14 (∞)	1.1%
f	Type 2, *ma*-initial	23	1.8%
g	Type 2, *o-/wo*-initial	∞	∞
h	Type 2, *cha*-initial	9 (∞)	0.7%
i	Type 2, *bwa- and tsa*-initial	22	1.8%
j	Simplex & compounds	476	38.9%
TOTAL		**1222** (∞)	**100%** (∞)

Each of the noun types in Table 4.2 will be discussed in more detail in the following sections. First to be examined are nouns that belong to what I have designated as *Type 1.*

4.3.1 Type 1 nouns (conversion, tone modification, NFV)

As introduced in Chapter 3, Section 3.3.2.1 above, some nouns are derived from verb stems by means of (covert) conversion, tone marking or by NFV only. This is shown in the singular examples in (1).[10] For a complete set of nouns of this type, see Appendix 2.

(1)		**Verb**	**Singular noun**	**Plural noun (NC2)**
	a.	gund-a	gundá	a-gundá
		ROOT-VFV	N.STEM	PL-drum
		“hit”	“a type of drum in Nyau dance”	“gunda drums”
	b.	gulul-a	gúlúle	a-gúlúle
		ROOT -VFV	N.STEM	PL-childish adult
		“dislocate”	“childish adult”	“childish adults”

[10] As stated in Chapter 3, I will continue to include plural nouns along with the singular examples to show the various types and functions of the noun prefixes; however, note that, unless otherwise stated, the focus in each set of examples is on the singular forms.

Stellenbosch University https://scholar.sun.ac.za

c.	gonth-a	gónth-i	a-gónthi
	ROOT-VFV	deaf-NFV	PL-deaf person
	“be deaf”	“deaf person”	“deaf people”
d.	gogod-a	gogoda	a-gogoda
	ROOT-VFV	N.STEM	PL-high-heeled shoe
	“knock, tap”	“high-heeled shoe”	“high-heeled shoes”
e.	gogod-er-a	gogodera	a-gogodera
	ROOT-APP-VFV	N.STEM	PL-kind of rifle
	“tap for”	“an old kind of rifle”	“kind of rifles”

As shown in (1), all the singular nouns are obtained without the aid of any prefix. However, they all control the AMs *m(u)-a-,* the ones associated with the traditional NC1. Crucially, the derivational processes illustrated in (1) are regular such that in everyday language use, novel nouns and proper names are created in this way.

However, some nouns derived by the means illustrated in (1) appear to be dependent stems such that they require a prefix even in their singular form. I discuss nouns derived from these dependent stems in sections 4.3.2 to 4.3.8.

4.3.2 Type 2 nouns derived via the prefix *m(u)-*

In the present section, I consider nouns that have been identified as characteristic of the traditional NC1. In my dataset, these nouns only represent 12% (150/1222) of all the lemmatised nouns in the *m(u)-a-*AC. Just as with the Type 1 nouns above, a subset of the nouns in the *m(u)-*group are derived by what are called *action/state deverbal and deadjectival nominalisation*, which refers to various types of derivational processes that tum events or states denoted by the verb or the adjectives into nouns (see Kishindo, 1985: 4–6; Comrie & Thompson, 2007: 335; Kramer, 2015: 187). Consider the following examples.[11]

(2)	**Stem**	**Singular noun**	**Plural noun**
a.	-lim-a	m-lim-i	a-lim-i (NC2)
	ROOT-VFV	SG-cultivate-NFV	PL-cultivate-NFV
	“cultivate”	“farmer”	“farmers”

[11] These examples will not be included in the appendix as they are generally accepted as typical of this class.

Stellenbosch University https://scholar.sun.ac.za

b.	-lond-a	m-lónda	a-lónda (NC2)
	ROOT-VFV	SG-N.STEM	PL-N.STEM
	“watch over”	“watchman”	“watchmen”
c.	-lesi	m-lesi	a-lesi (NC2)
	ADJ.STEM	SG-N.STEM	PL-N.STEM
		“lazy person”	“lazy people”

As proposed in Chapter 3, the prefix attaches to an already nominalised stem derived either by NFV, tone change or covert conversion. Take note that the prefixation in (2) is substitutive, i.e. the plural form substitutes the singular prefix. As I will show in the present chapter, some affixes follow the substitutive strategy while others are additive, i.e. the plural form adds on the singular prefix (Maho 1999:58). Examples illustrated in (2) are not controversial among Bantuists generally, but what may be contested is the analysis that the prefix only carries number features and not noun class features.[12] The distribution patterns shown in (1) and (2) suggest that the singular prefixes are only required by dependent nominal stems. For example, in some colloquial usage the stem *lónda* in (2b) can also be used without the singular prefix. As will become more evident in the course of the present chapter, the empirical evidence being considered here does not support the assumption that the prefix carries noun class features.

Although the prefix *m(u)-* was argued to be the main prefix for NC1, some previous studies have also observed that this prefix is constrained in many ways. For example, Kishindo (1985: 4–5) observes that not all verbs can be nominalized in this manner due to semantic constraints that the word formation rule (WFR) must be sensitive to. For example, the derived nouns in (3) may not be generally accepted as regular nominal forms in Chichewa despite satisfying the WFRs stipulated in Mchombo (1978, chap. 3).

(3)	a.	yend-a	*myendi	b.	tol-a	*mtoli
		walk-VFV	walker		pick-VFV	picker

Strikingly, there are other means that derive nouns that denote the very same referents intended in (3). For example, all verb stems in (3) can be nominalised by the prefixes *w-/o-* which are discussed in Section 4.3.7 below. Therefore, the constraints in (3) could be due to the fact that there already exists a lexical item that the action verb may refer to (a case of

[12] As already pointed out, Carstens (1991) and subsequent works have long argued for this view, but Bantuists have unfortunately not adopted it.

Stellenbosch University https://scholar.sun.ac.za

lexical blocking therefore). These constraints should serve as a clue that the prefix *m-* cannot be the sole prefix for derivational/inflectional processes of this type in the *m(u)-a-* group. It is therefore puzzling why research in Bantu has insisted on the prefix *m(u)-* as the sole prefix for NC1.

In addition to deverbal and deadjectival nominalization, the prefix *m(u)-* also derives nouns from both dependent and independent nominal stems. Consider the following examples.

(4)

	Stem	**Singular noun**	**Plural noun (NC2)**
a.	Malawi	m-Malawi	a-Malawi
	NOUN	SG-N.STEM	PL-N.STEM
	Malawi	“Malawian”	“Malawians”
b.	Chewa	m-Chewa	a-Chewa
	NOUN	SG-N.STEM	PL-N.STEM
	“Chewa tribe”	“a Chewa person”	“Chewa people”
c.	-zungu	m-zungu	a-zungu
	N.STEM	SG-N.STEM	PL-N.STEM
		“white person”	“white people”
d.	-nthu	mu-nthu	a-nthu
	N.STEM	SG-N.STEM	PL-N.STEM
	“being”	“human being”	“human beings”

In (4a-b), the prefixes are attached to independent nouns (name of a place and name of a tribe) to derive gentilic nouns (see also Harding, 1966: 26–35). In (4c), the prefix is attached to an idiosyncratic dependent noun stem, again to derive a gentilic noun. Equally in (4d), the same prefix attaches to another idiosyncratic dependent noun stem to derive a noun of human reference. These examples suggest that it is indeed the case that the prefix may carry more than just number features. In these specific examples, the alleged second set of features on the prefix could be associated with the gentilic interpretation of the nouns. However, in the traditional literature, this additional set of features has been regarded as core and has been associated with noun class features (see i.a. Welmers, 1971: 2; Kishindo, 1985: 2; Maho, 1999: 2–4; Mchombo, 2004: 3; van der Spuy, 2009: 3). Since the traditional analyses have mainly used these noun-types, these views have remained unchallenged.

It also appears that the requirement to add something to make the stem pronounceable does not have to be fulfilled by adding a prefix, but can also be fulfilled by adding other material. For example, the dependent nominal stem *-nthu* (“being”) in (4d) cannot stand on its own but requires either a singular or plural prefix. However, in some varieties of

Stellenbosch University https://scholar.sun.ac.za

Chichewa, the stem may also be used in second person reference without the prefixes but instead it takes the honorific or plural clitic (*-ni*), as in *nthu-ni* ("you person (HONORIFIC)/ you people"). In this case, the stem may optionally take the number prefixes suggesting that all that the stem *-nthu* needs is an extra syllable to become an independent morpheme.

So far, I have considered *m(u)*-prefixed nouns that have a human denotation. These are also the types of examples that have been frequently cited in the previous studies. This could be the reason why NC1 has been referred to as the *people/human* class in Chichewa and in many other Bantu languages (cf. Corbett & Mtenje, 1987: 4; Herbert, 1991; Matiki, 2001: 66, 68; Contini-Morava & Kilarski, 2013: 270 among many others). However, this is just part of the whole picture; just as in many other languages, derived agentive nominals also have non-human referents. For example, the English *-er*, which according to Marchand (1969: 215), "in most cases denote[s] a person, more specifically a male person" such as *potter*, *Londoner*, *banqueter,* etc. also derives nouns that do not have human denotation e.g. *a two-seater*, *two-decker*, *freighter* (cf. Lieber, 2004: 1; Yang, 2016: 106–121). Similarly, in Chichewa not all nouns derived by the prefix *m(u)-* have meanings denoting humans. Consider the examples in (5) where the nouns refer to non-human entities (cf. Watkins, 1937: 23).

(5)	**Verb expression**	**Singular noun**	**Plural noun**	
a.	lumikiz-a	m-lumikiz-i	a-lumikiz-i	
	ROOT-VFV	SG-join-NFV	PL-join-NVF	
	"join"	"conjunction"	"conjunctions"	Lit. "joiner(s)"
b.	nen-a	m-nen-i	a-nen-i	
	ROOT-VFV	SG-say-NFV	PL-say-NFV	
	"say"	"verb"	"verbs"	Lit. "one who says"
c.	lowa m-malo	m-lowammalo	a-lowammalo	
	enter in-place	SG-N STEM	PL-N STEM	
	"[VP enter the place]"	"pronoun/substitute"	"pronouns/substitutes"	

The evidence shown in (5) is in sharp contrast with the traditional view regarding NC1, that the prefix *m(u)-* derives nouns that have a specifically human denotation (see i.a. Mchombo, 1978; Kishindo, 1985; Corbett & Mtenje, 1987; Matiki, 2001). In fact, it appears that Chichewa differs remarkably from other Bantu languages in this regard. For example, Hlungwani (2012: 183) observes that in Xitsonga the prefix *mu-* for NC1 strictly derives nouns that refer to humans (see also Mugane (1997: 22) for Gikuyu). Similarly, Ferrari (2005: 176) concludes as follows: "In conformity with other Bantu languages, the Luganda noun

Stellenbosch University https://scholar.sun.ac.za

classes 1/2 contains only [+human] [+animate] nouns. In my database, 100% of the nouns of class 1/2 denote a [+human] person confirming the semantic nature of this class." This is certainly not the case in Chichewa.

What has traditionally been taken as the NCP for NC1 in Chichewa appears to be just one of the number-cum-derivational affixes. In fact, according to my dataset, *m(u)-* only contributes 12% of the *m(u)-a-*AC (≈NC1). We have also seen that the morpheme is constrained in the way it selects various stems. The prefix *m(u)-* in Chichewa is akin to the English derivational morphemes *-ist*, *-er,* etc. that also designate classes of 'agents' such as *scientist*, *novelist, helper, shredder, printer,* etc. (cf. Grinevald, 2000: 59; Senft, 2007: 679). As will become much clearer by the end of the present chapter, the traditional analyses mischaracterise the role of the prefix *m(u)-* to the extent that other prefixes are not recognised equally in Chichewa when they clearly need to be.

In the next section, I consider nouns in the *m(u)-a-*AC that are derived via the second type of prefix, namely the derivational prefix *ka-*.

4.3.3 Type 2 nouns derived via the prefix *ka-*

In the previous literature, this prefix has been deemed no longer productive in many Bantu languages (see i.a. Doke, 1927: 198; Cole, 1955: 11; Givón, 1971a: 35; Vail, 1971: 40; Lombard, 1985: 33; Maho, 1999: 74–75, 252–253). Strikingly, in my dataset, the prefix is actually more productive in Chichewa than the traditional prefix *m(u)-*. Nouns derived by the *ka-* prefix contribute approximately **30%** (365/1222) of all nouns that control AMs *m(u)-a-* (see Appendix 3). Like the *m(u)-* prefix, the prefix *ka-* derives nouns exhibiting a wide range of morphosyntactic and semantic characteristics. Three different types are illustrated here through examples (6), (7) and (8) below.

Like the previous *m(u)-* prefix, *ka-* also derives denominal, deadjectival and deverbal nouns. Consider the examples in (6).

(6)	**Stem**	**Singular noun**	**Plural noun**
a.	litsiro dirt	ka-litsiro DPF-dirt "a dirty person"	a-kalitsiro PL-dirty person "dirty people"
b.	chipapa winnowing basket	ka-chipapa DPF-winnowing basket "a type of mask dancer"	a-kachipapa PL-mask dancer "mask dancers"

Stellenbosch University https://scholar.sun.ac.za

c.	gwinthi	ka-gwinthi	a-kagwinthi
	stunted	DPF-stunted	PL-unripe pumpkin
		“a small unripe pumpkin”	“small unripe pumpkins”

Unlike the *m(u)-* prefix, plural marking in (6) is additive, i.e. the plural affix does not replace the prefix that derived the singular form. From the examples in (6), we can see that the derived nouns are generally descriptive of the source noun or adjective, for instance, denoting ‘person/thing that looks like …’, ‘person/thing that has …’ etc. However, this is not to be generalised to all possible examples. As for nouns formed from verbal stems, again there are a wide range of morphological and semantic properties. Consider examples (7a-b) where the prefix *ka-* attaches to a Type 1 noun as discussed in Section 4.3.1, specifically noting the tone variation between the source verb and the derived noun.

(7)	**Stem**	**Singular noun**	**Plural noun**
a.	pasul-a	ka-pasúle	a-kapasúle
	ROOT-VFV	DPF-N.STEM	PL-backbiter
	“untie/devastate”	“backbiter, spy”	“backbiters, spies“
b.	soz-a	ka-sóze	a-kasóze
	ROOT-VFV	DPF-N.STEM	PL-spy
	“lookout/see”	“spy” or “a type of mouse that looks out for danger”	
c.	budul-a	ka-budula	a-kabudula
	ROOT-VFV	DPF-N.STEM	pl-pair of shorts
	“remove tips of”	“pair of shorts”	“pairs of shorts”

The type of expressions represented in (7a-b) can be read either as verbal commands, e.g. “go cause disunity”, “go look out” or as nouns denoting somebody who engages in the activity denoted by the source verb. Take note that examples (7a-b) terminate with the indicative VFV *a-* while that in (7c) ends with the subjunctive VFV *e-.* The prefix *ka-*, like prefix *m(u)-* in section 4.3.2 above, can also attach to complex expressions such as VPs with reflexive markers (RFM) and other verbal extensions such as applicatives (APP). This is illustrated in (8).

(8)	**VP**	**Singular noun**	**Plural noun**
a.	dzi-otch-e	ka-dziotche	a-kadziotche
	RFM-ROOT-VFV	DPF-burn yourself	PL-brave person
	“burn yourself”	“brave person/ self-destructive being/ moth”	
b.	dzi-met-e	ka-dzimete	a-kadzimete
	RFM-ROOT-VFV	DPF-shave yourself	PL-bare necked chicken
	“shave yourself”	“bare necked hen”	“bare necked hens”

Stellenbosch University https://scholar.sun.ac.za

c.	pand-a-mneni	ka-pandamneni	a-kapandamneni
	ROOT-VFV-verb	DPF- be without-verb	PL-infinitive verb
	“be without-verb”	“infinitive verb”	“infinitive verbs”

In addition to deriving common nouns as illustrated above, *ka-* is also a very productive prefix for deriving proper names for persons, places, rivers, mountains, etc. This is shown in (9).

(9) a. Katelera (https://goo.gl/maps/bNfp7o723Xp)
b. Kampini (https://goo.gl/maps/8YqkvZDXNhr)
c. Katete (https://goo.gl/maps/fSCThsoo2kt)
d. Kaphula (https://goo.gl/maps/4pPNQLTTHPN2)

Just like the prefix *m(u)-* the prefix *ka-* gives an agentive meaning or denotes a description of a particular property roughly corresponding to the meaning of the expression that it attaches to. The derivation of these agentive nouns is not very different from those observed in other languages: the agents can be both human and non-human. This further contradicts the generally held assumption that nouns that trigger AM *m(u)-a-* (so-called *NC1*) comprise human agents only. If we are to maintain the traditional view that NC1 in Chichewa is human, perhaps these non-human denoting derived nouns should be treated as a case of **semantic extension**, as Dixon (1972: 306–311) proposes for an Australian language, Dyirbal (cf. Kilarski, 2013: 213–214).[13]

The prefix *ka-* illustrated in the foregoing discussion should, however, be distinguished from the four further homophonous ones, namely:

(i) the diminutive *ka-* associated with the traditional NC12 such as in *ka-munthu* (“a small person”); see Chapter 6, Section 6.3.1;
(ii) the manner nominalization prefix, e.g. *ka-yendedwe* (“manner of walking”); see Section 4.10;
(iii) the directional marker in imperatives that denotes future orientation e.g. *ka-gone* “go sleep”; and
(iv) what Hullquist (1988: 39) refers to as a non-productive *ka-*adverbializer occurring as a frozen element in adverbials such as *ka-ngati* (“how many times”), *ka-tatu* (“three times”) – which can be glossed as meaning “times” in this context.[14]

[13] For example in Dyirbal myth, birds are considered to be spirits of dead women and so are assigned to gender II with other female-denoting nouns and not to gender I with other animate nouns (Dixon, 1972: 308).

[14] However, I do not agree with Hullquist’s (1988: 39) analysis that this *ka-* is unproductive. This function of *ka-* is highly productive as it can attach to any adjective stem to convey the idea of “*n* times”.

Stellenbosch University https://scholar.sun.ac.za

These four different types of *ka-* are disambiguated by the denotative meaning of the derived word and also by the syntactic and pragmatic contexts in which they occur. With regard to the *ka*-facts that are centrally relevant here, I can conclude that the view that the prefix *ka-* is non-productive and unpredictable in Bantu languages (see i.a. Givón, 1971a: 35; Maho, 1999: 75, 252–253) does not hold for Chichewa. Next, I turn to nouns derived by the related prefix, *na-.*

4.3.4 Type 2 nouns derived via the prefix *na-/ná-*

The previous literature generally discusses the prefix *ka-* alongside the *na-* prefix as both are taken to be a relic of an earlier system (see Maho, 1999: 74). Some studies also allocate the nouns bearing the prefix *na-/ná-* to the traditional subclass NC1a, generally associating the prefix with kinship terms and proper nouns (see i.a. Givón, 1971a: 34–35; Bresnan & Mchombo, 1995: 245; Maho, 1999: 74–75; Demuth, 2000: 275). The differences between high toned *ná-* and the non-high toned *na-* have not been clearly investigated. However, the high toned *ná-* is predictably found with kinship terms and less predictably with non-kinship nouns.

Like the prefix *ka-*, the prefix *na-* is also found to be productive in Chichewa.[15] It contributes approximately 10**%** (125/1222) of all the nouns that belong to the *m(u)-a-*AC in the dataset. The *na-* prefix also shares the morphosyntactic and semantic characteristics of the prefixes *m(u)-* and *ka-*. For example, *na-* attaches to nominal, adjectival and verbal stems to create nouns descriptive of the source expression's denotation as shown in (10) (see also Appendix 4).

(10)	**Source expression**	**Singular noun**	**Plural noun**
a.	nyongo	ná-nyongo	a-nányongo
	gall bladder	DPF-gall bladder	PL-clitoris
		"clitoris"	"clitorises"
b.	-dzi-tch-e	na-dzitche	a-nadzitche
	RFM-root-VFV	DPF-give oneself a name	PL-namesake
	"give oneself a name"	"namesake"	"namesakes"

[15] Sometimes when a new phenomenon arises in Malawi, the names that are given are derived by either prefix *na-* or *ka-.* For example there was *na-chi-panti* (na-big-underwear: a name given to a mysterious serial killer who was always seen in underwear) (see i.a. Ashforth, 2014: 854), *na-mapopa* (na-sucker: a name given to suspected magical blood suckers) (Chinele, 2017), *ka-unjika* (ka-heap: a name given to second-hand clothes as they are often displayed in heaps at the market) (Nyondo, 2013). This is another source of evidence that these prefixes are productive in the synchronic lexicon.

Stellenbosch University https://scholar.sun.ac.za

c.	m-f-edw- a	na-mfedwa	a-namfedwa
	SG-ROOT(die)-PASS-FV	DPF-bereaved person	PL-bereaved person
	“bereaved person”	“bereaved person”	“bereaved people”
d.	ma-sipuni	na-masipuni	a-namasipuni
	PL -spoon	DPF-spoons	PL-water hyacinth
	“spoons”	“water hyacinth”	“many water hyacinth”

Although the prefix may attach to other nouns and verb phrases, this prefix shows a tendency to co-occur with expressions that already have a nasal prefix (especially prefixes *m-* and *ma-*), as shown in (10c-d). When attached to nominal expressions, *m-* and *ma-* serve as singular and plural markers for the source nominal stems, respectively. In such structures, the prefix *na-* attaches to an already nominalised expression, one with the singular prefix *m-* or with the plural prefix *ma-.* In all such cases, the derived nouns denote meanings related to properties of the relevant source noun e.g. ‘one who looks like…, one who possesses particular features, etc.’ In some cases, the prefixes *na-* and *ma-* appear as one unit, behaving as if it is a complex prefix, *náma-*.[16] However, the fact that *na-* also appears with other expressions without the *ma-* prefix (e.g. the singular affix *m-*) suggests that the prefix does not come as a complex unit, *nama-.*

Among the Chewa and other tribes from southern Malawi, the prefix *ná-* is also well known for deriving female clan names from male ones, as shown in (11).[17]

(11)	**Male clan names**	**Female clan names**
a.	Phiri	Ná-phiri
b.	Banda	Ná-banda
c.	Zunga	Ná-zunga
d.	Jere	Ná-jere

Note that the clan names consistently take the high-toned version of *na-.* It is due to examples such as those in (11) that some previous studies have divided the traditional NC1 into 1 and 1a, where the latter is identified as comprising kinship terms (see i.a. Maho, 1999: 64; Matiki, 2001: 66, 76).[18] However, according to my dataset, in Chichewa the *ná-* is not only for deriving kinship names, as shown in (10) and Appendix 4. In addition to this, just

[16] The view that *náma-* could be a complex prefix was suggested to me by Dr. Winfred Mkochi, personal communication.

[17] In many cases, in these communities, elders are not called by their first or surname, but by their clan names.

[18] Note that while *na-* was assigned to the subclass NC1a, all the other prefixes in this class such as *ka-* have not been treated the same. It is this type of unsystematic classification that has motivated the present study to describe the Bleek-Meinhof system as based on partial data.

Stellenbosch University https://scholar.sun.ac.za

like the prefix *ka*- above, *na*- is by far the most common productive prefix for deriving names of people, places, mountains, rivers, etc. such as those in (12).

(12) a. Nathenje (https://goo.gl/maps/UuM7Y2xaukJ2)
b. Namadzi (https://goo.gl/maps/jxeD5F6aSLk)
c. Namitete (https://goo.gl/maps/SrkSJrtuNZF2)
d. Namitembo (https://goo.gl/maps/MDgBXtdGnjo)

The prefix *na*- also appears to be closely associated with the first-person pronoun *ndi* ("I") in verbal expressions as in *ndi-bwera* ("I am coming"), *nda-bwera* ("I have come"), as some nouns alternate between the *nda* and *na*- forms. Consider examples in (13).

(13)	**Verb**	**Singular noun**	**Plural noun**
a.	lem-a	nda-lema/na-lema	a-ndalema/a-nalema
	ROOT-VFV	I am- tired	PL-easy chair
	"be tired"	"easy chair"	"easy chairs"
b.	lap-a	nda-ku-lapsa/na-kulapsa	a-ndakulapsa/a-nakulapsa
	ROOT-VFV	I am-OM-be contrite	PL-scorpion
	"be contrite"	"a type of scorpion"	"scorpions"
		Lit.: "I have learnt a bitter lesson from you."	

It is not yet clear whether the first-person pronoun *na*- is related to the prefix *na*- under consideration in the present section. There is obviously more to the various derivational affixes in Chichewa. However, the task at hand is not to provide a fully-fledged analysis of the nominalization processes in Chichewa, but, instead, to illustrate that nominal expressions in each AC are obtained by more than one derivational strategy. For this reason, I now turn to the fourth type of nominal derivation in Chichewa which involves the negative (NEG) marker *sa*-.

4.3.5 Type 2 nouns derived via the prefix *sa*-

The prefix *sa*- is generally known as a NEG marker in Chichewa which attaches to the verb stem to mark negation (cf. Mchombo, 1978: 311–313, 2004: 13, 28). The underlying form of this marker is assumed to be *si*- and it manifests allophonically as *si-, sa*- or *su*-, depending on the phonological context. The vowel on the NEG marker is observed to coalesce with the phonological features of the following vowel as in *s-a-bwera* [from *si-a-bwera*] ("he/she/they will not come"), *s-u-bwera* [from *si-u-bwera*] ("you will not come") or *si-ndi-bwera* ("I am not coming"). The vowel coalescing with the NEG is either the verb's subject marker (SM) or object marker (OM). As regards nominal derivation, the morpheme appears to perform a

Stellenbosch University https://scholar.sun.ac.za

similar function to the previous three prefixes *m-, ka-* and *na-,* except that *sa-* denotes a negative meaning, e.g. 'person/thing that **does not** …', as illustrated in (14).

(14)

	Source	**Singular noun**	**Plural noun**
a.	samb-a	sà-sámbà	a-sàsámbà
	ROOT-VFV	NEG-N.STEM	PL-one who does not bath
	"bath"	"someone who does not bath"	"people who do not bath"
b.	funs-a	sà-fúnsà	a-sàfúnsà
	ROOT-VFV	NEG-N.STEM	PL-one who does not ask
	"ask"	"someone who does not ask"	"people who do not ask"
c.	-mw-a mowa	sáá-mwàmowa	a-sáámwàmowa
	ROOT-VFV beer	NEG.SM-drink beer	PL-a kind of fish
	[VP drink beer]	"a kind of (bottle-nosed) fish"	"many bottle-nosed fish"
d.	sos-a	sí-m´-sosa	a-sím´sosa
	ROOT-VFV	NEG-OM-prepare the garden	PL-a type of bird
	"prepare the garden"	"a type of bird"[19]	"many birds of this type"

The examples illustrated in (14) may have both verbal and nominal readings. The difference is determined by tone marking. For example, if the first three series of syllables in the expression have the **Low-High-Low** tone pattern, then it is read as a noun whereas the **High-High-Low** tone pattern gives the expression a verbal reading (Mchombo, 1978: 313). However, this is only a common pattern; other patterns are also attested such as (14c, d) where the pattern is similar to the verbal **High-High-(Low)**. Although it is not initially very clear whether the nominalisation comes before or after the NEG *sa-* is attached, the tone patterns indicate that the nominalisation comes after the NEG prefixation. In this regard, then, the NEG *sa-* may not necessarily be on a par with the other prefixes discussed above.

Although this form of nominalisation is grammatically predictable and productive, some nouns formed in this way have been lexicalized and these contribute 1% (15/1222) to all nouns in the *m(u)-a-*AC (See appendix 5). In addition to these, in Malawi, *sa-* is another very common prefix found on names of persons, places, rivers, mountains, etc. as illustrated in (15).

[19] This type of bird is believed to herald the coming of the rain season. In some communities when this bird appears it is taken as a sign that one must start preparing the garden.

Stellenbosch University https://scholar.sun.ac.za

(15)	a	lima (“cultivate”)	Sà-líma (https://goo.gl/maps/atWHHeN46PH2)
	b.	kata (“stop raining”)	Sà-kata (https://goo.gl/maps/XwyyfUF1vhu)
	c.	pit-idwa/pit-wa (“go-PASSIVE”)	Sà-pítwa (https://goo.gl/maps/GNfrFrMZs6x)
	d.	kwata (“have sex/marry”)	Sà-kwáta (name of a person)

Next, I consider nouns in the *m(u)-a*-AC that carry the fifth type of derivational affix, namely the prefix *ma-.*

4.3.6 Type 2 nouns derived via the prefix *ma-*

Traditionally, the prefix *ma-* is known as the NCP for NC6 (which is the plural class assumed to pluralise nouns from NC5 and NC14), which I have designated as the *a*-AC in the present study. As will become clear below, some animate-denoting nouns formed by this prefix control AMs *m(u)-a-* (≈NC1), as shown in (16) (see also Appendix 6).

(16)	**Source**	**Singular noun**	**Plural noun**
a.	-nthu N.STEM “being”	ma-nthu DPF-N.STEM “mother ant”	a-manthu PL-mother ant “mother ants”
b.	tay-a ROOT-VFV “throw away “	ma-tay-a DPF-throw away “affluent person”	a-mataya PL-affluent person “affluent people”
c.	tsakamul-a ROOT-VFV “cause to fall down”	ma-tsakamula DPF-cause to fall down “rainmaker”	ø-matsakamula “rainmakers”

Like the other prefixes, *ma-* also attaches to nominal and verbal stems. In (16a) the prefix is attached to the nominal stem *-nthu* (“being”). Note that the same nominal stem in (4d) was shown to take prefixes *m(u)-* and *a-* and belonged to the same *m(u)-a*-AC. The dataset registered approximately 2% (23/1222) of lexicalised nouns of this type in the *m(u)-a*-AC.

This prefix is also common in deriving proper names such as those in (17).

(17)	a.	Mapeto	Lit. “the end / at the end of something”
	b.	Mathero	Lit. “end of something”
	c.	Malunda	Lit. “the enraged one”
	d.	Magomero	Lit. “where something ends”
	e.	Matola	Lit. “one who picks up things anyhow”

Since the prefix *ma-* is involved in the derivation of nouns of the type shown in (16-17) and also serves as a plural marker for nouns associated with the traditional NC5 and NC14,

Stellenbosch University https://scholar.sun.ac.za

there is a lot of confusion in the way these nouns are classified.[20] For example, the *Chichewa monolingual dictionary* has most *ma*-initial nouns placed in the so-called *Li-Ma* class (NC5/6) (see Section 4.6.4 for specific details).

In the next section, I present the sixth nominalization strategy for nouns that also control *m(u)-a*-AMs.

4.3.7 Type 2 nouns derived via the prefixes *wa-/a-* and *wo-/o-*

As discussed in Section 4.3.2, particularly as illustrated in example (3), some verb stems cannot be converted into agentive nouns by means of the prefixes considered so far in the *m(u)-a*-AC. Such verbal stems prefer to take what are traditionally considered as adjectival prefixes *wo-/o-* or *wa-/a-* to derive nouns of various kinds (cf. Nankwenya, 1992: 50–51; Matiki, 2000).[21] The prefixes *wo-/o-*or *wa-/a-* are also a result of coalescence between the nominal AM and the associative marker (ASC) *á* ("of"), the combination of which literally means "it/he/she of *x*" as illustrated in (18). The *wó-* and *wá-* mark singular whereas *ó-* and *á-* mark plural number. Consider the following examples.

(18)

	Verb stem	**Singular noun**	**Plural noun**
a.	nam-a	wó-nama	ó-nama
	ROOT-VFV	SG.ASC-lie	PL.ASC-lie
	"lie"	"liar"	"liars" (Lit. "she/he/them of lie")
b.	tsuts-a	wó-tsutsa	ó-tsutsa
	ROOT-VFV	SG.ASC-oppose	PL.ASC-oppose
	"oppose"	"opposer"	"opposers" (Lit. "s/he/them of opposition")
c.	yembekez-er-a	wó-yembekezera	ó-yembekezera
	ROOT-APP-VFV	SG.ASC-wait for	PL.ASC-wait for
	"wait for"	"expectant woman"	"expectant women" (Lit. "… of waiting for")
d.	vot-a	wó-vota	ó-vota (Eng.LW)
	vote	SG.ASC-vote	PL.ASC-vote
	"vote"	"voter"	"voters"

[20] The multifunctionality of this prefix is also observed in another Bantu language, Shona (Déchaine *et al.*, 2014: 20). These authors also argue that analyses that identify the prefix as having a dedicated function are "idealized descriptions" and they also observe that Shona "N-class prefixes are multi-functional, with the same prefix coding several contrasts."

[21] In another Bantu language, Xitsonga, a similar type of nominalisation is also observed. According to Hlungwani (2012: 187), the form *wa-* is described as a possessive concord. In Chichewa however, the form does not only function as a possessive concord, but generally as an adjectival prefix or associative marker for NC1 nouns (cf. Mchombo, 2004: 24).

Stellenbosch University https://scholar.sun.ac.za

Like the prefix *m(u)-* in Section 4.3.2, the prefixation strategy in (18) is substitutive, i.e. the plural form replaces the singular one. Therefore, we have two types of prefixes that are substitutive in the *m(u)-a-*AC, namely *m(u)-a-* and *wo-/o-*or *wa-/a-*. The adjectivally marked expressions in (18) can serve as modifiers of a noun, as in (19a) or as a typical stand-alone noun in (19b).

(19) a. mtsogoleri **wó-nama**
leader AM.ASC-lie
"The leader who is a liar."

b. A-na-meny-edwa ndi **wá-misala**. (Matiki 2000:55)
s/he-T/A-beat-PASS by AM.ASC-madness
"S/he/it was beaten by a mad person."

This is by far the least constrained prefix for deriving nouns of human reference, as it can be used to derive novel expressions from any verbal or adjectival stem. It is actually the escape route for referents that one does not have appropriate terms for, e.g. *wó-chuna* (from English verb *tune – i.e. "*someone who performs a tuning function of any kind").

Matiki (2000: 55) also observes that expressions such as those in (19) are "habitually used as noun[s] in Chichewa". This phenomenon, where adjectives function as nouns has been observed crosslinguistically, such that expressions of this type have been referred to by a variety of terms, such as "adnouns" (Martin, 1986), "adjectives used as nouns" (Yamamura, 2010: 344), "headless nominals" Bresnan (1995: 31), etc. (cf. Nkemnji, 1995: 160; Matiki, 2000: 55; Gil, 2013). Nkemnji (1995: 160) claims that the descriptive terms associated with these expressions suggest that they are a blend of adjectives and nouns. I agree with these observations and assume that these cases are simply another strategy for nominalisation; hence the outputs can be described as *nominalised adjectives.* In addition to the examples in (18-19), there are also nominalised adjectives that have generic interpretation. Consider the following (see also Nkemnji, 1995: 147; Yamamura, 2010: 344):

(20) a. ó-lemera b. ó-sauka c. ó-dwal-a d. ó-sa-phunzira
PL-be heavy PL-be poor PL-be sick PL-NEG-learn
"the rich" "the poor" "the sick" "the uneducated"

Nouns of this type are not lemmatised in the dictionary because of their grammatical predictability. However, everyday use of the language clearly shows that most of these words have been lexicalized. For example, signage such as *ó-dwala* ("patients"), *ó-yembekezera* ("expectant women"), *óyendapansi* ("pedestrians") is commonly found in

Stellenbosch University https://scholar.sun.ac.za

Malawi. In fact, there are no known regular words apart from these nominalised adjectives. Interestingly, all these nouns belong to the *m(u)-a-*AC.

Closely related to the nominalization involving *wo-/o, wa-/a-* prefixes is the seventh set of derivational prefixes *cha-*, *bwa-,* and *tsa-.* I discuss these in the next section.

4.3.8 Type 2 nouns derived via the prefixes *cha-*, *bwa-* & *tsa-*

The prefixes (*cha-, bwa-* and *tsa-*) will be discussed together because their morpho-semantic properties are closely related. Like the *wá-/wó-* prefixes; *cha-, bwa-* and *tsa-* also convey possessive and associative meanings, as illustrated in (21).[22]

(21)	**Source**	**Singular**	**Plural**
b.	makala charcoal	chá-makala DPF.ASC-charcoal “a kind of black edible mushroom”	a-chámakala PL-a kind of mushroom “black edible mushrooms”
c.	mpini hoe/axe handle	bwá-mpini DPF.ASC-hoe/axe handle “African giant rat, poor man, etc.”	a-bwámpini PL-African giant rat, etc. “African giant rats, etc.”
d.	munda garden	tsá-munda DPF.ASC-garden “estate owner/colonialist”	a-tsámunda PL-estate owner “estate owners/colonialists”

The manner in which *cha-, bwa-* and *tsa-* derive these nominals is not any different from the way prefixes such as *wá- or wó-* do this. The prefixes *cha-, bwa-* and *tsa-* also roughly denote ‘possession “of” a particular property’.[23] The prefixes *bwa-* and *tsa-* are not very productive: there are only 1.3% (16/1222) lexicalised nouns formed from *bwa-* and only 0.5% (6/1222) from *tsa-* in the dataset. However, the prefix *cha-* is very productive. Although there are only 0.9% (9/1222) entries in the dataset, there are many non-lexicalised nouns derived with the aid of the prefix *cha-*. The prefixes *cha-, bwa-, tsa-* are also found in some proper names, for example, *Chankhungu, Chamachete, Chanunkha, Bwandiro, Bwaila, Bwanje, Bwabwa, Tsabango, Tsangano, etc*.

[22] The prefix *bwa-* and close variants are also attested in other Bantu languages e.g. in the languages called Lomongo (C. 16) and Bubi (A. 31) (Hyman, 1980: 190).

[23] In some varieties of Chichewa, the prefix *bwa-* is also used as an AM associated with some *b*-initial nouns such as bowa ***bwa****nga* (mushroom mine -> “my mushroom”).

Stellenbosch University https://scholar.sun.ac.za

This concludes the discussion of Type 2 nouns in the *m(u)-a-*AC. The next section considers nouns that are not acquired by any productive or regular processes, specifically compounds and non-derived simplex nouns.

4.3.9 Compounds and simplex nouns in the *m(u)-a-*AC

In addition to the nouns derived through predictable means discussed in the preceding sections, there are also various nouns that do not undergo these predictable morphological processes, namely compound and simplex nouns. On the one hand, compounds are obtained by means of juxtaposing two or more (dependent or independent) stems of varying categorial statuses, for example noun+noun, verb+noun, as shown in (22) (see i.a. Mchombo, 1978: 181–293, 2004: 117–118; Nankwenya, 1992: 52; Bresnan & Mchombo, 1995: 222).

(22)	**Singular**	**Plural**
a.	msungi-chuma (N+N) keeper-money “treasurer”	a-msungichuma PL-treasurer “treasurers”
b.	gontha-mkutu (V+N) be deaf-in the ear[24] “a type of edible flying ant”	a-gonthamkutu PL-flying ant “flying ants”
c.	mkhala-pa-mpando (N-LOC-N) sitter-on-chair “chairperson”	a-mkhalapampando PL-chairperson “chairpersons”

The compounding processes illustrated in (22) may not be as predictable as the derivational processes discussed in the preceding sections. Since not every juxtaposition of lexical items shown in (22) would obtain acceptable nominal expressions, I treat compounding as an unpredictable word formation process.

The simplex lexicon, on the other hand, is not acquired by any overt morpho-syntactic processes. This category of noun also includes loan words that are not morphologically modified. Consider the following examples.

[24] The name comes from the myth that these insects may cause deafness if eaten, though many people (including the researcher) do eat them, both in a raw and fried state.

Stellenbosch University https://scholar.sun.ac.za

(23)	**Singular**	**Plural**	
a.	zakhali (“aunt”)	a-zakhali (“aunties”) (NC2)	
b.	huluku (“duiker”)	a-huluku (“duikers”) (NC2)	
c.	tekitivi (“detective”)	ma-tekitivi (“detectives”) (NC6)	Eng.LW
d.	venda (“vendor”)	ma-venda (“vendors”) (NC6)	Eng.LW

As introduced in Chapter 3, Section 3.3, the singular forms of nouns illustrated in (22-23) show no predictable morphological pattern. Compound and simplex nouns comprise approximately 39% (476/1222) of all the nouns that make up the *m(u)-a-*AC in my dataset. However, note that this excludes some nouns that are simplex, but whose initial syllables are accidentally homophonous with some of the prefixes discussed above. These have been excluded here for two reasons. First, it is not always straightforward to differentiate prefixes and spurious forms because some stems do not have transparent meanings. Second, as will become clear by the end of the present Chapter, there are word-initial characteristics that are associated with many ACs such that in such cases it does not matter whether the word-initial syllable is a prefix or a spurious form. This is a pattern observed to be common with *ka-* and *na-* initial nouns within the *m(u)-a-*AC.

As already stated, some previous studies have treated non-derived nouns as very limited in number, proposing that they have lost their prefixes over time (see i.a. Givón, 1971a: 35; Herbert, 1991: 105; Kishindo, 1998: 45; Maho, 1999: 67; and see also Chapter 2, Section 2.4.). According to my dataset, however, this claim does not seem to be well motivated, as approximately **39%** of the so-called NC1 nouns are non-derived. These nouns have a range of semantic and morphological characteristics.

With all this in place, the question that now arises is what underlying principles there are that determine whether a noun belongs to the *m(u)-a-*AC. I briefly consider this question in the next section.

4.3.10 The common characteristics for nouns in the *m(u)-a-*AC

In this section, I outline what appear to be the common underlying characteristics of nouns belonging to the *m(u)-a-*AC. Firstly, the existing literature has associated nouns triggering the AMs *m(u)-a-* (the so-called NC1) with the semantic attributes of [+human] and [+animate] (see i.a. Corbett & Mtenje, 1987: 9–10; Maho, 1999: 51; Demuth, 2000: 6; Katamba, 2006: 114–116). This view is borne out in my dataset. However, as was pointed out at relevant points, there are also several other agentive but inanimate nouns that are derived via the

Stellenbosch University https://scholar.sun.ac.za

various strategies discussed in the present chapter. So, what appears to be the case is that this class includes agentive nouns, which subsumes both animate and human nouns. The strongest piece of evidence supporting the view that this class comprises agentive nouns comes from the distribution of NFVs. The NFV *i-* is, for example, observed to derive agentive nouns while the NFV *-o* derives non-agentive nouns (see Mchombo, 1978: 105–118, 2004: 113–114; Kishindo, 1985, see also Chapter 5, Section 5.3 for specific statistical details). In addition to agentive meanings denoted by the NFVs, many prefixes such as *ka-, na-, cha-, sa-, wa-* and *bwa-* also have the properties of agents such as 'having or not having a property of something', 'ability to do something', etc. In this regard, the non-human or inanimate derived nouns found in the *m(u)-a-*AC broadly have agentive connotations. Therefore, the major semantic features characterising this AC are [+human], [+animate] and [+agentive], which are broadly referred to as *animacy* features in crosslinguistic studies (see i.a. Comrie, 1989: 9; Dahl & Fraurud, 1996; Yamamoto, 1999; Dahl, 2000; de Swart, Lamers & Lestrade, 2008; Becker, 2014; Igartua & Santazilia, 2018).

Note that the notion of animacy does not fully coincide with the biological dimension or dictionary denotation which generally makes a distinction between living and non-living things but may also involve a wide range of social-cultural categorisation, which may classify some inanimate entities as animate and vice versa (Comrie, 1989: 186; Mithun, 1999: 98; van Nice & Dietrich, 2003: 106; de Swart *et al.*, 2008: 135; Becker, 2014: 70). Therefore, due to the variations in what qualifies to be animate or inanimate, the term 'animacy' is used as a convenient label not necessarily depicting the biological distinction of animals and inanimates, human and non-human (Comrie, 1989: 186; Mithun, 1999; van Nice & Dietrich, 2003: 106; de Swart *et al.*, 2008: 135).

Further, animacy effects are often observed to manifest in a form of a continuum, which is variedly known as the *animacy hierarchy* or *agentivity hierarchy* (Comrie, 1989: 199; Payne, 1997: 150; Yamamoto, 1999, chap. 1; de Swart *et al.*, 2008: 132; Becker, 2014: 63–75). Although the animacy hierarchy is often represented as a three-step scale: human > animals (animate) > inanimate (see i.a. Payne, 1997: 150; Yamamoto, 1999: 2; Croft, 2003: 130), there are different cut-off points regarding what is animate or inanimate in different languages (van Nice & Dietrich, 2003: 106; de Swart *et al.*, 2008: 135). It is therefore worth noting that the semantic boundaries involved in the animacy hierarchy are fuzzy and vary from one language to another. Interestingly, various grammatical structures are observed to

Stellenbosch University https://scholar.sun.ac.za

follow the hierarchy such that certain generalizations may only apply to all cases above certain cut-off points of the animacy hierarchy (Dahl & Faurud 1996:47, de Swart). For example, in some languages, direct objects are grammatically marked only if they are higher on the animacy hierarchy (Dahl & Fraurud, 1996: 47). In the present study, I also argue that the Chichewa noun class system is organised along the Chichewa-specific animacy hierarchy (see also Chapters 5 and 7).

Having introduced the notion of the animacy hierarchy, it becomes easier to understand the further subcategories of the noun classification for nouns forming the *m(u)-a*-AC. First, if the derived nouns in this class have animate/agentive meanings (as shown by the +human, +animate, +agentive derived nouns above), I expect this to be the underlying property of the simplex lexicon which the derived lexicon is assumed to extend. To a large extent, this is the case. Out of 476 nouns comprising the irregular compound and simplex lexicon, 70% (333/476) have clear-cut animate denotations. The remaining 30% (143/476) cannot be said to have clear-cut animate/agentive properties, however they also occur in clusters that share certain common characteristics, e.g. plants, instruments, etc. Table 4.3 summarises these clusters.

Table 4.3: Sets of inanimate nouns in the m(u)-a-AC

Category	Statistical distribution	Reference
Letters of the alphabet	17% (24/143)	Excluding letters X and Q
Plants	32% (46/143)	Appendix 7A
Instruments	23% (33/143)	Appendix 7B
Non-count miscellaneous nouns	20% (29/143)	Appendix 7C
Inherently a-initial nouns	8% (11/143)	Appendix 7D

As shown in Table 4.3, the first set includes letters of the alphabet, which I suggest falls into the broader category of metalinguistic reference. In Chichewa and many other Bantu languages, any metalinguistic reference to a word controls the AM *m(u)-a-*. For example, in the following sentence '*Yenda' mmodzi wafufutika* ("one '*yenda'* has been rubbed off" – i.e. in a list of words, the word *yenda* is rubbed off). This pattern also applies to all reference to English-based numbers, such as *wani* ("one"), *thuu* ("two"), *fiili* ("three"), etc. In this regard, all metalinguistic reference, alphabet letters and numbers form a cluster of nouns that are inanimate but belong to the *m(u)-a*-AC.

Stellenbosch University https://scholar.sun.ac.za

The second cluster includes nouns referring to various types of instruments, which suggests that the assignment of the nouns to this class is based on this semantic attribute as there is no other common morphological or phonological basis for triggering AM *m(u)-a-*. Perhaps, some instruments are treated as agents because of their ability to perform actions. This kind of noun classification has also been attested in other languages (see i.a. Fortune, 1970: 95–96; Dixon, 1972: 306–311; Corbett, 1991: 33).

The third cluster of inanimate nouns consists of names of various plants. However, most of the plant names in this class are non-count, a property which has also been observed in what I have identified as non-count miscellaneous set in Table 4.3. In this regard, it is not clear whether they are assigned to this class because they are plant names or because they belong to the non-count category that identifies the fourth cluster. As will become clear when the other classes have been considered, the plant names and the non-count/mass sets of nouns are also found in the other ACs. However, in those other ACs, nouns with these characteristics are observed to be there based on other characteristics, for example their word-initial phonological features. So it appears that Chichewa includes some plants in its set of animates on the animacy hierarchy, which according to Corbett (1991: 14–24) is a pattern that is also observed in other languages such as Zande (a Niger-Congo language of Zaire), Ket (an isolate Siberian language), Ojibwa (an Algonquian language of United States) (see also Berlin, 1973; Yamamoto, 1999: 10–13; Becker, 2014: 288). Interestingly, in all these other languages, the quantity of plant-denoting nouns in the animate classes is very low. Nevertheless, the clusters into which these nouns fall suggest how the various languages structure their animacy hierarchies. In Chapter 5, I subject these quantitative distribution patterns to a rule-productivity metric to find out if these 'exceptions' to animacy would counterweight the productivity of the animacy criterion in Chichewa.

The fifth cluster includes a meagre collection of inherently *a*-initial nouns. These *a*- initial nouns differ significantly from those where the word-initial *a*- is a derivational or inflectional prefix. The latter type controls the AM *a-.* As will become clear from Chapter 5, the inherently *a*-initial nouns in the *m(u)-a-*AC are assigned to this class on a word-initial phonological basis since this AC also involves *a*- as one of the AMs. Therefore, although the classification seems to be semantically determined, there are elements of word-initial phonological criteria (see Chapter 5 for specific details).

Stellenbosch University https://scholar.sun.ac.za

At this point, the major goal is just to note the common characteristics underlying nouns in the *m(u)-a*-AC, so as to provide background to a more insightful analysis in Chapter 5. Suffice to say that the majority of nouns in this class are assigned to it based on a gradient of animacy features which I claim characterises the Chichewa animacy hierarchy. Therefore, the *m(u)-a*-AC can be argued to represent nouns that are higher on the animacy hierarchy.

According to the classification system used in the *Chichewa monolingual dictionary*, there are also several nouns that were considered to belong to the *m(u)-a*-AC, by association with the plural AM *a-.* However, not all plural nouns that control AM *a-* are in NC1; some of them belong to NC6 which pluralises nouns from NC5 and N14. Some of these nouns from NC6 were wrongly classified as belonging to the NC1/NC2 class. These misclassified nouns are discussed in Section 4.6.4 below.

In the next subsection, I summarise the discussion of the *m(u)-a*-AC.

4.3.11 Interim summary

This far, I have shown that nouns that belong to the *m(u)-a*-AC have a range of morpho-semantic characteristics; some simplex and others complex. The complex ones are also derived through a number of strategies, such as conversion, suffixation, tone modification, prefixation, etc. If the Bleek-Meinhof NCP-based approach was the one to go by, then we would have a proliferation of classes as each prefix would have to have its own noun class as is already the case with traditional classes NC1, NC1a, NC6, NC6a and the various traditional NCs that share the same AMs. However, what is observed to be a consistent pattern is that each AM in every AC participates in deriving/inflecting a subset of nouns in the respective classes but does not imply that it is the sole prefix for the ACs. Therefore, with regard to the *m(u)*-AC, all the considered noun-types with their diverse morphological properties belong to one AC, which is fundamentally determined by semantic properties. These nouns are primarily +human, +animate and +agentive, i.e. they are entities that cause or initiate an event or state (Kroeger, 2005: 54). I also stated that the set of entities that appear to comprise the animate cut-off point include some inanimate things such as letters of the alphabet, numbers, names of instruments and some plants. In summary, there are the following denotative as well as connotative semantic attributes that seem to be key in this class: agentive, human, animate, plants, and instruments. More importantly, I have concluded that these features are an instantiation of the animacy hierarchy in Chichewa.

Stellenbosch University https://scholar.sun.ac.za

In the next section, I consider nouns that trigger AM *li-*, which have been associated with NC5 in the Bleek-Meinhof schema.

4.4 The *li*-AC, which includes NC5

According to my dataset, there are **843** nouns that trigger AM *li-*. Like nouns in the *m(u)-a*-AC discussed above, the *li*-AC also features nouns with diverse morphological characteristics. I consider each of these noun-types in subsections 4.4.1 to 4.4.3.

4.4.1 Type 1 nouns (conversion, tone or FV modification)

The dataset shows that the *li*-AC also comprises some nouns that are derived from verb stems by (covert) conversion, tone marking or by means of FV modification - i.e. without the involvement of prefixes. The dataset indicates that 79% (85/107) of all derived nouns in this AC are prefixless. Consider the following examples (see also Appendix 8).

(24)	**Verb Stem**	**Singular Noun**	**Plural** (NC6)
a.	lemb-a ROOT-VFV “write”	lémba N.STEM “letters/scripts”	ma-lémba PL-letter “letters/scripts”
b.	ful-a ROOT-VFV “dig”	fule N.STEM “mound, heap of soil”	ma-fule PL-mound “mounds, heaps of soil”
c.	yankh-a ROOT-VFV “answer”	yankh-o reply-NFV “answer/reply”	ma-yankho PL-answer “answers/replies”
d.	lemb-ets-a ROOT-CAUS-VFV “cause to write”	lembets-o cause to write-NFV “dictation”	ma-lembetso PL-dictation “dictations”

As shown in (24), nouns of this type are derived by conversion, tone or FV modification (similar derivational options were observed for the *m(u)-a*-AC in Section 4.3.1). Importantly, this type of nominal derivation for nouns belonging to NC5 has not been considered systematically in the previous literature. According to Mchombo (1978: 117), the type of nominalisations illustrated in (24) “do not appear to conform to the required standard.” In other words, they lack the prefixes one would expect on the traditional analysis of Bantu NCs. To explain the lack of noun prefixes in derived nouns such as those in (24), Mchombo (1978: 115–118) claims that the initial prefix in NC5 has been lost and that this type of nominalisation only applies to verbs whose initial segments are plosives or liquids (cf.

Stellenbosch University https://scholar.sun.ac.za

Kishindo, 1985: 8). He further argues that those verbs that are not plosive- or liquid-initial such as *yankh-o* (“answer/reply”), *sewer-o* (“play”) and *funs-o* (“question”) “descended from earlier forms which had initial affricates rather than fricatives. … but somehow that did not lead to a reclassification of the nouns” (Mchombo, 1978: 117). However, Mchombo’s (1978) claims do not seem to explain the predictable aspects of this data, which do come to the fore on the analysis proposed here (see, in this connection, Section 4.16 where the relationship between nominalisation strategies and noun classification is discussed**,** and also Appendix 8). Mchombo’s (1978) views are expressed within the tradition that holds that all nouns in Chichewa ought to have the prefixes.

Another important property of derived nouns in this class is the high frequency of NFV *-o,* i.e. the final vowel associated with non-agentive nouns. According to the dataset, of the 73 derived nouns involving the NFVs *-i* and *-o* in this AC, 99% (72/73) of them terminate in NFV *-o* and only 1% (1/73) bears the NFV *-i.* This gives an indication that this AC takes nouns that are lower on the animacy hierarchy.

In the next section, I consider the second type of nouns in the *li*-AC.

4.4.2 Type 2 nouns, derived via the prefix *li-*

As discussed in Chapter 2, NC5 is generally analysed as a null prefix class, which is assumed to be the relic of an old noun class that derived nouns via the reconstructed prefix *li-*. It is also claimed that Chichewa specifically lost NC11 and that its prefix merged with NC5 *li-* (Matiki, 2001: 67). However, as I will show, these views follow from the Bleek-Meinhof approach’s preoccupation with prefixed nouns at the expense of other simplex and prefixless derived nouns. Evidence emerging in this study shows that the prefix ***li-*** is just one of the many derivational affixes in the *li*-AC. First, consider the nouns derived by the prefix *li-* in (25).

(25)	**Source**	**Singular**	**Plural** (NC6)
a.	-fupi	li-fupi	ma-lifupi
	ADJ.STEM	DPF-short	PL-width
		“width”	“more than one width”
b.	-kulu	li-kulu	ma-likulu (headquarters)
	ADJ.STEM	DPF-big	PL-headquarter
		“headquarter”	“headquarters”

Stellenbosch University https://scholar.sun.ac.za

c.	-ps-a	lu-psa	ma-lupsa
	ROOT-VFV	DPF-burn	PL-bush fire
	“burn”	“bush fire”	“bush fires”
d.	-kodz-a	li-kodz-o	UNCOUNTABLE
	ROOT-VFV	DPF-urinate-NVF	
	“urinate	“bilharzia”	

Note that, as introduced in Chapter 3, Section 3.3.2, the *li-* prefix is not necessarily the number affix as number marking is shown to be additive, i.e. the plural affix does not replace the prefix that derives the singular noun (cf. Price, 1958: 35; Maho, 1999: 58). Morphologically, this type of prefix is similar to many other prefixes in the *m(u)-a*-AC, such as *ka-, na-, sa-, bwa-, tsa-, etc.,* which also do not mark number. However, the *li*-prefixation nominalization strategy is not very productive in that it is not attested in creating new words over and above already lexicalised ones, and it is also not found to participate in more complex nominalization processes, e.g. attaching to VPs. Of course, there are several names of places, and rivers that I can speculatively assume to have been derived from this prefix, for example *Lilongwe, Likuni, Liwaladzi, Liwonde, Lilangwe, Linthipe, Lifidzi,* etc. However, the would-be roots to which the prefix would attach do not seem to have regular meanings compared to names derived by the prefixes *ka-, na-,* and *sa-* shown in examples (9), (12) and (15), respectively. Because the prefix *li-* is less predictable it is hard to distinguish between the nouns derived by prefix *li-* and those that are coincidentally syllable *li*-initial. I discuss the latter in Section 4.4.4.

Although the Bleek-Meinhof system only recognises the prefix *li*, in Chichewa, there are other prefixes that derive nouns for this class. I consider these in the next section.

4.4.3 Type 2 nouns, derived with the involvement of prefixes *dz-*, *d-* and NFVs

Contra the traditional assumptions, in addition to the prefix *li-*, there are also other prefix-like strategies that derive nouns that belong to the *li*-AC. First, consider the following examples (see also Harding, 1966, chap. 52; Carstens, 1997: 379).

(26)	**Source**	**Singular**	**Plural**
a.	ombol-a	d-ombol-o	UNCOUNTABLE
	ROOT-VFV	DPF-rescue-NFV	
	“rescue”	“rescue/ransom”	
b.	longosol-a	d-ongosol-o	ma-dongosolo
	ROOT-VFV	DPF-organise-NFV	PL-system
	“organise”	“system/order”	“systems”

Stellenbosch University https://scholar.sun.ac.za

c.	lip-a	d-ip-o	ma-dipo
	ROOT-VFV	DPF-pay-NFV	PL-fine
	“pay”	“fine, payment, pledge”	“fines/ payments/ pledges”
d.	-nthu	thu-nthu	ma-tunthu
	-ROOT	DPF-ROOT	PL-trunk, whole
	“being”	“trunk, whole”	“trunks”

As shown in (26), the prefix *d-* appears to function in the same way as the *li-* in (25): in some cases, the phoneme /*d*/ replaces the /*l*/ consonant to derive a noun. According to Demuth (1988: 78), in Sesotho, /l/ and /d/ are allophones where /d/ only appears before high vowels. However, this is not the case in Chichewa where these consonants exhibit phonemic variation, except in some less predictable cases.[25] In (27d), the root *-nthu* also appears with an idiosyncratic prefix *thu-* which is not observed to be productive at all.

In addition to the derivation processes illustrated in (25-26), Chichewa also features nouns that appear to involve singular-plural alternations that differ from those considered above. Consider the following (see also Harding, 1966, chap. 52; Carstens, 1997: 379).

(27)	**Singular (NC5)**	**Plural (NC6)**
a.	dzina (“name/noun”)	ma-ina (“names/nouns”)
b.	dzanja (“hand”)	ma-nja (“hands”)
c.	duwa (“flower”)	ma-luwa (“flowers”)
d.	bele (“breast”)	ma-(b)ele (“breasts”)
e.	tsamba (“leaf”)	ma-samba (“leaves”)
f.	phanga (“cave”)	ma-panga (“caves”)

Although the plural formation process illustrated in (27) appears to be predictable, the roots in these examples do not appear to participate in any regular morphological processes so that it is dubious whether the singular forms are indeed made up of prefix and stem. However, if the Bleek-Meinhof NCP-based approach is to be followed, these were also supposed to be identified as different noun prefixes and to form another noun class: similar inflectional processes have been considered to involve NCPs that derive distinct noun classes such as NC3 and NC4.

[25] There are rare cases where the /l/ and /d/ variation do no change meaning, for example *lumpha/dumpha* (“jump”). In more predictable cases however, the variation changes meaning, for example *dowe* (“green maize”)/ *lowe* (“water logged”), *luka* (“weave”)/ *duka* (“become cut into two”), which implies that /l/ and /d/ may not be described as clear-cut allophones.

Stellenbosch University https://scholar.sun.ac.za

As has been shown for all the examples in this class, the derivation- and number-marking processes are not very regular. In fact, the majority of nouns in this class appear to be of the non-derived type, i.e. those I have described as forming the simplex lexicon. This majority is considered next.

4.4.4 Simplex nouns that control AM *li-*

Since I identified 107 nouns in this class as exhibiting properties of derivation/inflection in Section 4.4.1, there are 736 nouns, including the singular forms illustrated in example (27) that need to be treated as non-derived. This implies that approximately 87% (735/843) of the nouns in this AC are either non-derived or obtained by unpredictable processes, e.g. compounding. This confirms the previous literature which identifies this class as predominantly prefixless (see i.a. Carstens, 1997: 379; Maho, 1999: 87). Some of these nouns are illustrated in (28).

(28)

	Singular	**Plural** (NC6)	
a.	bondo (“knee”)	ma-wondo (“knees”)	
b.	khosi (“neck”)	ma-kosi (“necks”)	
c.	vuzi (“pubic hair”)	ma-vuzi (“pubic hair”)	
d.	luso (“talent”)	ma-luso (“talents”)	
e.	lisiti (“receipt”)	ma-lisiti (“receipts”)	Eng.LW

Since this AC comprises a large number of simplex and prefixless nouns which have a wide range of word-initial characteristics, it is very unlikely that its underlying unifying criterion is based on word-initial properties (for instance the morpheme *li-*), as is the case with the other AC to be discussed below. As shown by the Type 1 nouns in the *m(u)-a-*AC, this class has a large number of nouns derived by means of the NFV *-o,* i.e. that are associated with non-agentive and inanimate nouns. In fact, according to the dataset, only 5% (43/843) of the nouns in this class have an animate and agentive denotation (see Appendix 9). This implies that 95% (800/843) of the nouns in the *li-*AC are inanimate. Interestingly, most of the animate nouns in this class have special characteristics, for example small insects, small kinds of fish, undesirable or frowned upon humans, etc. Consider the following examples.

(29)

	Singular	**Plural**
a.	bunda (“baby dove”)	ma-unda (“baby doves”)
b.	khanda (“baby”)	ma-kanda (“babies”)
c.	thekenya (“jigger-flea”)	ma-tekenya (“jigger-flea”)
d.	linthumbu (“very tiny type of ants - red driver ants”)	linthumbu (“red driver ants”)
e.	hule (“prostitute” (Afrikaans loan word)	ma-hule (“prostitutes”)

Stellenbosch University https://scholar.sun.ac.za

It appears that by invoking the inanimate AM the implication is to show that these animate entities have some peculiar features. This is also a common noun classification phenomenon observed in many other languages (see i.a. Fortune, 1970: 95–96; Dixon, 1972: 306–311; Corbett, 1991: 33). Considering this semantic pattern and the low frequency of agentive and animate nouns in this class, it would be reasonable therefore to conclude that this class represents elements lower in the animacy hierarchy, as introduced in Section 4.3.10 (see also Chapter 5, Section 5.3).

As already stated in Chapter 3, Section 3.2, nouns belonging to the *li*-AC are classified as belonging to NC5/6 (*Li-Ma*) in the *Chichewa monolingual dictionary*. However, what is assumed to be the plural NC6 features several nouns that are not derived via the singular noun channel (see Section 4.5.2.2 below, such that treating the plural-singular pairs as forming one overarching class makes a very wrong generalisation. In this regard, I treat nouns assumed to belong to the NC5/6 (*Li-Ma*) as belonging to two distinct ACs. In the next section, I discuss nouns that belong to the *a*-AC, which includes the traditional NC2 and NC6.

4.5 The *a*-AC, which includes NC2 and NC6

According to the Bleek-Meinhof classification system, nouns that trigger AM *a*- belong to two different noun classes, namely NC2 and NC6. The distinction is based on the assumption that NC2 nouns are prefix *a*-initial and also mark plural number for nouns assumed to belong to NC1, whereas NC6 nouns are prefix *ma*-initial and mark plural number for nouns assumed to belong to NC5 and NC14. However, the dataset used in the present study shows that there are diverse noun types that control AM *a*-. Some are indeed derived through the pluralisation process while others do not have any singular counterparts. It may thus not be entirely correct to consider all nouns that control AM *a*- as plural nouns of some other singular classes. The *a*-AC nouns are discussed in the following order: in Subsection 4.5.1, I consider the vowel *a*-initial nouns (NC2) whereas in Subsection 4.5.2, I focus on the *ma*-initial nouns, those typically assumed to originate from singular NC5 and NC14.

4.5.1 The *a*-initial noun types in the *a*-AC

There are two types of *a*-initial nouns in the *a*-AC, namely (i) plural nouns, collectively identified as NC2 in the traditional literature and (ii) honorific nouns, which have been associated with NC2 and its subclass NC2a on the Bleek-Meinhof schema (see i.a. Fortune,

Stellenbosch University https://scholar.sun.ac.za

1955: 66ff; Maho, 1999: 51, 93). Focusing on the plural nouns first, the examples discussed in Section 4.3 have shown that count nouns in the *m(u)-a-*AC are pluralised by means of the prefix *a-*. Several examples were given alongside each singular noun. I repeat some of these in (30).

(30)	**Plural nouns (*a*-AC)**	**From singular nouns {*m(u)-a*-AC}**
a.	a-nalimata (“geckoes”)	na-limata (“gecko”)
b.	a-nyamata (“boys”)	m-nyamata (“boy”)
c.	a-neni (“verbs”)	m-neni (“verb”)
d.	a-kalulu (“hares”)	kalulu (“hare”)
e.	a-galu (“dogs”)	galu (“dog”)

The examples illustrated in (30) represent a very regular pluralisation process in Chichewa. Many (but not all) count nouns from the *m(u)-a-*AC derive their plural forms via this mechanism (for nouns that do not follow this rule, see (31b, c) below).

However, not all *a*-initial nouns in the *a*-AC have a plural denotation; some *a*-initial nouns convey honorific meanings. There are clear differences between the plural and the honorific (HON) prefix *a-*. For example, honorific nouns may have a singular reading and may attach to proper nouns whereas the plural prefix *a-* always yields a plural reading and may not pluralise proper nouns. In fact, some nouns bearing the HON *a-* prefix do not take the *a-* as their plural marker, as shown in (31a-c) (honorific expressions are discussed in more detail in Chapter 6, Section 6.3.3).

(31)	**Honorific singular nouns**	**Plural counterparts**
a.	a-Obama HON-PERSONAL NAME “[HON] Mr. Obama”	NO PLURAL i.e. “the one and only Obama”[26]
b.	a-pulezidenti HON-president “[HON] the president”	ma-pulezidenti PL-president “presidents”
c.	a-dokotala HON-doctor “[HON] the doctor”	ma-dokotala PL-doctor “doctors”
d.	a-phunzitsi HON-teacher “[HON] the teacher”	a-phunzitsi PL-teacher “teachers”

[26] However, if it happens that one wants to refer to two *Obamas* the Chichewa equivalent is *ma-Obama*

Stellenbosch University https://scholar.sun.ac.za

The honorific nouns in (31a-c) do not take *a-* as the plural marker, which implies that the HON markers in these examples are unambiguously singular. In (31d), however, the HON and plural nouns are ambiguous. Interestingly, these honorific nouns take a full range of AMs for the *a*-AC in all domains; however, the interpretation is not plural, but unambiguously singular as shown in (32).

(32) a. **A**-dokotala **a**-thu **a**-wa-samutsa dzulo
HON-doctor AM-POSS.1ST PRSN PL-AM-transfer yesterday
"Our [HON] doctor has been transferred yesterday."

Against this backdrop, I would like to avoid assuming the traditional generalisation that nouns that trigger the AM *a-* are made up of the plural prefix *a-* only. Rather, as I will conclude later, what appears to be the case is that prefix *a-* and generally syllable *ma*-initial nouns control AM *a-* whether they are countable or uncountable.

However, this generalisation does not extend to a set of nouns that are inherently *a*-initial which have been observed not to take plural agreement marking. These were introduced in Section 4.3.10 (see also Appendix 7D). These inherently *a*-initial nouns consistently control the *m(u)-a-* AMs whether singular or plural. It thus appears that the agreement marking in Chichewa is also sensitive to the distinction between prefixed *a-* and inherently *a*-initial syllabic forms.

In the present subsection, I have shown that the prefix *a*-initial nouns that belong to the *a*-AC comprise plural nouns and singular/plural honorific nouns. In the next section, I consider another set of nouns in the *a*-AC, namely the *ma*-initial nouns.

4.5.2 The *ma*-initial noun-types in the *a*-AC

The *ma*-initial nouns are also observed to occur in at least four different morphological types, namely (i) countable nouns with the plural prefix *ma-*; (ii) mass nouns with prefix *ma-*; (iii) inherently *ma-* initial nouns and (iv) manner nominals. Each of these will be introduced in separate sub-sections below.

4.5.2.1 Nouns with the plural prefix ma-

As stated above, nouns associated with NC5 and NC14 mark their plural number by means of the prefix *ma-*. However, according to the dataset, these plural nouns may also derive their singular counterparts from other NCs such as NC1 and NC9, a pattern that is observed

Stellenbosch University https://scholar.sun.ac.za

to be very common with loan words (cf. Taraldsen, 2010: 1524 for similar pattern in Zulu). Consider the following examples.

(33)	**Plural nouns {*a*-AC (≈NC6)}**	**From singular nouns**		
a.	ma-piri (“mountains”)	phiri (“mountain”)	NC5	
b.	ma-uta (“arrows”)	uta (“arrow”)	NC14	
c.	ma-polofesa (“professors”)	polofesa (“professor”)	NC1	Eng.LW
d.	ma-suti (“suits”)	suti (“suit”)	NC9	Eng.LW

In (33) the *ma*-initial syllable is a clear-cut plural marker, which derives plural nouns from their singular counterparts. As shown in (33), the singular forms are from four different traditional singular NCs.

However, in addition to the plural prefix *ma*- shown in (33), there is another set of *ma*-initial nouns that do not have singular counterparts, either because the derived noun is a mass noun or because the *ma*- is not necessarily functioning as a number prefix. These are discussed in the next section.

4.5.2.2 Ma-initial nouns without singular counterparts

The second category includes *ma*-initial nouns that do not have singular counterparts (see Appendix 10). These *ma*-initial nouns can further be categorised into two subtypes. The first type includes nouns where the *ma*-is a clear-cut derivational/number prefix which derives nouns from the nominalised verbal, adjectival, and ideophonic stems. However, these nouns do not have singular counterparts as shown in (34).[27]

(34)	**Source expression**	**Derived noun**
a.	uk-a ROOT-VFV “wake up”	ma-uka DPF-N.STEM “genital warts”
b.	min-a ROOT-VFV “blow one’s nose”	ma-mina DPF-N.STEM “nasal discharge”
c.	-njenje IDEOPHONE “the way something looks when shaking”	ma-njenje DPF-N.STEM “palsy, shaking of one’s body”

[27] This general lack of corresponding singular nouns is also attested in other Bantu languages (see i.a. Taraldsen, 2010, n. 8).

Stellenbosch University https://scholar.sun.ac.za

Although the uncountable nouns in (34) are derived from words of other categories, there are no singular counterparts to these nouns. This is further evidence that nouns in the *a*-AC do not have a one-to-one relationship with the singular classes.

The third type of *ma*-initial nouns include nouns where the word-initial *ma*- does not show any derivational properties; therefore, this *ma*- may not necessarily be regarded as a number or derivational prefix. These nouns are illustrated in (35).

(35) a. madzi (“water”) b. mate (“saliva”) c. malo (“place”)
d. mantha (“fear”) e. manyazi (“shame”)

As shown in (34-35), it is not entirely correct to identify nouns forming the *a*-AC as involving plural count nouns. The fact that noun-initial *ma*- corresponds to several distinct morphemes has also been observed in some previous studies (see i.a. Fortune, 1970: 97; Welmers, 1971: 13; Hyman, 1980: 180; Maho, 1999: 249). Due to the tendency to associate the form *ma*- with the NC6 prefix, Welmers (1973: 163), for example, proposed that NC6 be split into NC6 and NC6a. In Welmers’ classification, the regular plural prefix *ma*- was proposed to be the noun prefix for NC6, while the inherent *ma*- under discussion in the present section was associated with the non-count (liquid/mass) nouns designated NC6a (see i.a. Fortune, 1970: 97; Welmers, 1971: 13; Hyman, 1980: 180; Maho, 1999: 249). However, these views cannot be sustained for Chichewa where we can clearly see that there are at least five different types of *ma*-initial nominals (see (37) below for a summary) and not just the two identified in Welmers (1973).

In the next section, I consider the fourth type of *ma*-initial nouns, namely manner nominals.

4.5.2.3 Ma-initial manner nominals

As already introduced in Chapter 3, Section 3.3.2.3, in Chichewa and other Bantu languages (e.g. Shona (Fortune, 1985: 101), Zulu (Kunene, 1974)) , there is a set of verbal nouns that are derived from passivized verbal expressions (Bresnan & Mchombo, 1995: 190; Mchombo, 1998: 170, 2004: 115–117). The examples illustrating *ma*-initial manner nominals are repeated in (36).

(36)	**Verb stem**	**Passivized verb stem**	**Manner nominal**
a.	yend-a	yend-edw-a	ma-yend-edw-e
	ROOT-VFV	ROOT-PASS-VFV	DPF-ROOT-PASS-VFV
	“walk”	“be walked”	“manner of walking”

Stellenbosch University https://scholar.sun.ac.za

b.	val-a	val-idw-a	ma-val-idw-e
	ROOT-VFV	ROOT-PASS-VFV	DPF-ROOT-PASS-VFV
	"wear"	"be dressed"	"manner of dressing/ dressing style"

The manner nominalisation process is very regular such that any verb stem can be turned into a manner nominal in the way illustrated in (36). Of interest, in the present study is that these nouns also control AM *a-*, just like the other *ma-* and *a-* initial nouns discussed above.

To summarise what we have seen in Section 4.5.2, then, five different types of nouns associated with the noun-initial form *ma-* can be identified in Chichewa. These are as follows:

(37) a. Animate nouns derived from the derivational prefix *ma-* which control AMs *m(u)-a-* discussed in Section 4.3.6),
b. Plural count prefix *ma-* (illustrated in plural nouns in Section 4.5.2.1),
c. Non-count (liquid/mass) nouns associated with the prefix *ma-* (e.g. 34),
d. Idiosyncratic syllable *ma-* initial nouns (e.g. 35) and
e. The *ma-*initial manner nominals (discussed in the current section, cf. Section 3.3.2.3)

As will become clear in Chapter 5, the derivational morphemes identified in (37a) and (37c) are the same prefix; it is only that some nouns are assigned to the *m(u)-a-*AC on a semantic basis while others are assigned to the *a-*AC on word-initial phonological basis, a pattern observed with many other ACs. I will show that the nasal-vowel symmetry (i.e. *m-* and *a-*) observed in the *a-*AC is also observed with other ACs, namely *u-*AC and *i-*AC, i.e. there appears to be a previously unrecognised general organizational principle in play here.

In this section, I have shown that nouns belonging to the *a-*AC, traditionally associated with NC2/6, exhibit diverse morphological and semantic properties. However, what appears to be the major criterion for controlling AM *a-* is word-initial properties: more specifically, nouns that are *ma-* and prefix *a-*initial.

In the next section, I turn to nouns that control AM *u-*, i.e. those associated with the traditional NC3 and NC14.

4.6 The *u-*AC, which includes NC3 and NC14

The focus in the present section is on nouns that control AM *u-.* The AM *u-* has also been associated with two different traditional NCs, namely NC3 and NC14. The reasons for identifying these as different classes were (i) the different prefixes on the nouns and (ii)

Stellenbosch University https://scholar.sun.ac.za

semantically, NC14 is assumed to comprise abstract nouns. However, I argue that the two criteria for identifying these as two distinct classes are not well motivated for Chichewa. I demonstrate this in the following two subsections.

4.6.1 *M(u)*-initial nouns that trigger AM *u*-, those associated with NC3

According to the dataset, nouns that control the AM associated with the traditional NC3 include both derived and non-derived lexical items. I consider the derived nouns first because they have been extensively described in previous works. As already introduced in Chapter 3, Section 3.3.2, nouns being considered in this group have the prefix *m(u)*- attached to nominalised or inherent nominal stems as shown in (38).

(38)

	Source expression	**Singular**	**Plural**
a.	yes-a ROOT-VFV “measure”	mu-yes-o SG-ROOT-NFV “measurement”	mi-yeso PL-measurement “measurements”
b.	gugu IDEOPHONE “pounding sound”	m-gugu SG-N.STEM “rhythm, sound of thumping feet”	mi-gugu PL-N.STEM “rhythms, etc”
c.	pend-a dzuwa ROOT-VFV sun “aim the sun”	m-pendadzuwa SG-N.STEM “sunflower”	mi-pendadzuwa PL-N.STEM “sunflowers”
d.	-tatu ADJ.STEM “three”	m-tátu SG-N.STEM “a type of traditional game”	mi-tátu PL-N.STEM
e.	-zimu N.STEM	m-zimu SG-N.STEM “spirit/ghost”	mi-zimu PL-N.STEM “spirits/ghosts”

The derivational processes illustrated in (38) have been considered as the defining property of the traditional NC3. The prefix *m(u)*- is viewed as being homophonous with that of NC1, i.e. two distinct forms are postulated, carrying different noun class features (see i.a. van der Spuy (2009: 198) for Zulu; Kishindo (1985) for Chichewa; Alcock & Ngorosho (2004: 8) for Kiswahili). According to Kishindo (1985: 6), the derivation processes for this class in Chichewa involve the attachment of the prefix to verb stems that “denote some kind of

Stellenbosch University https://scholar.sun.ac.za

activity and the resultant nouns have the meaning 'object of V-ing'" (cf. Kunene 1976:152).[28] However, take note that this prefix attaches to a wide range of nominalised stems, e.g. those deriving from adjectives, ideophones, verbs, etc. Although, the previous literature considers the prefix *m(u)-* in this class to be different from that found in the *m(u)-a-*AC, I argue that the nouns from these two different traditional classes use the same singular prefix *m(u)-;* they only come to belong to different classes following the animacy hierarchy. More specifically, nouns that are higher on the animacy hierarchy belong to the *m(u)-a-*AC while those low on the hierarchy belong to the *u-*AC.

Although, the word-initial *m(u)-* is seen to be morphologically attached to the various expressions in (38), there are some nouns that are evidently not derived through prefixation. According to the dataset, some assumed roots do not seem to be intuitively meaningful and they are not seen to participate in any regular and predictable derivation with other prefixes to derive other word categories (39a-b). As for the count nouns, the only observed morphological modification is the number marking. Consider the nouns in (39).

(39)	**Singular**	**Plural** (NC4)
a.	mowa ("beer")	
b.	moni ("greeting(s)")	
c.	moto ("fire")	miyoto ("fires")
d.	mudzi ("village")	midzi ("villages")
e.	muvi ("arrow")	mivi ("arrows")

The major argument for assuming that nouns of the type in (39) contain the prefix *m(u)-* and a stem is the number marking that changes the vowel of the first syllable into an *-i-,* which seems to indicate that the first syllable is a prefix. Even though the pluralization process can be taken as evidence to consider the initial syllable as a prefix, the lack of predictability and regularity of the individual roots does not support the view that all these nouns derive from verbal or adjectival roots. Therefore, the number marking strategy illustrated in (39) could also be analysed as a form of root modification. Alternatively, this could be viewed as a rule that has been overgeneralised to non-inflectional roots. However, whichever analysis one

[28] Contini-Morava & Kilarski (2013: 270) say, with reference to Kiswahili, that this class (*m-/mi-*) nominalizes verbs referring to 'a verbal process' e.g., *mfuo* 'a hammering' (< *-fua* 'to hammer', including nominalizing suffix *-o*).

Stellenbosch University https://scholar.sun.ac.za

adopts, it does not in any way undermine the analysis that I am proposing about the assignment criteria for noun classes in Chichewa.

One unique aspect of the nouns in this sub-group within the *u*-AC is that they are all *m(u)*-initial. Semantically, the class is composed of approximately 93% (583/627) inanimate nouns and only 7% (46/627) animate nouns (see Appendix 11A). Some of the animate nouns are shown in (40). Since the alleged roots to which the singular and plural prefixes attach do not have any known meaning, the examples will not be glossed morpheme by morpheme.

(40)	**Singular nouns**	**Plural nouns**
a.	m-zenga (“civet cat”)	mi-zenga (“civet cats”)
b.	m-lamba (“mudfish”)	mi-lamba (“many mudfish”)
c.	m-leme (“bat”)	mi-leme (“bats”)
d.	m-kunga (“eel”)	mi-kunga (“eels”)

It appears that the kind of animate nouns that fall under this type are mainly the less desirable or small-sized kind. In fact, the human-denoting ones found in the category represented in (40) can also control AMs for the *m(u)-a*-AC without causing ungrammaticality. There are also *m(u)*-initial nouns that alternate between the *u*-AC (≈NC3) and the *m(u)-a*-AC (≈NC1) depending on whether one is referring to non-human or human entity, respectively. Consider the following examples.

(41)	**Noun**	***u*-AC (≈NC3)**	***m(u)-a-AC* (≈NC1)**
a.	mpendadzuwa m-penda-dzuwa SG-monitor-sun	“sunflower”	“timekeeper” (human)
b.	mchizawanthu m-chiza-wanthu SG-heal-people	“food aid”	“healer/doctor” (human)
c.	muongolera mu-ongol-er-a SG-steer-APP-VFV	“seam of mat”	“one who directs, drives” (human)

The pattern illustrated in (41) is predictable such that one would know whether the referent is human or non-human by virtue of the AM.

Another important observation deriving from the dataset is that the traditional NC3 also takes a wide range of deverbal abstract nouns. Some of these nouns are shown in (42).

Stellenbosch University https://scholar.sun.ac.za

(42)	**Verb stem**	**Singular noun**	**Plural noun**
a.	gwiriz-an-a ROOT-REC-VFV “agree”	m-gwiriz-an-o SG-agree-REC-NFV “agreement”	mi-gwirizano PL-agreement “agreements”
b.	pikisan-a ROOT-VFV “compete”	m-pikisan-o SG- compete-NFV “competition”	mi-pikisano PL- competition “competitions”
c.	v-an-a ROOT-REC-VFV “hear”	m-va-n-o SG-hear- REC-NFV “unity”	NON-COUNT

The derivation illustrated in (42) is a predictable abstract nominalisation process in Chichewa; as such I will not consider its statistical distribution. The fact that this class also contains abstract nouns has also been noted for many other Bantu languages (cf. Maho, 1999: 64, 69). These facts, however, undermine the traditional analyses, which designate NC14 as the abstract class. As will become clear in the next section, it does not help to describe nouns associated with NC3 as involving both abstract and concrete nouns. However, one thing that can be stated with a high level of certainty about this sub-class within the *u*-AC is that it comprises nouns that are syllable *m(u)*-initial.

Having discussed the *m(u)*-initial nouns in the *u*-AC, I now turn to another set of nouns that also belongs to the *u*-AC, those associated with the traditional NC14.

4.6.2 *U*-initial nouns that trigger the AM *u-*, those associated with NC14

Previous studies have identified this class as comprising nouns derived by the prefix *u-*, which includes abstract nouns, with the result that it is, as I have noted above, commonly referred to as the *abstract class* (see i.a. Maho, 1999: 56, 77; Matiki, 2001: 72; Mchombo, 2004: 114; Katamba, 2006: 15). However, as I have observed in the foregoing discussion, there are also abstract nouns within the *m(u)*-syllable initial nouns (i.e. NC3). As will be shown in the following sections, this is also the case with some nouns associated with the traditional NC5, NC6, NC7 and NC9. This distribution pattern of abstract nouns in many other NCs has also been observed for many other Bantu languages (see i.a. Worsley, 1954: 286; Maho, 1999: 77). The present section thus effectively provides further evidence that complicates the traditional assumptions that NC14 is an abstract denoting class.

To start with, the dataset had 618 entries for this subgroup. However, there were many nouns that had been wrongly assigned to this class. I explain the misallocation in Section

Stellenbosch University https://scholar.sun.ac.za

4.6.4, below. When the misclassified nouns were removed, the dataset reduced to only 216 entries, making it the smallest traditional NC in the system. The nouns in this group fall into a number of categories, both morphologically and semantically. First, there are nouns that are derived via the prefix *u-*, such as those in (43). Like many other prefixes discussed above, the *u-* attaches to nominalised adjectival, and verbal stems. Nouns derived by this means do indeed denote abstract entities.

(43)	**Source**	**Singular**	**Plural**
a.	bwenzi friend	u-bwenzi SG-friend “friendship”	ma-ubwenzi (NC2/6) PL-friendship “friendships”
b.	-lesi ADJ.STEM	u-lesi SG-ADJ.STEM “laziness”	
c.	tola-nkhani [VPpick news]	u-tolankhani SG-pick news “journalism”	

However, there are many other non-derived nouns that are syllable *u*-initial. Many of them do not denote abstract entities, such as those in (44) below.[29]

(44)	**Singular**	**Plural**
a.	*uta* (“bow”)	ma-uta (NC2/6)
b.	*una* (“hole”)	ma-una (NC2/6)
c.	*ufa* (“flour”)	
d.	onga (“gun powder”)	
e.	ombwe (“poisonous type of beans used for killing fish”)	

As shown in (44), the singular nouns are apparently coincidentally *u*-initial, as they do not show any derivational properties. Also note that some nouns are *o*-initial. Most nouns of the category represented in (44) are concrete. Of the 216 syllable *u-*/*o*-initial nouns in the dataset, 62% (133/216) refer to abstract entities. The remaining 29% (62/216) do not denote abstract entities (see also Appendix 12). Therefore, these abstract features cannot plausibly be the basis on which children assign these nouns to the *u*-AC.

In addition to the syllables *m(u)-*, and *u-* and *o*-initial nouns, there is also a third set of nouns that control the AM *u-*. I introduce these in the next subsection.

[29] For similar patterns in other Bantu languages, see also Mugane (1997: 82–83) and Maho (1999: 76–77).

Stellenbosch University https://scholar.sun.ac.za

4.6.3 Nouns with other word-initial characteristics that trigger AM *u-* (≈NC3/14)

There is a set of 12 nouns starting with the bilabial phonemes */b/, /bw-/*, */w-/,* and another set of 15 nouns with various word-initial properties that also belong to the *u*-AC. Consider a representative sample of these in (45) (see also Appendix 13).

(45) a. bowa (“mushroom”) b. bweya (“fur/wool”)
c. wonga (“gunpowder”) d. fumbefumbe (“shavings, sawdust”)

The nouns in (45) are not accidentally in the *u*-AC as they share the *labial* place feature with *m(u)*-initial nouns and those beginning with *u-* and *o-*. I return to this matter in more specific detail in Chapter 5, Section 5.4.2.

I have so far highlighted three important characteristics of nouns that control the AM *u-* which were previously associated with NC3 and NC14. I am arguing that these two classes do not have any unique semantic properties that would motivate identifying them as different classes. The morphological differences are also not reliable, as I have shown in this section: if the morphological differences are key, I would have to postulate more than three different classes, as there are more than three different word-initial characteristics that form this AC. From the agreement classification perspective, however, it is possible to argue that these two traditional NCs are actually only one class as they all control the same AM, *u-*.

In fact, the connection between NC3 and NC14 has been observed in many previous studies, but it has always been interpreted differently to what is being suggested here. For example, Hyman (1980: 82–83) reports that in the Proto-Western Bantu class, the noun prefix for NC3 was **u-*. Take note that the prefix *u-* is the one that is usually associated with NC14 in the traditional analysis. Schadeberg (1990: 28), in turn, analyses what would be NC14 in Chichewa as NC3 in Umbundu (a zone R Bantu language). From the perspective of the present study, these conflicting analyses simply imply that there is no proper motivation for identifying each of these noun groups as belonging to different noun classes. Maho (1999: 145), however, defends the approach that identifies these two classes as distinct as follows:

> In quite a number of Bantu languages, noun classes 3 and 14 employ identical sets of concords, while retaining separate noun prefixes… . There are also some languages in which these two classes have completely merged, concords as well as noun prefix (notated '3=14'). … Had we looked for target genders and lumped […] [them – PKM] into a single noun class, valuable dialectological data would have been lost in the process.

Stellenbosch University https://scholar.sun.ac.za

In the context of the present study, I argue that Maho's (1999) concerns do not arise since prefixes in these classes are found to be mere derivational or inflectional affixes that do not carry noun class features. However, what these studies confirm is that NC3 and NC14 could be one class in Bantu more generally, and not just in Chichewa. Maho's (1999) views clearly points in that direction.

What appears to be key for the *u*-AC is that it mainly takes inanimate nouns that either have the initial syllable *m*(u)-, or initial labial vowels (*u-, o*-)- or consonants (*b-, w*-). However, according to the dataset, these criteria do not apply to many loan words, which are conspicuously absent in this class. As will become clear in Chapter 5, the cluster of word-initial phonemes */m/, /u/, /o/, /b/* and */w/* are also phonologically related; they fall under [+*labial*], [-*coronal*] noun-initial phonemes which contrast with other ACs whose noun membership is determined by [+*coronal*]-initial phonemes. I reserve this discussion to the relevant Section (5.4), in Chapter 5. Next, I provide an account of the nouns that were allocated to the traditional NC14, which I have identified as belonging to different ACs.

4.6.4 Some misclassified nouns in NC14

The raw dataset excerpted from the *Chichewa monolingual dictionary* had 227 nouns that were classified as belonging to the so-called *U-Ma* class (NC14/6). These nouns were allocated to this class on the basis of the plural form that they take, that is, the prefix *ma*-. As already introduced above, many pedagogical grammar books identify noun classes as involving singular and plural pairings. For this reason, NC14 is mnemonically known as the *U-Ma* class. The labels *U*- and *Ma*- are from the singular (NC14) and plural (NC6) prefixes, respectively. Further note that the *U-Ma* class is taken not to have a plural marker of its own, but to share the plural form with class 5/6 (the so-called *Li-Ma* class). Because plural-based classes are taken to be derived from their singular counterparts, many nouns without singular forms have caused a lot of classification problems in the traditional approach adopted in the *Chichewa monolingual dictionary*.

The second reason for this misclassification could relate to the fact that the AMs for nouns in the *u*-AC (≈NC3 & NC14) are homophonous with the AM for the *m(u)-a*-AC (≈NC1) in

Stellenbosch University https://scholar.sun.ac.za

certain agreement domains, especially on possessive pronouns and adjectives. This is illustrated below.[30]

(46) a. Munthu **wá**-nu **wó**-yera
person (NC1) AM.3RD PER-POSS.HON AM.ASC-white
"Your light complexioned person"

b. Uta **wá**-nu **wó**-yera
bow (NC14) AM.3RDPER-POSS.HON AM.ASC-white
"Your white bow"

c. Fodya **wá**-nu **wó**-yera
tobacco (NC14)AM.3RD PER-POSS.PL AM.ASC-white
"Your white tobacco"

The examples in (46) would suggest that the nouns *munthu, uta* and *fodya* belong to the same AC. On the basis of this evidence, one would assign *uta* or *fodya* (46c) to either NC1 or NC14, which is the case in the *Chichewa monolingual dictionary*. However, these nouns do not share homophonous AMs in other domains, such as, on the numeral, demonstrative, verb SM and relative pronoun, as shown in (47).

(47) a. Muthu **m**-modzi **a**-mene **a**-na-thawa.
person (**NC1**) AM-NUM AM-REL AM-T/A-run
"One person who run away."

b. Uta **u**-modzi **u**-mene **u**-na-thyoka.
bow (**NC14**) AM-NUM AM-REL AM-T/A-break
"One bow that broke."

c. *Fodya (**NC14**) **u**-modzi **u**-mene **u**-na-phukira

d. Fodya **m**-modzi **a**-mene **a**-na-phukira
tobacco (**NC1**) AM-NUM AM-REL AM-T/A-sprout
"One tobacco plant that sprouted."

From the examples in (47), I can show that *fodya* does not fully behave like the NC14 *uta*, but rather it behaves like *munthu* of NC1. To determine the AC for such nouns, one therefore has to consider a full range of agreement domains. For this reason, all the nouns of the type in (47d) were tested specifically on the non-homophonous domains. After doing this, I found that 95 entries that were classified as belonging to NC14 actually belong to the traditional

[30] Recall that the forms *wa-* and *wo-* are a result of coalescence between the AM *a-* / *u-* and the ASC *á* (*u+a = wa, u+o = wo, a+a = wa, a+o = wo*), as introduced in Section 4.3.7.

Stellenbosch University https://scholar.sun.ac.za

NC1 (see Appendix 14). Some of these nouns were classified as belonging to both NC1 and NC14, especially the *ka*-initial nouns and some letters of the alphabet.

Concluding the discussion focusing on the *u*-AC: I have shown that the traditional NC3 and NC14 are one class as there is no reliable semantic or morpho-phonological basis for identifying them as two or three distinct classes. The major criteria for nouns in this AC are that they are inanimate nouns that begin with labial phonemes, such as *m(u)-, mo-*, *u-, o-, b-, w-*, etc. Take note of the emerging pattern regarding nasal-vowel symmetries: in the *a*-AC there is nasal *m-* and vowel *a-*; in the present section, there is nasal *m-* and vowel *u-.*

In the next section, I consider nouns that control AM *i-,* which has been associated with NC4 and NC9 in the Bleek-Meinhof schema.

4.7 The *i*-AC, which includes NC4 and NC9

The *i*-AC comprises nouns from two distinct traditional NCs, namely NC4 and NC9. On the one hand, NC4 is traditionally understood to include plural nouns which are counterparts to singular nouns of NC3. On the other hand, NC9 is assumed to comprise singular nouns which are homorganic nasal-initial.[31] In addition to these traditionally identified noun characteristics (i.e. N*, mi-*), the dataset also has nouns with other word-initial characteristics that control AM *i-*, such as nouns that begin with the following phonemes: vowels (*/i/* and */e/*)*;* consonants (*/s/, /t/, /d/*, *etc.*) and the syllable */me/.* Each of these noun-types will be introduced in separate sections below.

4.7.1 The plural prefix *mi-* and *me-* initial nouns

According to the dataset, there are two types of *mi*-initial nouns in Chichewa. The first type includes nouns where *mi-* is a plural prefix. The second type is where the noun-initial *mi-* is not a plural marker but an inherent word-initial syllable. It is this latter group where we also find *me-* initial nouns. The first type of *mi*-initial nouns has been the major identification basis for nouns belonging to the traditional NC4, while the *me*-initial nouns were not accounted for at all in the Bleek-Meinhof system. As described in Section 4.6.1, nouns associated with

[31] A homorganic nasal is a "nasal that adapts phonologically to an immediately following consonant" (Maho, 1999: 59).

Stellenbosch University https://scholar.sun.ac.za

the traditional NC3 take the prefix *mi-* to mark plural as shown in (48). Due to their predictable nature, these nouns are not lemmatised in the dictionary.

(48)	**Plural nouns (NC4)**	**From singular nouns (NC3)**
a.	mi-tengo (“trees”)	mtengo (“tree”)
b.	mi-nda (“gardens”)	munda (“garden”)
c.	mi-yoyo (“lives”)	moyo (“life”)

As shown in (48), the plural nouns are derived from singular nouns, either through prefixation or root modification. However, this does not imply that all *mi*-initial nouns are plural nouns. According to the dataset, some of the nouns have singular readings, while others are uncountable or mass nouns which serve both singular and plural readings. The dataset registered 2% (30/1453) entries of this type, some of them are represented in (49) (see also Appendix 15).

(49)	**Singular/uncountable (*i*-AC)**	**Plural counterpart**	
a.	mimba (“abdomen, pregnancy”)	mimba (“abdomen”)	(NC8/10)
b.	micheta (“immature maize”)	micheta (“immature maize”)	(NC4)
c.	mitala (“polygamy”)	UNCOUNT	
d.	mita (“meter”)	ma-mita (“meters”)	(NC2/6) Eng.LW
e.	minibasi (“minibus”)	ma-minibasi (“minibuses”)	(NC2/6) Eng.LW
f.	mendulo (“medal/pile of logs”) Eng.LW		

As shown in (49a-c), the nouns do not change form whether they are referring to singular or plural entities. However, what may change are the AMs: in the case of (49a), when denoting plural entities, the noun belongs to the *zi*-AC (≈NC8/10). However, in (49b), the noun does not change noun class whether singular or plural. In (49d-e), the coincidentally *mi*-initial singular nouns of foreign origin also control the same AM as the *mi*-initial plural nouns. Take note also that the English loanwords are pluralised using other means, i.e. by the prefix *ma-*, with the consequence that the pluralised noun eventually belongs to the *a*-AC. As I will show in the next subsection, the AM *i-* does not disambiguate between singular and plural. For example, consider a sentence with a covert subject in (50).

(50) **i**-kuon-ek-a
AM-see-STATIVE-VFV
“**It/they** are visible.”

In (50), the AM may be controlled by a singular noun e.g. *mbuzi* (“goat”) or a plural noun e.g. *mitengo* (“trees”) or an uncountable noun e.g. *mimba* (“abdomen”). This pattern is also observed in other ACs, such as *a*-AC.

Stellenbosch University https://scholar.sun.ac.za

Next, I consider nouns that also trigger the AM *i*-, but are homorganic nasal-initial.

4.7.2 Homorganic nasal-initial nouns (NC9)

Pioneer studies in Bantu noun classification have associated this noun class with reconstructed prefixes **n*- (Meeussen, 1967: 97), **ny*- (Guthrie, 1971), **thin* (Bleek, 1869: 282), **li-ni* (Meinhof & Van Warmelo, 1932: 39) and **li-ne*- (Welmers 1973) (cf. Maho, 1999: 171; Batibo & Kgolo, 2016: 21–22). However, modern studies view the class as having the reduced prefix forms N- or null (Ø-) (cf. Maho, 1999: 59; Katamba, 2006: 109; Choti, 2015; Batibo & Kgolo, 2016: 21).[32] The upper-case notation signifies the underlying form which is manifested as a nasal prefix *n*-, Ø, or homorganic **n+stop consonant-** or **m+stop consonant-**initial syllables (cf. Alcock & Ngorosho, 2004: 9). The various manifestations depend on language-specific phonological rules and noun-particular surrounding phonological environments. These nouns are illustrated on the basis of a two-stage derivation process in (51).

(51)

	Source	**Prefixation stage**	**Homorganic stage**
a.	gow-a	N-gów-e	ngowe
	ROOT-VFV	DPF-N.STEM	N.STEM
	“pull down”	“hooked stick used for plucking fruits/ firewood”	
b.	thir-ir-a	N-thíríra	nthíríra
	ROOT-APP-VFV	DPF-N.STEM	N.STEM
	“irrigate”		“irrigation”
c.	kuman-a	N-kuman-o	nkhumano
	ROOT-VFV	DPF-meet-NFV	N.STEM
	“meet”		“meeting/ meeting place of lines, rivers”
d.	badw-a	N-bádwa	mbádwa
	ROOT-VFV	DPF-N.STEM	N.STEM
	“be born”		“citizen/ descendant”
e.	pats-a	N-pats-o	mphátso
	ROOT-VFV	DPF-give-NFV	N.STEM
	“give”		“gift”

[32] According to Hyman (1980: 193) it is assumed that the nasal prefixes are a later innovation of some Bantu languages. According to Meinhof (1932), the bilabial /m/ and the underspecified /N/ prefixes are diachronic relics of Pro-Bantu syllables /mu/ and /ni/, respectively (Choti, 2015: 37).

Stellenbosch University https://scholar.sun.ac.za

As shown in (51) the prefixation of the nasal follows the same pattern as shown in the other classes above. The nasal also attaches to an already nominalised stem, namely those belonging to the Type 1 nouns discussed in Chapter 3, Section 3.3.2.

Motivation for postulating an underlying N- in cases such as (51d-e) comes from a phenomenon called *nasal place assimilation* which is also observed in many other languages. For example, the distribution of the English negative prefixes, *in-* and *im-*, such as those in (52), is taken to be an instance of allophonic variation, where "the labial [m] shows up before a labial ([p] in *impossible*), and the alveolar [n] shows up before the alveolar ([t] in *intolerant*)" (Roca & Johnshon, 1999: 92). This explains why **imtolerant* and **inpossible* are unacceptable in (52e) and (52g), respectively.

(52)	a.	intolerant	e.	*imtolerant
	b.	indefinite	f.	*imdefinite
	c.	impossible	g.	*inpossible
	d.	imbalance	h.	*inbalance

Considering neutral environments, e.g. where the adjectival stem starts with a vowel as in *in-ability, in-operable*, it is possible to determine that the prefix form is the alveolar [n]. In this regard, [n] is taken to be the lexical form of the prefix (Roca & Johnshon, 1999: 93).

There are at least three pieces of evidence to support the hypothesis that /n/ is also the lexical form of the homorganic nasal prefix associated with NC9 in Chichewa. Firstly, as was discussed in Sections 4.3.2 and 4.6.1, some nouns in the *u*-AC are derived via the prefix *m(u)-.* One interesting property of the prefix in these classes is that it does not trigger nasal assimilation and concomitant stop aspiration. Consistently, the *m-* is either realised as a syllabic nasal or there is vowel insertion to give *mu-, ma-* or *mi-* in *m(u)-a*-AC (≈NC1), *a*-AC (≈NC2) or *i*-AC (≈NC4), respectively.[33] This is not the case with the homorganic nasal prefix; here the nasal prefix *m-* is accompanied by homorganic coalescence and optional aspiration. The nasal assimilation and aspiration are spin-off effects of a sound that readjusts to suit the following alveolar and labial stop phonemes. This may also account for the fact that the homorganic nasal and the syllabic nasal are minimal pairs, as shown in (53) below. The nouns to the left in (53) are nouns with a homorganic nasal prefix (i.e. derived from underlying nasal *N*) and those to the right are nouns derived from a syllabic nasal */m/*.

[33] See Chapter 5, Section 5.4 for specific details regarding the nasal-vowel symmetries

Stellenbosch University https://scholar.sun.ac.za

Note that there is no semantic relationship between the two groups of nouns illustrated in (53).

(53)	**Homorganic nasal prefix (*i*-AC)**	**Syllabic nasal prefix**
a.	mba.le (“plate”)	m-ba.le (“a relation of somebody”) **(NC1)**
b.	mbu.li (“ignorant person”)	m-bu.li (“a second exit of a mouse hole”) **(NC3)**
c.	mphi.ra (“rubber”)	m-pi.ra (“ball”) **(NC3)**
d.	mpha.sa (“mat made out of reed”)	m-pa.sa (“a type of fish”) **(NC3)**
e.	mphi.ni (“incision skin marks”)	m-pi.ni (“handle e.g. of a hoe”) **(NC3)**

As shown in (53), the nasal *m*-initial nouns in NC1 and NC3 do not trigger homorganic nasal coalescence despite being followed by stops.

The second piece of evidence comes from the fact that all **309** *m*-homorganic-initial nouns in the *i*-AC are followed by the stops */b/, /p/, /f/* and */v/*. This supports the claim that the */m/* is a result of place of articulation nasal assimilation and that this assimilation is not induced by the bilabial nasal /m/ itself.

The third piece of evidence comes from cross-linguistic evidence. Although the present study largely focuses on the grammar of Chichewa, the fact that previous studies have consistently observed that NC9 comprises nouns bearing the initial nasal */n/* or *null* in other Bantu languages could be evidence to support the hypothesis that the underlying form of the NC9 prefix is /n/ (see i.a. Hyman, 1980: 199; Maho, 1999: 59; Katamba, 2006: 109). In the absence of this hypothesis, one might, like Hyman (1980: 205), wonder “whether a homorganic nasal might have been a nominalizer.” However, this view cannot be sustained considering the predictable derivational processes illustrated in (51) above.

Although the three strategies of deriving nouns in this class are shown to be predictable, as illustrated in (51), they do not apply to all nouns in this class. There are many nouns, especially the *n*-initial type, where the nasal-initial syllable does not show any evidence of being a derivational or inflectional prefix. Some of these nouns are shown in (54).

(54) a. njala (“hunger”)
b. njati (“buffalo”)
c. nungu (“porcupine”)
d. nunsu (“part of something that is worn out”)
e.

nol-o	from the verb	nol-a
sharpen-NFV		ROOT-VFV
“whetstone”		“sharpen”

Stellenbosch University https://scholar.sun.ac.za

Take note that the initial *n-* in (54e) is part of the verb stem and that the noun is derived by means of the NFV *-o*. Suspicions that some of these nouns are not obtained by productive processes are also expressed in some previous studies. For example, Herbert (1978: 135) argues that the synchronic status of NC9 in some languages is a result of language change which has seen the prefix being "incorporated into the stem due to the non-alternating nature of these forms" such that the old prefixes are now part of the noun stem (cf. Bates, 1926: 10).[34] Worth noting, however, is the fact that Herbert's claim is made within a traditional analysis of the kind that I am in the process of showing cannot be supported.

As regards the classification criterion of the homorganic nasal-initial nouns into the *i*-AC, many studies claim that NC9 takes animate nouns (see i.a. Maho, 1999: 51; Demuth, 2000: 6; Katamba, 2006: 114–116). However, it is also generally accepted that the common factor connecting the nouns in NC9 is the nasal element which often forms a homorganic relationship with the first consonant of the stem. In addition to homorganic nasal-initial nouns, I observe that the *i*-AC also comprises nouns that are vowel *i*-initial. I consider vowel *i*-initial nouns in the next section.

4.7.3 Vowel *i*-initial nouns

The *i*-AC also contains nouns that are vowel *i*-initial. Like the homorganic nasal, the vowel *i-* occurs either as a derivational prefix or as an inherent word-initial syllable, as evidenced in some loan words. In some studies, the vowel *i-* has been analysed as an epenthetic vowel (EV) which attaches to CV stems as a strategy to achieve word minimality conditions.[35] Consider the following examples.

(55)		**Source**	**Singular**	**Plural**
	a.	-f-a	i-m-fa	imfa (NC10)
		ROOT-VFV	EV-PX-die	"deaths"
		"die"	"death"	

[34] For other accounts of language change regarding nasal prefixes, see Guthrie (1967), Bell (1972) and Batibo & Kgolo (2016).

[35] It is generally observed that many languages require that words be of a certain minimum size – e.g. bisyllabic, bimoraic, etc. (McCarthy & Prince, 1996: 1). The word minimality condition is very common in Bantu languages and is achieved by different strategies, such as vowel epenthesis (see i.a. Mutaka & Hyman, 1990; Downing, 1994; Carstens, 1997: 380; Mkochi, 2009). The general fact that Bantu has a minimal-word constraint also points us to the fact that phonological (prosodic) considerations are likely to play a key part in the grammar-defining systematic regularities that the child will need to learn.

Stellenbosch University https://scholar.sun.ac.za

b.	-mpso N.STEM	i-mpso EV-N.STEM “kidney”	impso (NC10) “kidneys”
c.	-vi	imvi “grey hair”	imvi (NC10) “grey hairs”
d.		inswa “flying termite”	inswa (NC10) “flying termites”
e.		injini “engine”	injini (NC10) / ma-injini (NC2/6) (Eng.LW) “engines” / PL-engine
f.		intaneti (Eng.LW) “internet”	

As shown in (55a) the *i-* is prefixed to a dependent derived nominal stem. The prefix *i-* also attaches to monosyllabic noun stems so as to satisfy word minimality conditions, which is clearly the case in examples (55a -c) (see i.a. Kanerva, 1990: 40; Carstens, 1997: 380), a phenomenon that I argue also applies to other singular prefixes in the other ACs. However, in the other ACs, the singular prefixes are not restricted to monosyllabic stems, which implies that dependency is not always based on syllable count but could be motivated by some other factors, for example morpho-syntactic or semantic.

Although the previous literature did not consider these nouns in any systematic manner, the presence of the prefix *i*-initial nouns in the *i*-AC is instructive. Firstly, the morpheme *i-* is also the AM for this AC. As shown in the other ACs above, the AMs also participate in deriving or inflecting a subset of nouns in the respective ACs (see also Chapter 3). Secondly, as introduced above, this is the third nasal-vowel pairing symmetry (*N, mi, i, e*), in addition to the (*m, mu, mo, u, o*) and (*ma, a*). In Chapter 5, I will show how these clusters form three phonologically determined noun classes, which are further distinguished along the phonological features [+coronal], [-labial], [-back], [-round] and [-coronal], [+labial], [+back], [+round]. This pattern also shows that if the traditional NC3 and NC14 are different classes, then equally the traditional NC9 should be split into two NCs, one taking the *N*-initial and the other *i*-initial nouns. However, as I am arguing, it is not insightful to determine NCs based on the noun’s morphological characteristics alone.

In the next section, I outline another set of examples of nouns that control AM *i-* but have a range of word-initial characteristics.

Stellenbosch University https://scholar.sun.ac.za

4.7.4 Nouns with other word-initial characteristics in the *i*-AC

In addition to the nouns that are *N-, mi-, me-,* and vowel *i*-initial, the *i*-AC also contains many other nouns with a variety of word-initial characteristics, as illustrated in Table 3.6 below (see also Appendix 16). The less regular sets are mainly composed of loan words, most of which optionally control AMs *m(u)-a-* (≈NC1) and *li-* (≈NC5).

Table 4.4: Other systematic word initial characteristics

Noun initial characteristics	Examples	Frequency	Percentage
s-initial	sawa (“ground nut”) sayansi (“science”) (Eng.LW)	**96**	**6.8%**
t-initial	tangani (“kidney”) tepi (“tape”) (Eng.LW)	**34**	**2.4%**
k-initial	kampani (“company”) (Eng.LW)	**20**	**1.4%**
b-initial	basi (“bus”) (Eng.LW)	**16**	**1.1%**
f-initial	filimu (“film”) (Eng.LW)	**16**	**1.0%**
g-initial	galimoto (“car”) (from English ‘car motor’)	**10**	**0.7%**
h-initial	habu (“hub”) (Eng.LW)	**10**	**0.7%**
j-initial	jekeseni (“injection”) (Eng.LW)	**9**	**0.6%**
p-initial	phundudwa (“sloughed off snake’s skin”)	**9**	**0.6%**
a-initial	ambulasi (“ambulance”) (Eng.LW)	**8**	**0.6%**
w-initial	wotchi (“watch”) (Eng.LW)	**5**	**0.4%**
l-initial	lole (“lorry”) (Eng.LW)	**5**	**0.3%**
y-initial	yembe (“mango”)	**5**	**0.3%**
v-initial	vesiti (“vest”) (Eng.LW)	**4**	**0.3%**
e-initial	epuloni (“apron”) (Eng.LW)	**3**	**0.2%**
Total		**250**	**17.6% (250/1416)**

As shown in Table 4.4, the distribution of some of these word-initial characteristics appears to be systematic, especially *s-* and *t*-initial sets. Although the less frequent ones appear to be less systematic, I will show in Chapter 5 that their occurrence here is not necessarily accidental, but does satisfy some underlying phonological criteria, for example possession of [+coronal] and [+high] features.

Stellenbosch University https://scholar.sun.ac.za

Previous studies have claimed that there is a relic of semantic content in the prefixes and that NC9 contains the names of most animals (see i.a. Bell, 1972: 33; Herbert, 1978: 125; Batibo, 1987: 4; Batibo & Kgolo, 2016: 27). However, according to my dataset, only 19% (262/1416) of nouns that control AM *i-* have animate denotation. As I will show in Chapter 5, the word-initial phonological characteristics of these nouns seem a much more plausible basis for their co-classification - more specifically, the relevant elements appear to be nouns that begin with the following phonemes */mi-/,* homorganic nasal /N/, */i-/*, */s-/, /t-/,* etc. At present, I can only observe that these word-initial phonological characteristics appear to be key. Therefore, it may not necessarily be correct to claim that the animate-denoting nouns are assigned to this class on a semantic basis. Rather, any noun that satisfies these phonological properties is assigned to this class, irrespective of its meaning. I return to this matter in Chapter 5.

In the next section, I turn to yet another AC, the *chi*-AC.

4.8 The *chi*-AC, which includes NC7

In the dataset, the *chi*-AC has **1075** entries. In agreement with the previous literature, I observe that all the nouns in this class are *ch*-initial (see i.a. Corbett & Mtenje, 1987: 4; Matiki, 2001: 66; Mchombo, 2004: 6). The morphosyntactic and semantic properties of this class are quite diverse, however. There are three broad types of nouns associated with the form *ch-*. The first type comprises nouns formed from the derivational-cum-number prefix *chi-,* which derives nouns from nominalised adjectival and verbal stems/phrases. The second type includes what in the present study I consider to be non-derived nouns, which happen to be coincidentally *ch*-initial. The third type is the evaluative prefix *chi-* which derives both denotative augmentatives and connotative meanings related to speaker-hearer perspectives, such as pejoration, amelioration, etc. In the present chapter, I will mainly focus on the first two, as the evaluative *chi-* is discussed in detail in Chapter 6, Section 6.4. Each of these three broad types also exhibit a range of group-specific morphological and semantic properties.

4.8.1 Nouns derived with involvement of the prefix *chi-*

In nouns that are derived via the prefix *chi-,* the prefix functions either as a number marker or as a derivational morpheme or as both. Consider the examples in (56) where the prefix

Stellenbosch University https://scholar.sun.ac.za

chi- is shown to derive singular nouns from nominal, adjectival, verbal and ideophonic expressions.

(56)

	Source	**Singular (NC7)**	**Plural (NC8)**
a.	-nthu N.STEM “being”	chi-nthu SG-N.STEM “thing”	zi-nthu PL-being “things”
b.	Chewa tribe of the Chewa	chi-Chewa SG- tribe of “Chewa language/culture”	zi-Chewa PL-Chewa “varieties of Chewa language”
c.	sul-a ROOT-VFV “mint”	chi-t-súl-o SG-mint-NFV “metal”	zi-tsúlo PL-N.STEM “metals”
d.	dul-a ROOT-VFV “cut”	chi-dúle SG-N.STEM “summary, short cut”	zi-dúle PL-N.STEM “summaries, etc.”
e.	-dy-a ma-kanda ROOT-VFV PL-baby “eat babies”	chi-dyamakanda SG-eat babies “sugar daddy”	zi-dyamakanda PL-eat babies “sugar daddies” (Lit. “babies eater”)
f.	wewe IDEOPHONE “the way dogs cry”	chi-wewe DPF-N.STEM “rabies”	UNCOUNTABLE

Like the derivational processes observed in the other ACs, in (56a) the prefix *chi-* attaches to the nominal stem to derive singular nouns. The prefix may also attach to stems that have been nominalised by conversion, tone or FV modification, as shown in (56c-e). In (56f), the prefix is attached to the nominalised onomatopoeic expression *wewe.*

In addition to the *chi-* prefix deriving nouns for this class, there are also other nouns that are derived by the prefix *ch(a/o)-*. This is the same prefix that was discussed in Section 4.3.8, which dealt with nouns that control AMs *m(u)-a-* (≈NC1). The major difference in classification is that those in the *m(u)-a-*AC have animate meaning while those in (57) have inanimate meanings. Consider the following examples.

(57)

	Source	**Singular derived noun**	**Plural derived noun**
a.	met-er-a ROOT-APP-VFV “shave with”	chó-métera SG.ASC-N.STEM “shaver”	zó-métera PL.ASC-N.STEM “shavers”

Stellenbosch University https://scholar.sun.ac.za

b.	ku-mw-a	chá-kumwa/ chó-kumwa	zá-kumwa (NC8)
	INF-ROOT-VFV	SG.ASC-to drink (“something to drink”)	PL.ASC-to drink
	“drink”	“drink”	“drinks”
c.	bwalo	chá-bwalo	NO PLURAL
	court/arena	SG.ASC-court/arena (“of court”)	
		“fine paid to the chief for being in contempt of the chief’s court”	
d.	mamba	chá-mamba	NO PLURAL
	“scales”	SG.ASC-scales (“something with scales”)	
		“a type of plant with hairy leaves”	

The prefixes *cho-* and *cha-* are also the least constrained prefixes for deriving names of novel objects or something that one does not know the name of while knowing its function – the function word can simply be prefixed by *cho-* or *cha-.* Take note that the nouns derived in (57c-d) do not change form between singular and plural; hence the absence of plural forms. However, when the derivation involves mass nouns, the prefix takes the plural form only, as shown in (58).

(58)	**Verb stem**	**Derived noun**
a.	bal-a	zó-bala
	ROOT-VFV	PL.ASC-breed
	“breed/produce”	“fruits” Lit.: “those that are bred”
b.	kamb-ir-an-a	zó-kambirana
	ROOT-APP-REC-VFV	PL.ASC-talk to each other
	“talk to each other”	“deliberations” Lit.: “those that are talked to each other”

Such nouns do not trigger AM *ch-* (≈NC7), but its plural counterpart, *zí-* (≈NC8). The allomorphic variation of *chi-, chá-* and *chó-* is also manifested in the plural forms as *zí-, zá-* , and zó-, respectively. As was mentioned in Section 4.3.8, the vowel on *chá-* coalesces with ASC marker á (“of”) or any other vowel which forms part of the source stem; hence the variations.

Also, of specific interest to the present study is the fact that this AC, like the *u-, li-* and *a-* ACs discussed above, also contains several abstract nouns. The abstract nouns in the *chi-* AC follow the same derivational process as the other concrete nouns, as shown in (59).

(59) a.	-lendo	chi-lendo	b.	gwiriz-an-a	chi-gwirizan-o
	N.STEM	DPF-N.STEM		ROOT-REC-VFV	DPF-agree-NFV
	“foreignness, strangeness”			“agree”	“unity”
c.	kok-a	chi-kóka			
	ROOT-VFV	DPF-N.STEM			
	“pull”“attraction”				

Stellenbosch University https://scholar.sun.ac.za

The presence of abstract noun formation processes here further challenges the analyses that postulate a separate abstract class, NC14. In summary, the derived nouns contained in the *chi*-AC are thus either derived by the prefixes *chi-, chá-* and *ch*ó-. Semantically, the derived nouns are mainly inanimate, and they can also have concrete or abstract denotation.

4.8.2 Simplex *ch*-initial nouns that control AM *chi-*

Although the derivation processes in this class are very productive and predictable, some of the nouns do not show any derivational properties. Consider the examples in (60) below, which do not deliver meaningful roots, and which can also be seen not to participate in other processes if the nouns in question are divided into prefix-stem components.

(60)	**Singular**	**Plural**
a.	chisa (“nest”)	zisa (“nests”)
b.	chitsa (“tree-stump”)	zitsa (“tree-stumps”)
c.	chaka (“year”)	zaka (“years”)
d.	cheya (“body hair”)	
e.	chuku (“moulds”)	
f.	chonde (“soil fertility”)	

Note that I have also highlighted similar patterns with some nouns that control AMs, *u-, li-* and *a*-above. I also consider the plural marking in (60a-c) to involve root modification rather than prefixation, as was the case with examples (28) and (39) in sections 4.4.4 and 4.6.1, respectively. As shown in (60d-f), some nouns do not even inflect for number. It should be reasonable therefore to argue that not every noun is divisible into prefix and stem in this group; therefore, the word-initial *ch-* is not always a prefix as is generally assumed.

All nouns illustrated in this section are *ch*-initial and also trigger AM *chi-.* So far, this is the first AC where all its nouns control what is popularly known as *alliterative agreement*. The other ACs considered so far only have subsets of nouns that are alliterative. The consistency of the *ch*-initial nouns in the *chi*-AC suggests that the word-initial form is the major criterion for nouns to be assigned to this class.

Semantically, the *chi*-AC has 89% (953/1075) inanimate-denoting nouns. In addition to that, the animate nouns in this class are mainly those with some special characteristics such as small insects (e.g. wasps, termites), humans and animals with undesirable or superhuman features (e.g. drunkard, very rich person, etc.; see Appendix 17). Just like the *li*-AC, the *chi-*

Stellenbosch University https://scholar.sun.ac.za

AC can be said to only accept animate entities that are placed lower in the Chichewa animacy hierarchy.

According to the dataset (see Table 4.1), the six ACs considered so far are the only ones with a sizeable number of entries. The remaining six ACs have very few entries (i.e. lexicalised items) not exceeding 16. This quantitative difference in the lemmatised entries in the dictionary also coincides with observations made in the previous literature regarding *primary* and *secondary prefixation*: in the present study, I claim that the former refers to lexicalised items while the latter references morphosyntactically productive expressions. I also observe that the latter group include a variety of expressions including those that are not usually associated with the nominal category, such as CPs, PPs. As I will show below (Section 4.15), these ACs should not be considered at the same level as the ACs comprising clear-cut nominal expressions. Carstens (1997: 370) also observes this distinction and describes these expressions as representing *non-core phenomena* in relation to the NC system. Although the expressions from these ACs will be considered in more detail in Chapter 5, I briefly introduce each in the present chapter.

In the next section, I consider the *zi*-AC, which is associated with the traditional NC8 and NC10.

4.9 The *zi*-AC, which includes NC8 and NC10

The AM *zi-* is triggered by at least six different types of expressions, namely (i) plural nouns that can be associated with traditional NC8, NC10 and NC4, which are plural counterparts of singular NC7, NC9 and NC3, respectively; (ii) mass/uncountable *zi*-initial nouns, (iii) complementiser phrases (CPs); (iv) conjoined expressions with conflicting noun class features; (v) unspecified agreement triggers where the AM serves expletive-like functions and (vi) plural evaluative expressions that convey meanings such as augmentative, pejorative, endearment, etc. These six agreement trigger-types will be illustrated in three groups below.

First, the plural nouns falling into this class have a variety of morphological features: some bear the plural prefix *zi-* while others do not change their morphological shape. Those that take the prefix *zi-* are associated with the traditional NC8 while those that do not change form are associated with the traditional NC10. Consider the following examples.

Stellenbosch University https://scholar.sun.ac.za

(61)	**Plural noun (*zi-* AC)**		**From singular noun**	
a.	zi-tsulo (“metals”)	NC8	chi-tsulo (“metal”)	NC6
b.	zi-zimba (“charms”)	NC8	chi-zimba (“charm”)	NC6
c.	minga (“thorns”)	NC10	minga (“thorn”)	NC4
d.	mimba (“pregnancies”)	NC10	mimba (“pregnancy”)	NC4
e.	nyama (“animals”)	NC10	nyama (“animal”)	NC9
f.	mphuno (“noses”)	NC10	mphuno (“nose”)	NC9

All the plural-denoting nouns in (61) trigger AM *zi-.* Note that their singular counterparts are from three different traditional classes, namely NC4, NC6 and NC9. Apart from examples (61a-b), the other examples do not change their morphological shape to mark plural, which implies that number marking in these nouns is morphologically ambiguous. From the present section onwards, I will be showing that the prefix *zi-* is the default AM in Chichewa; hence its association with nouns that are morphologically ambiguous between singular and plural, conjoined NPs with conflicting class features and other underspecified agreement triggers.

Second, in the present study, the *zi-*AC is also observed to contain nouns that are not derived from a singular form. However, there are only 0.2% (14/5828) entries of nouns of this type. Specifically, these include mass and uncountable nouns, such as the following.

(62) a. zambi (“blame, sin”)
b. zewe (“slowness”)
c. zi-kanda
DPF-scratch
“scabies”
d. zó-khoma-khoma
DPF.ASC-harmer-harmer
“carpentry”

The nouns in (62) do not have singular forms as they are all mass/uncountable nouns. Therefore, they differ significantly from those in (61).

In addition to being triggered by overt nouns with particular characteristics, the AM *zi-* is also triggered by CPs and furthermore commonly occurs as the default AM where the agreement features of the trigger are conflicting or underspecified. Consider the examples in (63).

(63) a. [CP **Za**-kuti mu-bwer-a] **za**-dziw-ik-a tsopano.
ASC-that 2ND PRS.SG-come-VFV SM-T/A-know-STAT-VFV now
“That you are coming (it) is known now.”

b. [**Munthu** ndi **chiwala**] **zi**-ku-thawa.
person (**AM = a**) and grass hopper (**AM = chi**) SM-T/A-run away
“A person and a grasshopper are running away.”

Stellenbosch University https://scholar.sun.ac.za

c. **Zi-**dze pano ndi zá-tonse.
SM.T/A-come here is AM.ASC-3RD PRS PL
"What will come here is for all of us"
Idiom: "We will all be responsible for what will happen here."

d. **Zi**-ku-oneka kuti kugwa mvula.
AM-T/A-appear that T/A-fall rain
"It appears that it is going to rain."

In (63a) the subject of the sentence is the CP *zakuti mubwera* while in (63b) the subject is the conjoined NP *munthu ndi chiwala*, which includes nouns that control different AMs (*m(u)-a-* and *chi-*); as such the verb takes the 'default' AM *zi-*. In (63c-d), the subject is not specific and the AM *zi-* functions like a dummy affix. Although Bresnan & Kanerva (1989: 12) argue that there are no expletive or impersonal subjects in Chichewa, I will show that the AMs *zi-, ku-* and *pa-* optionally serve as expletive AMs. Expletive-like AMs have been observed to exist in other Bantu languages (see i.a. Buell, 2007: 109; Marten, 2010: 254–255; van der Spuy, 2014: 64).

As shown in the preceding discussion, the *zi-*AC differs in interesting ways from the other ACs. First, the AM for this class functions as the default, implying that it is not triggered by some specific controller type. Second, singular *zi-*initial nouns do not control this AM, which again implies that these expressions are not assigned to the *zi-*AC based on their word-initial phonological feature as is the case with the other classes discussed above. Rather, it functions as a default AM (see Chapter 6, Section 6.5). Third, there are very few entries in the dictionary of nouns that control the *zi-*AM, the only entries attested are nouns that have been lexicalised from the regular plural marker *zi-*.

The only case where the assignment of expressions is based on clear-cut word-initial properties involve the sixth type, namely plural augmentatives (AUGs). Consider the following.

(64) a. zi-a-na
PL.AUG-PL-child
"big children"

b. zi-mitengo
PL.AUG-PL.tree
"big trees"

Plural augmentative nouns like those in (64) constitute the only semantically predictable component of the *zi-*AM. I defer more detailed discussion of these nouns until Chapter 6.

In summary, there are six different types of expressions associated with the AM *zi-*, while the Bleek-Meinhof system has only two different classes, namely NC8 and NC10. However,

Stellenbosch University https://scholar.sun.ac.za

as I have tried to show in this section, there is no motivation for distinguishing the six groups I have considered in turn. I therefore maintain that all the six groups belong to one AC.

Next, I consider the second morphologically predictable AC, which is associated with the traditional NC12.

4.10 The *ka*-AC, which includes NC12

There are two different types of expressions that are observed to trigger the AM *ka*-. The first group includes nouns that are traditionally known as diminutive (DIM) nouns, specifically singular DIMs, which have been associated with NC12. Although the DIM nouns are considered in detail in Chapter 6, example (65) serves to illustrate what *ka*-initial DIM nouns look like in Chichewa.

(65)	**Noun**	**Singular DIM**
	mwana	ka-mwana
	child	SG.DIM-SG.child
		“small child”

In (65), the noun *mwana* of the *m(u)-a*-AC is marked with the singular DIM with the consequence that the noun changes class from *m(u)-a*-AC to *ka*-AC. As shown in (65), the other common morphological property of the more regular classes is that the prefixation is largely additive and that the prefixes select full NPs.

The second type of expression in the *ka*-AC are the *ka*-initial manner nominals. These nouns take the same form and give the same interpretation as the *ma*-initial manner nominals that were discussed in Section 4.5.2.3 above. Consider the following examples.

(66)		**Verb stem**	**Passivized verb**	**Manner nominal**
	a.	yend-a	yend-edw-a	ka-yend-edw-e
		ROOT-VFV	ROOT-PASS-VFV	DPF-ROOT-PASS-VFV
		“walk”	“be walked”	“manner of walking”
	b.	-dy-a	dy-edw-a	ka-dy-edw-e
		ROOT-VFV	ROOT-PASS-VFV	DPF-ROOT-PASS-VFV
		“eat”	“be eaten”	“manner of eating/diet”

The manner nominal derivation processes illustrated in (66) are very regular in Chichewa. Although the dictionary does not lemmatise these nouns, some of them are almost lexicalised, e.g. *kadyedwe* (“diet”) in (66b). These have not been considered systematically

Stellenbosch University https://scholar.sun.ac.za

in the previous literature, so they do not have a class of their own in the Bleek-Meinhof system.

Both the DIM and the manner *ka*-nouns trigger the AM *ka-.* Interestingly, there is no conflict with the *ka*-initial nouns of the *m(u)-a*-AC (≈NC1). However, as will be argued in Chapter 6, the DIM *ka-* and the manner *ka*-prefixes are morpho-semantically different and are simply accidentally homophonous.

As shown in (65), the plural DIM triggers a different AM, *ti-*. The *ti*-AC forms the third morphologically predictable AC. I briefly introduce this AC next.

4.11 The *ti*-AC, which includes NC13

The *ti*-AC is one of the less complex classes as it mainly includes plural DIM nouns, such as those in (67).

(67) a. ti-mbale
PL.DIM-plate
"small plates"

b. ti-a-galu
PL.DIM-PL-dog
"small dogs"

However, in a very rare occurrence, there are two nouns in the dataset that optionally trigger AM *ti*-(cf. Watkins, 1937: 40). These are illustrated in 68).

(68) a. tulo ("sleep") b. tuvi ("excreta, faeces")

The nouns in (68) optionally trigger AM *li-* (≈NC5). These nouns appear to belong to the *li*-AC on a semantic basis and to the *ka*-AC on word-initial phonological basis.

In the next section, I introduce the *ku*-AC, which is the fourth morphologically predictable AC.

4.12 The *ku*-AC, including NC15 and NC17

This AC comprises three types of expressions, namely nominal infinitives (NC15), *ku*-initial locatives (LOC) (NC17) and CPs.[36] Nominal infinitives are derived from verbal expressions by attaching the infinitival prefix *ku-*, as illustrated in (69).

[36] The term *nominal infinitive* rather than the traditional *class 15 nouns* is adopted from Du Plessis (1982a,b) and Visser (1989).

Stellenbosch University https://scholar.sun.ac.za

(69)	**Verb stem**	**Nominal infinitives**
a.	imb-a	ku-imba
	ROOT-VFV	INF-sing
	“sing”	“singing”
b.	lamul-a	ku-lamul-ir-a
	ROOT-VFV	INF-command-APP-VFV
	“command”	“commanding /leading”
c.	kond-a	ku-dzi-konda
	ROOT-VFV	INF-RFM-love
	“love”	“selfishness/loving oneself”
d.	mv-a	ku-sa-mva
	ROOT-VFV	INF-NEG-hear
	“hear”	“not being able to hear/stubbornness”

In (69), the infinitive expressions are derived by prefixing the infinitival affix *ku-* to various projections of the verb expression. For example, in (69a) the prefix is attached to the simplex verb stem *imba*; in (69b) the prefix is attached to the complex verb stem *lamulira*, one carrying the applicative suffix *-ir-*. In example (69c), the prefix is attached to the complex verb stem, *-dzikonda,* one carrying the reflexive marker (RFM) *dzi-*, whereas in (69d) the infinitival affix *ku-* is attached to the negated verbal complex deriving a negative infinitival nominal expression. All *ku*-initial infinitival expressions trigger alliterative AM *ku-.* Nominal infinitives are not lemmatised in the dictionary as they are highly regular.

Although locative expressions are discussed in detail in Chapter 6, I just introduce them here because they are part of the larger AC system of Chichewa. Locative phrases (LOCPs) are obtained by attaching locative markers (LOCS) to either adverbial or adjectival stems or NPs. As shown in the glossing in (70a-b), in non-lexicalised LOCPs, LOCS are free morphemes with nominal complements (cf. Bresnan & Mchombo, 1995: 195–208; Carstens, 1997: 368; Salzmann, 2004: 15–17). This section only focuses on *ku*-initial locatives of the kind shown in (70).

(70)	a.	ku Dubai	b.	ku sukulu	c.	ku-seli
		LOC NAME OF PLACE		LOC school		LOC-ADV.STEM
		“to/in Dubai”		“to/at school”		“behind”

Although the locative expressions in (70) are not lemmatised in the dictionary, there are a few idiomatic locatives whose meanings are not transparent in the way those in (70) are. The dataset has only six of these, including the following (see Appendix 18).

Stellenbosch University https://scholar.sun.ac.za

(71) a. ku-mwezi
LOC-moon
“menstruation”

b. ku-maso
LOC-face
“private parts”

Expressions like those in (71) also trigger the *ku-* AM just like the *ku*-nominal infinitives.

As mentioned above, the third type of expression that has been associated with the *ku*-AC are CPs, such as that in (72) (cf. Bresnan, 1995: 34ff).

(72) **Ku**-ku-onek-a [CP**kuti** **ku**-gwa mvula].
SM-T/A-seem-FV that SM.T/A-fall rain
“It appears [that it is going to rain].”

Although the *ku*-prefix is phonologically identical to the first syllable of the complementiser *kuti* (and to the SM encoding the (quasi-referential) subject of the “weather” predicate in the embedded clause), the concordial agreement seems to be a mere coincidence as it is also possible for the AM *ku-* to appear without the corresponding trigger. Consider the example in (73) by way of illustration:

(73) **Ku/zi**-ku-onek-a ngati nyengo i-sintha.
AM-T/A-appear-VFV like weather AM-change
“It appears like the weather is going to change.”

In (73), the AMs *ku-* and *zi-* do not have any overtly realised agreement trigger, meaning that we are dealing with an unaccusative predicate lacking an external argument; *zi-* and *ku-* in (73) thus serve the same function as expletive *it* or *there* in the corresponding English translation. In these constructions, the AM *zi-* and *ku-* appear to function as a dummy AM, further confirming my view (introduced in Section 4.9) that these AMs also function as dummy elements.

Significantly, the patterns observed with the *ku*-AC are also found in the *pa*-AC, which I introduce next.

4.13 The *pa*-AC, including NC16

The *pa*-AC is the fifth morphologically predictable AC. Consider example (74).

(74) pa chola
LOC handbag
“on the handbag”

b. pa mwala
LOC rock
“on the rock”

c. pa-tali
LOC-ADJ.STEM (long)
“far”

Stellenbosch University https://scholar.sun.ac.za

As with *ku-*, the *pa*-AC also contains some idiomatic locative-derived expressions whose meanings are not predictable, like those in (74). The dataset has 13 such expressions, for example (75) and Appendix 19 (cf. Watkins, 1937: 41).

(75) a. pa-thupi
LOC-body
"pregnancy"

b. pa-ny-o
LOC-ROOT (defecate)-NFV
"anus"

Both the regular locative and the lexicalised types trigger the AM *pa-.*

The AM *pa-* is also observed to occur in expressions where the controller is not obvious, such as the following.

(76) a. **Pa**-li vuto.
AM-COP problem
"There is a problem."

b. **Pa**-li anthu ena a-madya ma-koswe.
AM-COP people some AM-T/A-eat PL-rat
"There are some people who eat rats."
Literally: "There are some people, they eat rats."

Just like the expletive AMs in (63c-d) and (73) above, the AM *pa-,* in (76) has no specific agreement trigger and may not have a specific location in time and space. In this regard, the examples illustrated in (76) can be regarded as another expletive use of the AM *pa-*.

In the next section, I introduce the sixth morphologically predictable AC in Chichewa, that involving the AM *mu-*.

4.14 The *m(u)*-AC (≈NC18)

The *mu*-AC in Chichewa is associated with the *m(u)*-initial locative expressions, which form the NC18 on the Bleek-Meinhof schema. These include the following.

(77) a. mu msika
LOC market
"in the market"

b. m nyumba
LOC house
"in the house"

Like the other locative ACs, there are very few idiomatic *m(u)*-initial locative nouns. My dataset has only two of these, as shown in (77) (cf. Watkins, 1937: 42).

(77) a. m-khwapa
LOC-armpit
"armpit" Lit.: "in the armpit"

b. m-kati
LOC-ADV.STEM
"inside"

Stellenbosch University https://scholar.sun.ac.za

Although this LOC prefix is homophonous with the prefixes for the *m(u)-a*-AC and *the u*-AC, there is no ambiguity in interpreting the locative *m(u)*- and the non-locative *m(u)*-. The locative *m(u)*- conspicuously denotes location or direction while the non-locative *m(u)*- has nominal reference. I return to this matter in more detail in Chapter 6, Section 6.4. In the next section, I provide an interim summary of the preceding discussion focusing on the various ACs in Chichewa.

4.15 Interim summary

To summarise the discussion focusing on the various ACs in Chichewa, three points stand out. First, the observed differences between what was referred to as *noun-internal distribution properties* vs. *noun-external distribution properties* (Gagliardi & Lidz, 2014: 59) are borne out in Chichewa. By structuring the agreement system based on noun-external information, I have arrived at a different classification system from the Bleek-Meinhof system which is largely based on noun-internal information. As a consequence, Bleek-Meinhof-based analyses arrive at a very narrow noun class system that omits many expressions that trigger the same set of AMs. In the present study, I have taken *noun-external distribution information* as my point of departure, focussing on agreement rather than *noun-internal distribution information*. The ACs that emerge from my consideration thus include expressions with diverse characteristics.

Second, although the noun-internal distribution information of the various contents within the various ACs is not as homogenous as is traditionally assumed, I have shown that the expressions share certain common semantic and/or phonological properties. More importantly, nouns denoting entities higher on the animacy hierarchy are observed to belong to the *m(u)-a*-AC while the rest of the classes take nouns that denote entities that are lower on the hierarchy. I will consider the nature and significance of these shared properties (i.e. noun-internal distribution properties) in more detail in Chapters 5 and 7.

Third, while some classes are composed of typical nominal expressions, other classes are composed of a range of agreement trigger-types. In this regard, in the next section, I consider the differences between the ACs composed of the nominal expressions and those that comprise more diverse complex phrases.

Stellenbosch University https://scholar.sun.ac.za

4.16 Noun classes vs. agreement classes

As shown in the foregoing discussion, there are three broad types of expressions that trigger various AMs. The first type includes morphologically less predictable nouns as illustrated in *m(u)-a-*, *li-, a-, u-, chi-* ACs. The second type includes more complex expressions with regular and predictable morphosyntactic processes, such as DIMs, locatives and CPs. The third group includes cases where the AM functions as a dummy form, i.e. the AM is used in constructions that lack overt controllers or contain controllers with conflicting agreement features. This distinction regarding classes made up of typical nouns and others including a mixture of complex expressions was first noticed quantitatively in the dataset as summarised in Table 4.1, where the first six ACs include large number of entries, with the lowest being 357 tokens while the other six ACs have scanty lemmatised entries with the highest being 16 tokens.

In some previous studies, a similar distinction has also been made where the first group has been identified as including nouns with *prefixes in primary function* as opposed to the second group which is assumed to comprise nouns with *prefixes in secondary function* (see i.a. Fortune, 1955: 54, 1970: 87ff; cf. also Harding, 1966: 24; Givón, 1972: 11; Maho, 1999: 98). However, note that to identify the second group as forming noun classes is to exclude equally important expressions such as CPs and PPs. As I will show in Chapter 6, in addition to including non-nominal expressions, the secondary classes have a different morphosyntactic structure to those containing only typical nominal expressions. Following the distinction observed in the dataset and the distinction made in some previous literature (see i.a. Fortune, 1955: 54, 1970: 87ff; Harding, 1966: 24; Maho, 1999: 98), I propose that these ACs should be analysed as forming two interrelated subsystems. Specifically, I propose that the ACs involving typical nouns should be identified as forming *agreement-based noun classes* (ANCs) while those including non-nominal expressions should be referred to as *general agreement classes* (GACs). According to the evidence considered in the present chapter, the latter is more general as it includes CPs, unspecified agreement controllers, locatives and DIMs while ANCs are a limited set which only obviously suggest the inclusion of nominal expressions. The classification proposed here is summarised in Table 4.5. I will, however, continue to identify the *agreement-based noun classes* by their AMs so as to make the nomenclature more meaningful. Take note that, although the *zi-*AC has very low lemmatised entries, it has been included under the noun class system because

Stellenbosch University https://scholar.sun.ac.za

it is used as a default AM where the nominal agreement features are underspecified, which implies that it is used as the default AM for the ANC system as well as for the GAC system (see Chapters 5 and 6 for more details).

Table 4.5: ANCs and GACs in Chichewa

	ANCs **(traditional 'primary classes')**		***GACs*** **(traditional 'secondary classes')**
1.	m(u)-a-*agreement-based noun class (*m(u)-a-*ANC)*	1.	*ka*-agreement class (*ka*-GAC)
2.	u-*agreement-based noun class (*u-*ANC)*	2.	*ti*-agreement class (*ti*-GAC)
3.	li-*agreement-based noun class (*li-*ANC)*	3.	*ku*-agreement class (*ku*-GAC)
4.	a-*agreement-based noun class (*a-*ANC)*	4.	*pa*-agreement class (*pa*-GAC)
5.	chi-*agreement-based noun class (*chi-*ANC)*	5.	*mu*-agreement class (*mu*-GAC)
6.	i-*agreement-based noun class (*i-*ANC)*		
7.	zi-*agreement-based noun class (*zi-*ANC)*		

Note that there is a discrepancy between the agreement based classes proposed by Corbett & Mtenje (1987: 8) and those proposed in Table 4.5. One of the sources of discrepancy lies in the fact that Corbett & Mtenje identified the AM *i-* as deriving two distinct homophonous ACs. Further, they also included the DIM *ka-* and *na-* as forming part of the core noun class system. What is also missing in their work is the subsystem that I have identified as a GAC. As far as Chichewa is concerned, then, the account that their proposed system analysis gives cannot be complete. Some previous studies also do not consider expressions comprising the GAC (the so-called *secondary classes*) as core to the typical noun class system and they are usually excluded or treated separately (see i.a. Watkins, 1937: 40–44; Corbett & Mtenje, 1987: 8–9; Carstens, 1991: 42–45, 1997: 370; Maho, 1999: 53; Matiki, 2001: 75–82). In the present study, I will attempt to account for the properties of both subsystems. The properties of these ANCs and GACs will thus be considered separately in Chapters 5 and 6, respectively.

In the next section, I provide a summary of two other important observations made in the present Chapter, namely those relating to noun formation processes and noun classification.

4.17 Nominalisation and noun class features in Chichewa

In the foregoing discussion, I have shown that nominalisation processes within each AC in Chichewa are much more diverse than previously thought. The facts seem to speak against the long-held traditional hypothesis that each noun class is largely determined by a specific

Stellenbosch University https://scholar.sun.ac.za

prefix in Chichewa. The facts, instead, suggest that, for the most part, classification is triggered by the semantic and word-initial characteristics of the expressions triggering agreement. The prefix seems merely to be participating in the lexical derivation and inflection systems of the language.[37] However, in some ACs, the prefix coincidentally has word-initial phonological characteristics that qualify a particular noun to belong to the relevant noun class. For example, the *u*-AC takes nouns that are syllables *m*-, *mu*-, and *u*- and *o*-initial, which also takes inanimate nouns that are derived by prefixes *m(u)*- and *u*-. Therefore, there is a need to draw a clear distinction between derivation/inflection and noun classification processes in Chichewa. For example, while the traditional literature identifies NC1 as being made up of *m*-initial nouns, there are many nouns in NC1 with other word-initial characteristics, reflecting a variety of nominal derivation strategies. This diversity of nominalisation strategies is observed to be widespread in all the other ACs too; in fact, there is no single class that has only one way of deriving the nouns it contains. Table 4.6 summarises the various morphological characteristics of each AC discussed in the present chapter.

Table 4.6: A summary of morphological properties of expressions in each ANC /GAC

Class	Derivation strategy	Morpheme	Example	Gloss
m(u)-a-*ANC (≈NC1)*	conversion	ø/tone	gogoda	"high-heeled shoe"
	prefixes	*m(u)-*	*m-londa*	"guard"
	prefixes and suffix	*m(u)-* and *-i*	*m-sodz-i*	"fisher(man/woman)"
	prefix	*ka-*	*ka-fadala*	"snouted beetle"
	"	*na-*	*na-nkalizi*	"centipede"
	"	*sa-*	*sa-khuta*	"glutton"
	"	*ma-*	*ma-taya*	"affluent person"
	prefixes	*w(o/a)-/o-/a-*	*wo-lumala*	"person with disability"
	prefix	*cha-*	*cha-mbampha*	"a type of termite"
	"	*bwa-*	*bwa-ntasa*	"a type of frog"
	"	*tsa-*	*tsa-munda*	"colonialist, estate owner"
	"	*-i*	*gonth-i*	"deaf person"
	compounding	ø	*msungi-chuma*	"treasurer"
	non-derived	ø	*galu*	"dog"

[37] Watkins (1937: 22–23) provides support to this view as he argues, "The frequent lack of noun prefixes in Chichewa must be well remembered when speaking of noun systems […] we cannot regard the prefix, nor its absence, as the sole or primary determinant of a noun class. It is rather the whole system of concordances, which […] may or may not include such affixes that must be taken as the fundamental basis of classification. This systematization is conventional and largely, though perhaps not entirely, arbitrary, as are other types of classification."

Stellenbosch University https://scholar.sun.ac.za

Class	Derivation strategy	Morpheme	Example	Gloss
u-*ANC* *(≈NC3 and NC14)*	prefixes	*m(u)-*	*m-tchetcha*	“elephantiasis”
	prefixes and suffix	*m(u)* and *-o*	*m-sonkh-o*	“tax”
	non-derived	*ø*	*bwato*	“canoe”
	prefix	*u-*	*u-bwezi*	“friendship”
	prefix and suffix	*u-* and *-o*	*u-batiz-o*	“baptism”
	non-derived	*ø*	*ufa*	“flour”
	compounding	*ø*	*mpenda-mphepo*	“wind vane”
li-*ANC* *(≈NC5)*	conversion	*ø*	*lemba*	“letter/script”
	suffix	*-o*	*lot-o*	“dream”
	prefixes	*li-*	*li-kulu*	“headquarters”
	prefix and suffix	*li-* and *-o*	*li-kodz-o*	“bilharzia”
	prefix and suffix	*d-* and *-o*	*d-ombol-o*	“rescue/ransom”
	compounding	*ø*	*pheru-pheru*	“butterfly”
	non-derived	*ø*	*phazi*	“foot”
a-*ANC* *(≈NC2 and NC6)*	prefix	*a-*	*a-nyamata*	“boys”
	prefix	*ma-*	*ma-uta*	“arrows”
	non-derived	*ø*	*ankolo*	“uncle (Eng.LW)”
	honorific marking	*a-*	*a-Phiri*	“[HON] Mr/Mrs Phiri”
	prefix and suffix	*ma-* and *-o*	*ma-dy-o*	“appetite, delicious food”
	compounding	*ø*	*madyera-mphoto*	“corruption, bribery”
	non-derived	*ø*	*malo*	“place”
chi-*ANC* *(≈NC7)*	prefix	*chi-*	*chi-badwa*	“thing”
	prefix and suffix	*chi-* and *-o*	*chi-bay-o*	“pneumonia”
	prefix	*ch(o/a)-*	*cho-vala*	“garment”
	compounding	*ø*	*chidya-onga*	“white stork”
	non-derived	*ø*	*chonde*	“fertility”
i-*ANC* *(≈NC4 and NC9)*	suffix	-o	*nol-o*	“whetstone”
	conversion	*ø*	*nyónga*	“energy”
	prefix	*N-*	*nth-irira*	“irrigation”
	prefix and suffix	*N-* and *-o*	*ny-imb-o*	“song”
	prefix	*i-*	*i-m-fa*	“death”
	prefix	*mi-*	*mi-tengo*	“trees/prices”
	non-derived	*ø*	*njala*	“hunger”
	compounding	*ø*	*mphamba-mutu*	“headband”
	non-derived	*ø*	*mimba*	“Stomach/pregnancy”
zi-*ANC* *(≈NC8 and NC10)*	prefix	*zi-*	*zi-phe*	“poisons”
	prefix and suffix	*zi-* and *-o*	*zi-pats-o*	“fruits”
	pluralisation	*ø*	*nyama*	“animals”
	compounding	*ø*	*zilubwe-lubwe*	“hysteria”
	non-derived	*ø*	*zambi*	“sin(s)”

Stellenbosch University https://scholar.sun.ac.za

Class	Derivation strategy	Morpheme	Example	Gloss
ka-*GAC* *(≈NC12)*	pluralisation	*ka-*	*ka-nyumba*	"a small house"
	nominalisation	*ka-*	*ka-yankhulidwe*	"manner of speaking/dialect"
ti-*GAC* *(≈NC13)*	prefix	*ti-*	*ti-nyumba*	"small houses"
	non-derived	*ø*	*tulo*	"sleep"
ku-*GAC* *(≈NC15 and NC17)*	infinitive prefix	*ku-*	*ku-gula*	"buying"
	locative marker	*ku*	*ku munda*	"at/to the garden"
	lexicalised	*ku-*	*ku-madzi*	"Nyau cult camp site"
pa-*GAC (≈NC16)*	locative marker	*pa*	*pa denga*	"on the roof"
	lexicalised	*pa-*	*pa-thupi*	"pregnancy"
mu-*GAC (≈NC18)*	locative marker	*mu*	*mu chipinda*	"In the room"
	lexicalised	*mu-*	*m-kati*	"inside"

Table 4.6's summary of ACs and the morphological characteristics of nouns that feature in these classes differs significantly from the ideal one-to-one relationship between prefix and noun class assumed in the Bleek-Meinhof system. Therefore, I reject this traditional view; instead I follow the views expressed by Lieber (2004: 9) regarding the lexicon and lexicogenesis:

> First, I believe that noninflectional word formation – derivation, compounding, and conversion – serves to create lexemes and to extend the simplex lexicon; for that reason, I believe that the meanings it expresses ought to reflect the semantic distinctions that are salient in the simplex lexicon. That is, to the extent that we find semantic classes that are significant in distinguishing the behavior of underived lexemes, we might expect derivation, compounding, and conversion to extend those classes.

In this regard, I consider the simplex lexicon in Chichewa as being augmented through a number of derivational strategies. The derivational processes are also not always straightforward as they can sometimes be non-productive, irregular and unpredictable (Comrie & Thompson, 2007: 381). Setting aside highly productive expressions such as plural nouns, locatives, AUGs and DIMs, most of the examples discussed in this chapter revolve around nominalisations where the nouns either name an activity or state, or name an argument) (e.g. agentive, instrumental, manner nouns, objectives of various kinds; (cf. Comrie & Thompson, 2007). In all these cases the derived nouns are also seen to retain properties of the verbal, adjectival, ideophonic, and nominal expressions they are derived from, a phenomenon observed in many other languages (see i.a. Lieber, 2004; Comrie & Thompson, 2007: 335–345). However, in Bantu, unlike in other languages such as English, the derived nouns must eventually belong to a particular AC. Therefore, the noun

Stellenbosch University https://scholar.sun.ac.za

classification system is a separate grammatical process from word formation processes. I propose a detailed ANC system in Chapter 5. Before that, in the following section, I review the main findings of the present Chapter.

4.18 Conclusion

The purpose of the present chapter was to discuss the properties of expressions forming the Chichewa ACs that were first introduced in Chapter 3. These twelve ACs were presented in section 4.2, whereafter the composition of each was discussed in turn (sections 4.3 to 4.14). There are many patterns that are observed to be contrary to the traditional assumptions about nominal structure, noun classification and agreement phenomenon in Chichewa.

With regard to morpho-syntactic aspects, I have shown that there is a variety of nouns in the Chichewa nominal lexicon. For example, there are non-derived nouns (i.e. a simplex lexicon), compounds, conversions, prefix-derived nouns and suffix-derived nouns. With regard to ACs, I further observed that some classes predominantly contain morphologically less predictable nouns (as shown in sections 4.3 to 4.8) while other classes include more complex, but morphologically predictable expressions that include NPs, CPs, PPs, etc. Agreement, then, is not only triggered by nominal expressions, but also by complex phrases and underspecified controllers. Following the observed differences in agreement trigger-types, I proposed that some ACs should be appropriately identified as *agreement-based noun classes* (ANCs) while the other ACs are *general agreement classes* (GACs). What appears to be the case is that agreement properties are fundamental in establishing the details of the nominal classification system, but certain very general "trigger" properties are also relevant (essentially, whether the trigger is part of the "content nominal" class or part of the "functional controller" class).

Another significant observation different from the traditional views relates to how the so-called NCPs and singular-plural number noun class pairing is analysed. First, since different ACs encompass expressions of various morphological types, I showed that the AMs for each AC also play the role of derivational/inflectional prefix for a subset of the expressions comprising the respective ACs. However, this subset is not a defining property of the entire respective ACs, as was the case in the Bleek-Meinhof system. Second, the much-observed singular-plural pairing cannot be generalised to the entire AC system in Chichewa. As I have

Stellenbosch University https://scholar.sun.ac.za

shown, some ACs have both singular and plural nouns, while other classes have their plural nouns in more than one AC elsewhere. Therefore, neither noun prefixes nor number marking come to the fore when one tries to establish a consistent basis for the Chichewa noun and agreement class systems.

With regard to semantic patterns, I have shown that the AC system is also structured on the basis of semantic features like [±human], [±animate] and [±agentive]. These features were argued to characterise the crosslinguistically observed animacy hierarchy, which has the tendency of placing human-, animate- and agent-denoting nouns higher on the hierarchy and the less animate and inanimate lower on the hierarchy. I showed that the *m(u)-a*-ANC contains nouns that are higher on the animacy hierarchy while the rest of the classes are predominantly characterised by nouns that are lower on the animacy hierarchy. A major class distinction was therefore proposed which identifies one group as the animate class and the other as the inanimate class.

While the traditional analyses view nominalization and noun class facts as intricately interlinked, I have argued that the nominal classification system is not tied to a particular derivational/inflectional strategy as each class contains both derived and underived expressions. With these facts in place, it should now be clear that the traditional Bleek-Meinhof noun class system only accounts for a subset of the nouns in each AC, owing to its preoccupation with nouns that bear prefixes. The present chapter has therefore set out and motivated an alternative descriptive typology, which is prerequisite for more insightful theoretical explanation of the facts in forthcoming chapters. In the next chapter, I propose and motivate in more detail the ANC system that emerges from the dataset considered in the present chapter.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 5

Noun-internal criteria behind the agreement-based noun class system

5.1 Introduction

Having introduced the various problems with the existing analyses of the Bantu noun class system in Chapter 2 and having considered the Chichewa empirical evidence in Chapters 3 and 4, the third goal of the present study is to propose a set of noun-internal distributional properties underlying the Chichewa ANC system. In the present Chapter, I broadly argue that, contrary to the traditional views discussed in Chapter 2, the typical noun class system in Chichewa in fact employs two interdependent subsystems – one which is semantically determined and another one that is phonologically determined. It is these two criteria that can help us to understand why some nouns have regular semantic properties, while others only have regular word-initial characteristics.

To establish the validity of this argument, I consider the challenge posed by Chichewa's noun class system from a child language learning perspective; specifically, my objective is to postulate an analysis based only on features that are plausibly acquirable. This implies appealing only to plausible input and cognitive resources available in language acquisition (see again Chapter 1, Section 1.3). The proposed analysis is structured as follows. In Section 5.2, I discuss the assumptions underlying the corpus empirical evidence vis-à-vis the type of evidence found in a real language-learning setting. In Section 5.3, I outline the kind of classification system that arises from the semantic criterion, specifically organised along the *animacy hierarchy* introduced in Chapter 4, Section 4.3.10. Here, I show that two super agreement-based noun classes (super-ANC) emerge: one *animate* and the other *inanimate*. In Section 5.4, in turn, I show that the inanimate super-ANC further classifies the nouns into subclasses based on their word-initial phonological characteristics. This gives rise to five phonologically determined noun classes. In Section 5.5, I evaluate the productivity status of the proposed semantic and phonological rules using Yang's (2016) rule productivity metric, the Tolerance Principle. In Section 5.6, I introduce the seventh noun class that emerges in cases where the agreement features are conflicting or underspecified. In Section 5.7, I discuss the properties of the emerging mixed ANC system. The merits of this proposed analysis are then highlighted in section 5.8. Finally, I summarise the Chapter in Section 5.9.

Stellenbosch University https://scholar.sun.ac.za

5.2 Linguistic evidence and the ecological setting of language acquisition

Faced with the empirical facts outlined in Chapter 4, the question that arises is how a child learning Chichewa goes about constructing the grammar underlying the proposed ACs. Before I attempt to answer this question, a brief clarification is in order as to how the corpus linguistic evidence compares to the evidence that children plausibly have access to in the process of language acquisition.

As introduced in Chapter 1, it should be apparent that no real-life language acquisition setting can provide the kind of comprehensive and well-structured evidence-set presented in Chapter 4. It is actually generally accepted that the linguistic evidence that is available to children is very partial – this is the classic *poverty of the stimulus* argument (Chomsky, 1975, 1980; Legate & Yang, 2002; Berwick, Pietroski, Yankama & Chomsky, 2011)[1]. In this regard, the large Chichewa nominal corpus presented in Chapter 4 should thus certainly not be equated with the set of evidence that children work with. Yang (2016: 109–224) argues that grammars are more readily learnable with a relatively small sample of language data and that if children were presented with all the linguistic evidence at once they could conceivably find it harder to converge on a stable grammar (cf. Newport, 1990; Elman, 1993; Perfors, Tenenbaum & Regier, 2006). Yang (2016: 224) further claims that "language learning not only needs to overcome impoverished data but is also enabled by impoverished data."

Therefore, I start off with the assumption that the primary linguistic data available to children for the acquisition of the Chichewa AC system is impoverished relative to the universe of Chichewa data that could potentially be mustered to demonstrate its properties. At this point, a Chichewa child language corpus comprising the most frequent early nouns would have helped to draw more fine-grained conclusions about the phenomenon. However, I could not find such corpus for Chichewa. I am therefore only, in Section 5.8 and in Chapter 7, able to indicate what one might expect to see or investigate in a corpus of this type; in-depth consideration of what Chichewa-acquiring children encounter during their early exposure to their native language must therefore be left to future investigation. In the interim, I will assume that the data presented in this study features regular patterns which are also found in the more partial data children are exposed to and that these patterns are the basis on

[1] For contrary views about the *poverty of the stimulus* hypothesis see Pullum & Scholz (2002).

Stellenbosch University https://scholar.sun.ac.za

which children acquire the principles underlying the Chichewa AC system. The advantage of taking this kind of perspective on the problem is that we can focus on simple traces of regularity rather than the entirety of what is encoded in adult grammars. The basis for this approach is the hypothesis that the ultimately very complex Chichewa noun and agreement class systems are built up "Goldilocks"-style – with acquirers systematically paying attention to "input that is neither too simple nor too complex" (Kidd, Piantadosi & Aslin, 2014; Biberauer, 2018b: 136). In "Goldilocks" terms, we would expect salient semantic and phonological patterns to matter in the early structuring of the Chichewa noun class and agreement systems.

Taking into account the empirical evidence presented in Chapter 4, it would seem that there are two possible triggers that could indicate to the child that Chichewa nouns belong to distinct categories. Specifically, following Gagliardi & Lidz (2014: 59 ff), I assume that children are guided by both the so-called *noun-internal distribution information* (i.e. semantic and phonological properties shared by the nouns in a particular class) and *noun-external distribution information* (i.e. AMs in their syntactic context). On the one hand, *noun-internal distribution information* is reflected in the recurrence of certain word-initial phonemes on nominal expressions such as nasals and vowels, which would suggest a phonological/sound-based categorisation. *Noun-external distribution information*, on the other hand, is systematically reflected on some words, such as adjectives, numerals, verbs, etc., which necessarily reflect the so-called AMs which vary with noun-type. This latter feature is made more salient by the fact that most AMs within the ANC are sonorant; hence phonologically distinct. In addition to the sonority aspect, some AMs are also alliterative with the noun's syllable-initial phonemes, which also make them prosodically salient. The prosodically salient word-initial phonemes (both on the noun and modifiers/verbs) and the alliterative concord could be key cues that contribute to the perceptual salience of the phonologically based categorisation and agreement patterns, which are the key properties of the Chichewa ACs proposed in Chapters 3 and 4 (see also Chapter 7). Given that words are arbitrary form-meaning pairings, one might also expect salient aspects of the meaning side of these pairings to play a role in grammatical development. And this does seem to be the case for Chichewa where features like humanness [±human], animacy [±animate] and agentivity/causation [±agentive] clearly play a prominent role in structuring the noun class system. In principle, the phonological and semantic cues may also interact in different ways

Stellenbosch University https://scholar.sun.ac.za

(see Chapter 7, Section 7.3). For ease of exposition, what I will do here is first consider the semantic sub-system before turning to the phonological one.

5.3 The semantically determined super-ANC system in Chichewa

As already indicated in the introduction, what is under consideration here are the noun-internal features that qualify particular nouns to belong to their respective ANCs. However, I will continue to identify these classes using the nomenclature proposed in Chapter 4, where each ANC is identified by the form of the AM that it controls. For ease of references, these ANCs are repeated in summary form in Table 5.1.

Table 5.1: A summary of agreement-based noun classes in Chichewa

	ANCs	**Associated traditional noun class**
1.	*m(u)-a*-ANC	≈NC1
2.	*li*-ANC	≈NC5
3.	*i*-ANC	≈NC4 and NC9
4.	*u*-ANC	≈NC3 and NC14
5.	*a*-ANC	≈NC2 and NC6
6.	*chi*-ANC	≈NC7
7.	*zi*-ANC	≈NC8 and NC10

As was shown in Chapter 4, each ANC comprises nouns associated with various morphological, semantic and phonological characteristics. Crucially, however, each of these classes is also observed to exhibit highly characteristic semantic and/or word-initial phonological properties.

Focusing on semantic properties first, a broad distinction emerges between nouns in the *m(u)-a*-ANC and the other ANCs.[2] In fact, there are at least five types of clues suggesting a core distinction between the *m(u)-a*-ANC and the other five classes. The first piece of evidence that Chichewa has a semantically determined super-ANC comes from the statistical distribution of animate/agentive and inanimate/non-agentive nouns across the various ANCs. Consider Table 5.2.

[2] It should be stated that the manner in which semantic criterion manifests here differs radically from the way it is applied in the traditional analysis. Therefore, I will show here that the traditional views about semantic basis of the various noun classes in Chichewa are not compatible with the findings that I present here.

Stellenbosch University https://scholar.sun.ac.za

Table 5.2: Distribution of animate/agentive and inanimate/non-agentive nouns in the ANCs

	Total nouns	**Animate/agentive**	**Non-agentive/Inanimate**
***m(u)-a*-ANC**	1222	88.2% (1079/1222)	11.7% (143/1222)
***li*-ANC**	843	5% (43/843)	95% (800/843)
***i*-ANC**	1453	19% (262/1416)	81% (1154/1416)
***u*-ANC**	839	7% (58/839)	93% (783/839)
***a*-ANC**	∞	∞	∞
***chi*-ANC**	1075	11% (122/1075)	89% (953/1075)
***zi*-ANC**	Default AM		
Total	**5828**		

The distribution of animate/agentive and inanimate/non-agentive in Table 5.2 is instructive. There is a high preference for animate/agentive nouns to belong to the *m(u)-a*-ANC compared to the other classes, and the reverse pattern holds in relation to inanimates/non-agentive. Also recall that the nouns identified as inanimate/non-agentive in the *m(u)-a*-ANC occur in clusters (e.g. plants, instruments, etc.) which may, if we take into account the properties of the *animacy hierarchy*, be ascribed agentivity on cultural grounds (see i.a. van Nice & Dietrich, 2003: 106; de Swart *et al.*, 2008: 135). In Chichewa, the way in which the animacy hierarchy regulates the "extension" of animate properties to inanimates is also observed in circumstances where animate pronouns are employed to refer to select inanimates with which a speaker feels a particular affinity. For example, a fisherman may refer to his canoe as *uyu ndi mzanga* ("this one [+HUMAN] is my friend").

The second piece of evidence comes from the distribution of the agentive and non-agentive suffixes *-i* and *-o,* respectively. Looking at all deverbal nouns that include the NFVs *-i* and *-o*, the picture that emerges is that summarised in Table 5.3.

Table 5.3: Distribution of agentive suffix -i and non-agentive suffix -o

	Agentive	**Non-agentive**					**Total**
	m(u)-a-NC	*u*-N	*li*-NC	*chi*-NC	*i*-NC	*a*-NC	Total	
NFV *-i*	**38**	1	1	8	4	0	**14**	52
NFV *-o*	**4**	98	72	74	47	15	**306**	310
Total	42	99	73	82	51	15	320	**362**

As shown in Table 5.3, the NFVs are distributed in a systematic way. Of all deverbal nouns ending with NFV *-i*, 73% (38/52) belong to the *m(u)-a*-ANC while the remaining 27% (14/52)

Stellenbosch University https://scholar.sun.ac.za

is shared among the other five ANCs. Of all the nouns derived via the NFV *-o*, only 1% (4/310) belong to the *m(u)-a*-ANC whereas 99% (306/310) belong to the rest of the non-agentive classes. What appears to be key here is that the distribution in Table 5.3 makes [+agentive] a rule-worthy property for the *m(u)-a*-ANC, whereas the [-agentive] rule is a rule-worthy property for the rest of the classes as a collective. Based on this consideration, then, it appears that we have two *super-ANCs*[3] in Chichewa, one agentive and the other non-agentive.

Importantly, Table 5.3 does not include the subset of *u*-initial nouns belonging to the *u*-ANC (≈NC14). These deverbal nouns are excluded because this subset of the *u*-ANC has a predictable morphological process that derives nouns from animate nouns to produce descriptive nouns such as the following.

(1)	**Animate noun**	**NC14 derived noun**
a.	galu (“dog”)	u-galu (“foolish person” Lit.: “dog like behaviour”)
b.	polofesa (“professor”)	u-polofesa (“professorship”) Eng.LW
c.	m-sodzi (“fisherman”)	u-sodzi (“fishing, fishing industry”)

Although example (1c) seems like a deverbal noun, note that the *u-* prefix is a predictable derivational affix which derives nouns from other nouns as shown in (1a-b). In this regard, I have excluded examples of this type in preparing the statistics of deverbal nouns presented in Table 5.2: all deverbal nouns in the agentive class can also, in principle, be turned into the type represented in (1c), but this is clearly a regular **further** derivation process, producing derived nouns that should not be counted alongside the “basic/primary” deverbal nouns that I am concerned with here. Including nouns like those illustrated in (1) in the basic deverbal nouns would result in the loss of two real generalisations, namely that relating to the agentive/non-agentive pattern just outlined, on the one hand, and that relating to the regular *u*-derivation process illustrated in (1c) on the other.

The third piece of evidence that supports the proposed super-ANC view comes from nouns that share the relevant word-initial phonological features but belong to different classes. For example, as already stated in the discussion of the relevant ACs in Chapter 4, nouns that share certain word-initial characteristics belong to the same class, except in cases where

[3] I propose the term *super-ANC* to refer to overarching noun classes which are further subcategorised on the basis of further criteria, see (4) and (10) below for diagrammatic representations.

Stellenbosch University https://scholar.sun.ac.za

the noun has agentive, human or a generally animate denotation, in which case it is very likely to belong to the *m(u)-a*-ANC. For example, in Chapter 4, Sections 4.3.2 and 4.6.1, it was shown that some nouns in the *m(u)-a*-ANC and *u*-ANC share the singular number prefix *m(u)-*. However, those that have an agentive denotation can be shown to belong to the animate super-ANC (*m(u)-a*-ANC), while those that have non-agentive meanings belong to other nasal-initial ANCs, on a phonological basis, see Chapter 4, examples (40) and (41). This has also been observed for other word-initial noun characteristics. Consider, for example, the following nouns with similar word-initial characteristics (comprising homophones and near-homophones), which belong to the animate super-ANC or elsewhere within the inanimate super-ANC on the basis of semantic contrasts.

(2)	**Animate super-ANC**	**Inanimate super-ANC**
a.	**mtantha** (“unfaithful person”)	**mtantha** (“elevation”) (*u*-ANC) (≈NC3)
b.	**mi**nisita (“minister”)	**mi**nisitile(“ministry”) (*i*- ANC) (≈NC4/9)-Eng.LW
c.	**mi**liyoneya (“millionaire”)	**mi**liyoni (“million”) (*i*-ANC) (≈NC4/9)-Eng.LW
d.	**sa**pota (“supporter”)	**sa**poti (“support”) (*i*-ANC) (≈NC4/9)-Eng.LW
e.	**simba** (“palm civet”)	**simba** (“initiation camp”) (*i*-ANC) (≈NC4/9)
f.	**chi**tute (“pouched rat”)	**chi**tupa (“identity card”) (*chi*-ANC) (≈NC7)
g.	**cha**nga (“night ape”)	**cha**ngu (“speed, briskness”) (*chi*-ANC) (≈NC7)

Examples (2b-d) are English loanwords with similar word-initial characteristics but belong to different classes, first based on their meanings and then based on their word-initial phonological characteristics. Specifically, when they have human denotation, they trigger AMs *m(u)-a-* and when they have non-human denotation, they trigger other AMs based on their word-initial phonological features.

Similar distribution patterns are also observed with ANCs that do not have regular word-initial characteristics. According to the dataset, there are two ANCs that contain a large number of nouns with varied word-initial characteristics, namely the *m(u)-a*-ANC (≈NC1) and *li*-ANC (≈NC5), which suggests that most of the nouns in these classes are not based on word-initial phonological features. As shown in Chapter 2, Section 2.2, I (2003: 21–22) also observed that the contrasts illustrated in (2) are found between these two classes. The examples in (30) in Chapter 2 are repeated with additional contrasting pairs in (3).

Stellenbosch University https://scholar.sun.ac.za

(3) Nouns that alternate between *m(u)-a*-ANC and *li*-ANC based on their meanings.

	Noun	***m(u)-a*-ANC (≈NC1)**	***li*-ANC (≈NC5)**
a.	bwande	"a kind of frog"	"a kind of hairstyle for post-partum women"
b.	búlu	"donkey"	"a lumpy bit of uncooked flour in mielie-pap"
c.	goli	"goalkeeper (Eng.LW)"	"yoke, handcuffs"
d.	bowo	"roan"	"hole"
e.	tambala	"rooster/cock"	"coin"
f.	gompho	"an underground insect"	"very short pair of trousers"
g.	fulu	"tortoise"	"a big tall handleless gourd for storing water"
h.	godi	"porcupine"	"holes in a board game called bawo"
i.	gaka	"pangolin"	"scaly skin found on crocodile skin"
j.	mamba	mámba ("cobra")	mamba ("fish/snake scales")

Note that most of these nouns do not even have a tone distinction as in (3j), which implies that they are morpho-phonologically the same and that they belong to different classes purely on the basis of the semantic differences between them, which cause them to be grouped into different ANCs. The other significance of the homophones and near-homophones in (2) and (3) is that a learner will need to notice the semantic distinctions to use the correct AMs.

The fourth piece of evidence in support of the semantically-based super-ANC comes from proper names and personified inanimates, as in fables and tales. All proper names and personified nouns from the inanimate super-ANC trigger AMs associated with the animate super-ANC. This phenomenon has been observed in many Bantu languages (see i.a. Corbett, 1991: 98; Maho, 1999: 64; Matiki, 2001: 66, 78).

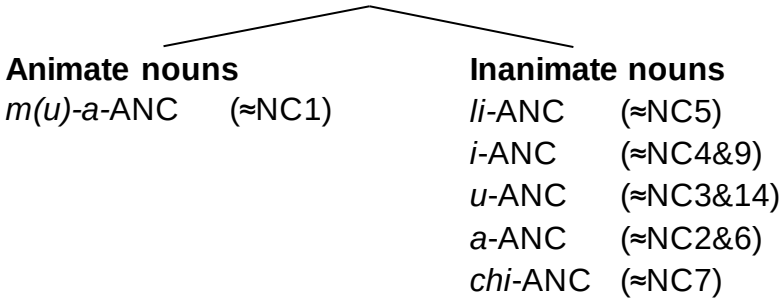
The fifth piece of evidence for postulating the super-ANC in Chichewa comes from the connotative meanings carried by the AMs. For example, the AMs (*u-; li-; i-* and *chi-*) found within the inanimate super-ANC are strongly associated with inanimate (and thus non-human) meanings, to the point where speakers may use these AMs as a strategy to insult others – e.g. referring to somebody as *i-* ***chi*** ("this, NON-HUMAN"), *i-**yi*** ("this, NON–HUMAN"), *i-**li*** ("this, NON –HUMAN"), etc.

The five pieces of evidence in support of the super-ANCs discussed here point, I argue, to the fundamental nature of Chichewa noun class system. Primarily, nouns may belong to two super-ANCs based on the semantic features [human]/[animate]/[agentive] (**animate**) and

Stellenbosch University https://scholar.sun.ac.za

[non-human]/ [inanimate]/[non-agentive] (**inanimate**). This is schematised by the diagram in (4).

(4) The semantically-determined super-ANC system in Chichewa



The Bleek-Meinhof classification system does not recognise the broader classification represented in (4), thus obscuring what seems to be a fundamental organisational principle in the Chichewa nominal domain. As I will demonstrate below, the inanimate super-ANC can be further articulated on the basis of word-initial phonological features. As already discussed (Sections 1.2.2 and 2.5.2), the observation that the Bantu noun class system references semantic and phonological considerations has been recurrent in the analyses of many Bantu languages (see i.a. Burton & Kirk, 1976; Batibo, 1987; Orr, 1987; Spitulnik, 1989; Kgukutli, 1994; Selvik, 1996, 2001; Contini-Morava, 1997; Demuth, 2000; Palmer & Woodman, 2000; Braver & Bennett, 2015). For example, it has generally been said that agentive, animate and human-denoting nouns belong to NC1 and that the various non-agentive, inanimate features are responsible for classifying nouns into the other noun classes. Some of the cited inanimate features include plants, trees, liquids, abstract, etc.; however, none of these features appear to apply in any consistent manner (see again Chapter 2, Section 2.4.3). Therefore, what has been missing to date is an account of how the various inanimate/non-agentive entity-denoting nouns come to distribute among the various distinct classes within the broader inanimate category that I am in the process of showing to be phonologically consistent. This is the objective I turn to next.

5.4 The phonologically determined inanimate noun classes

As indicated in the preceding section, there are five classes that make up the broader inanimate super-ANC (see again (4)). Within this broader category, the nouns appear to be further grouped based on their word-initial phonological characteristics. That some nouns are classified on the basis of their word-initial characteristics in Chichewa is not a novel observation (see i.a. Mchombo, 1978: 104; Matiki, 2001: 76). As shown in (4), there are four

Stellenbosch University https://scholar.sun.ac.za

ANCs in this subcategory that have systematic word-initial characteristics. Each of these will be considered in separate sections below.

5.4.1 The *i*-ANC

As introduced in Chapter 4, Section 4.7, nouns belonging to this class exhibit a range of word-initial characteristics. These are summarised in Table 5.4.

Table 5.4: Word-initial characteristics of the i-ANC.

Word-initial property	**Example**	**Distribution**
Homorganic nasal *N-*	*nyama* (“meat”)	**79.4%** (1154/1453)
Syllable *mi-, me*	*mimba* (“abdomen”)	**2.1∞%** (30∞/1453)
Vowel *i-,*	*imfa* (“death”)	**1.3∞%** (19∞/1453)
Consonants *s-, t-, f-, p-, b-, etc.*	*sitima* (“train, ship”) *tikiti* (“ticket”), etc. see Table 4.4	**17.2%** (250/1453)

Although these word-initial phonemes appear to be unrelated, they actually share profound phonological features. For example, the most frequent phonemes such as /N/, /s/, /t/, /i/, etc. share the phonological natural class feature [(+cor)onal] (cf. Chapter 7, Section 7.2.2). There are, however, two word-initial types given in Table 5.4 that do not, at first sight, seem to meet the [+cor] specification. The first is the front vowel /i/, which according to McCarthy (1988: 85) is described as [-cor]. However, this question of [+cor] being associated with vowels has also been considered differently in the literature: starting with the work of Clements (1991) and later works by Hume (1992, 1996) and Flemming (2003), it has been argued that the front vowel /i/ involves coronal features (cf. Johnson & Reimers, 2010: 237).[4] This initial element therefore also fits the generalisation I am making here. In Chapter 7, Section 7.2.2, I will further show that the front vowel /i/ clusters with [+cor] phonemes in other crosslinguistic phonological phenomena. Therefore, even if one were to argue that the front vowel /*i*/ is [-cor] there is crosslinguistic evidence that this vowel has the tendency to cluster with [+cor] phonemes in other phonological phenomena (cf. Gierut, Cho & Dinnsen, 1993: 223; Chen & Kent, 2005: 513). The presence of syllable-initial /*mi*/ and /*me*/ in Table 5.4 also require more detailed discussion.

[4] According to Johnson & Reimers (2010:237), coronal sounds are made with the blade or tip of the tongue and include the following phonemes /t/, /d/, /s/, /z/, /θ/, /d/, /ʃ/, /ʒ/, /n/, /l/, /ɹ/, /j/ and front vowels.

Stellenbosch University https://scholar.sun.ac.za

The interesting pattern regarding *m*-initial nouns is that it is not only the syllable **onset** (/m/) that determines the AC of *m*-initial nouns, but also the nucleus segment, the vowel. In Chichewa, there are three types of nasal *m*-initial nouns that are further distinguished by the syllable nucleus segment. To see this more clearly, consider the following examples.

(5) **Nasal-initial noun classes differentiated by either the syllable onset or the nucleus segments.**

	Noun	**AM-numeral**	**Gloss**	**Noun class**	**Distinguishing segment**
a.	**mph**a.sa	**i**-modzi	"one reed-mat"	***i*-ANC** (≈NC9)	syllable onset
b.	**mph**i.ra	**i**-modzi	"one piece of rubber"	***i*-ANC** (≈NC9)	syllable onset
c	**m.**pa.sa	**u**-modzi	"kind of fish"	***u*-ANC** (≈NC3)	syllabic nasal
d	**m.**pi.ra	**u**-modzi	"one ball"	***u*-ANC** (≈NC3)	syllabic nasal
e.	m**i.**nga	**i**-modzi	"one thorn"	***i*-ANC** (≈NC4)	syllable nucleus
f.	m**e.**sa	i-modzi	"one knife" (Afr. LW)	***i*-ANC** (≈NC9)	syllable nucleus
g.	m**u.**nga	**u**-modzi	"one thorn"	***u*-ANC** (≈NC3)	syllable nucleus
h.	m**o.**no	**u**-modzi	"one fish trap"	***u*-ANC** (≈NC3)	syllable nucleus
i.	m**a.**lo	**a**-modzi	"one place"	***a*-ANC** (≈NC6)	syllable nucleus

In (5), the distinguishing element is shown in bold face. First, consider the sets of examples (5a,b) and (5c,d): the homorganic nasal *mph*-initial and the syllabic nasal *m*-initial forms lead to two different ANCs. This is the consistent criterion determining the distinction between the *i*-ANC (≈NC9) and the *u*-ANC (≈NC3) as regards nouns of the type illustrated in (5a-d). The second set of examples (5e-i) are all nasal *m*-initial. What differentiates them into three different NCs are their initial syllable **nucleus** segments. In this regard, the classification of (5e-i) is determined by the syllable nucleus being one of *i-, e-, u-, o-* and *a-.* As shown in (5), the vowels /*i*/ and /*e*/ assign the nouns to the *i*-ANC, vowel /*u*/ assigns to the *u*-ANC and the vowel /*a*/ assigns the noun to the *a*-ANC. Therefore, while the nasal onset distinguishes these nouns from nouns that are non-nasal *m*-initial, the syllable nucleus segment distinguishes these nouns further into their respective classes. This gives us the following sets.

Stellenbosch University https://scholar.sun.ac.za

(6) a. m**a** ***a*-ANC**
b. m**i**, m**e** ***i*-ANC**
c. m**u**, m**o** ***u*-ANC**

As can be seen in (6b), the front vowel /*i*/ is the active determiner in the /*mi*/ initial nouns. This is further motivation for recognising the vowel /*i*/ as one of the systematic phonological criteria for the *i*-ANC.

Returning to the question regarding nouns that appear as if they are not [+cor] in Table 5.4, it is shown that the nasal *m*-initial nouns are actually assigned the *i*-ANC on the basis of the vowel /i/ which I have already shown is [+cor] and that it clusters with [+cor] consonants. The clustering of coronal consonants and front vowels is further motivated in Chapter 7, Section 7.2.2. What is not clear at this moment is why there are a few nouns that are not necessarily [+cor]-initial, for example, those with initial /p/ and /b/. Although these phonemes are not prevalent in this class, their presence is also phonologically motivated. In Chapter 7, I discuss the various phonological feature hierarchies that predict the clustering of certain phonemes. For example, what is common in all these phonemes is the place feature or in terms of the *feature geometry*, they all fall under the *place node* (see Chapter 7, Section 7.2.2).

The symmetrical distribution of nasal-vowel pairing illustrated in (6) has also been observed to be a widespread phenomenon in Bantu languages (see i.a. Bell, 1972). However, Bell (1972) proposes a hypothesised diachronic trend for this pattern, arguing that the syllables are undergoing syncopation – that is, the nasal-initial syllables are losing their vowel segment turning them into syllabic nasals. Strikingly, Bell (1972: 37) claims that in Nyanja (a variety of which is under investigation in the present study) NC1 and NC3 have been heavily syncopated. However, according to the pattern shown in (6), these claimed changes seem to have not affected the symmetrical distribution of nasal-vowel pairing in Chichewa.

In summary, the *i*-ANC has systematic patterns and is not at all a ‘garbage' class where anything goes, as reported in some studies (see i.a. Orr, 1987; Matiki, 2001: 79). The *i*-ANC also differs in interesting ways from the *u*-ANC, which I turn to next.

5.4.2 The *u*-ANC

The *u*-ANC also exhibits all the properties of a phonologically-determined classification. As introduced in Chapter 4, Section 4.6, nouns that make up this class feature initial nasal

Stellenbosch University https://scholar.sun.ac.za

syllables */m/, /mu/, /mo/,* vowels */u/, /o/* and, in a few cases, bilabial */b/, /bw/*, */w/*. Consider their distribution summarised in Table 5.5.

Table 5.5: Distribution of word-initial phonemes for nouns in the u-ANC

Word-initial property	**Example**	**Distribution**
Bilabial nasal *m(u/o)-*	*mzele* (“line, queue”), *moto* (“fire”)	**72%** (605/839)
Labial vowels *u-*, *o-*	*ufa* (“flour”), *onga* (“gun powder”)	**25%** (212/839)
Bilabials *b-*, w-,	*bongo* (“brain”)	**2%** (15/839)
Velar *g-*	*gamba* (“big red-winged grasshopper”)	**0.4%** (4/839)
Dentals *f-, d-*	*fwifwi* (“kind of climbing herb”)	**0.3%** (3/839)

As shown in Table 5.5, the most systematic word-initial characteristics are the syllabic nasal *m-* and the back-vowel *u-*. Note that, like the *i*-ANC above, this class also features a nasal and a pair of back vowels as the basis for its defining properties. As already shown in (5-6), the back vowels *u-* and *o-* also determine the classification of nouns that are syllable *mu-* or *mo*-initial as they all belong to the *u*-ANC.

At this point, it should be clear that there is an important symmetry between the major sets of word-initial characteristics – (*N-*, *i-*) and (*m-, u-)* - for the *i*-ANC and *u*-ANC, respectively (see also Chapter 7, Section 7.2.2). This symmetry is a strong piece of evidence for the proposed classification mechanism and also for considering these two classes as **contrasting** noun class pairs. The cluster pair (*N-*, *i-*) and [*m-, u-*) contrast in respect of the phonological features [+cor], [+front], [-labial], and [-cor], [+back], [+labial], respectively. Labials and back vowels do not involve tongue body manoeuvres of the sort associated with coronal phonemes (Flemming, 2003: 5). Consequently, the occurrence of the labial (*b-, w-*) initial nouns in the *u*-ANC can no longer be described as accidental. Therefore, what we now have are two phonologically determined noun classes (*i*-ANC and *u*-ANC) which contrast based on their word-initial features ([±cor], [±back], [±labial]). As I will show in Chapter 7, Section 7.2.2, the observed consonant-vowel symmetry is also predicted by feature hierarchy theories, such as *feature geometry* (see i.a. Clements, 1985; McCarthy, 1988; Gierut *et al.*, 1993).

Significantly, a similar symmetrical distribution is also observed with nouns that are *ma-* and *a-* initial. I discuss these in the next section.

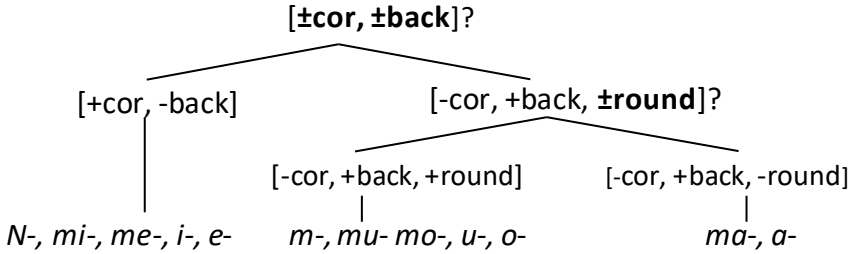
Stellenbosch University https://scholar.sun.ac.za

5.4.3 The *a*-ANC

As introduced in Chapter 4, Section 4.5, the *a*-ANC includes nouns that are *ma*- and *a*-initial. The noun-initial *ma*- and *a*- serve various functions. For example, the *a*- may be a plural or HON marker, while the *ma*- appears as a plural marker, a manner nominal prefix or an idiosyncratic word-initial syllable. This noun class does not have any other shared word-initial characteristics apart from the *ma*- and *a*.

As shown above (cf. (5-6)), the vowel *a*- on the *ma*-initial nouns is the active criterion that distinguishes these nouns from the other *m*-initial nouns, such as those beginning in *me-*, *mi-*, *mo-*, and *mu-*. The three predominantly nasal- and vowel-initial based noun classes can therefore be a schematised as in (7).

(7)



The symmetry observed between the nasals and the cardinal vowels, I argue, provides justification for combining the traditional noun class pairs (NC3 and NC14), (NC2 and NC6) and (NC4 and NC9) into three noun classes *u*-ANC, *a*-ANC and *i*-ANC, respectively. Further to the nasal- and vowel-based classification in (7), nouns beginning with other consonants are also classified on the same phonological basis, i.e. with reference to the features [±cor], [± labial]. This is shown in Table 5.6.

Table 5.6: Consonant-vowel symmetry in the phonologically determined noun classes

	***i*-ANC**	***u*-ANC**	***a*-ANC**
Nasal and Vowel	*/N/, /mi/, /me/, /i/, /e/*	*/m/, /mu/, /mo/, /u/, /o/*	*/ma/, /a/*
Consonants	*/s/, /t/, /z/, /f/*, etc.	*/b/, /w/, /f/*	NONE

The pattern summarised in Table 5.6 thus explains the systematic presence of certain consonant-initial nouns in what are predominantly nasal-vowel initial classes.

Next, I consider the *chi*-ANC.

Stellenbosch University https://scholar.sun.ac.za

5.4.4 The *chi*-ANC

As shown in Chapter 4, Section 4.8, the *chi*-ANC is the least controversial class as regards the basis on which it is postulated. All of the 1075 *ch*-initial nouns in the dataset trigger AM *chi-*. Some previous studies have claimed that the noun class assignment criteria underlying this class are morphological and phonological (Corbett & Mtenje, 1987: 5). I confirm these observations. However, I am inclined to argue that the assignment of nouns belonging to this class is **primarily** on **phonological** grounds, with the non-vocalic syllable-initial phonemes **/tʃ/** (**ch-)** being those of central relevance. The only *chi*-initial nouns in this group that appear to govern agreement in morphologically predictable ways are the complex *chi*-NPs that include evaluatives (see Chapter 6, Section 6.3.2).

Take note that the *chi*-initial phoneme is also [+cor], which would, at first sight, also qualify it to fall into the *i*-ANC, as shown in (8). However, it is distinguished from the *i*-ANC on the basis of [±ant(erior)] features. The articulation of *ch-* does not involve the tip of the tongue which makes it [-ant] phoneme. Unlike the nasal-based noun classes above, the *chi*-ANC shows that it is strictly the syllable onset that determines agreement in this case; it does not matter which vowel follows the *ch-*. However, the AM on which the nomenclature is modelled manifests by default as *chi-*. Consider the following examples.

(8)		**Noun**	**AM-numeral**	**Gloss**
	a.	chamba	chi-modzi	“one kind of music/dance”
	b.	cheke	chi-modzi	“one cheque” (Eng. LW)
	c.	chinthu	chi-modzi	“one thing”
	d.	chola	chi-modzi	“one handbag”
	e.	chulu	chi-modzi	“one anthill”

The only instance where the *ch*-initial noun would trigger a different AM, is when the noun has strong human/agentive connotations (e.g. *chemwali* -> “sister”) or where the inanimate-denoting *ch*-initial noun has been personified (e.g. *chijere* -> “bracelet” as a name of a person or a character in a folktale). In such cases the noun belongs to the animate, *m(u)-a-*ANC.

In the next section, I consider how the relevant word-initial phonological features interact in general.

Stellenbosch University https://scholar.sun.ac.za

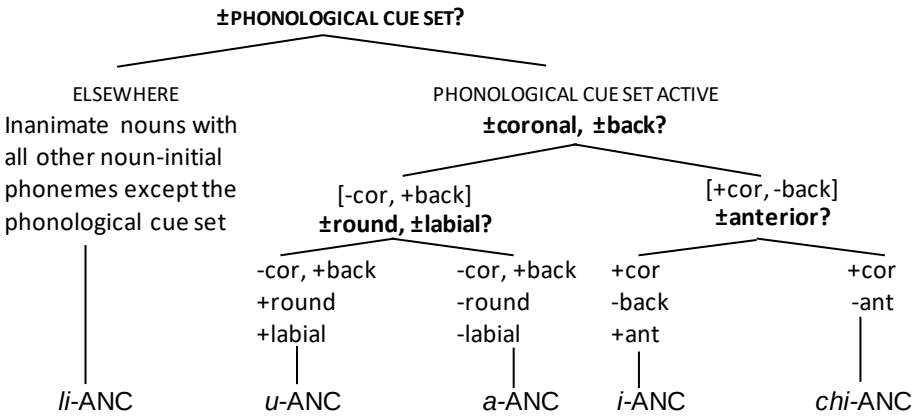
5.4.5 The phonological cue set in Chichewa

Having considered all the phonologically determined noun classes, it appears that there is a word-initial *phonological cue set* in Chichewa that functions as a discriminative stimulus to determine how nouns are classified. This phonological cue set includes the following phonological features.

(9) a. [+cor], [-back], [+ant]: *N-, i-, e-, mi-, me-*
b. [+cor], [-back], [-ant]: *ch-*
c. [-cor], [+back], [+labial], [+round]: *m-, u-, o-, mu-, mo-*
d. [-cor], [+back], [+labial], [-round]: *ma-, a-*

The phonological cue set in (9) subdivides the inanimate nouns in (4) into five distinctive classes, with the fifth one coming next. There are two proposed classes that I have not discussed to date, namely the *li*-ANC and the *zi*-ANC. I discus the *li*-ANC now and consider the *zi*-ANC later in Section 5.5. Regarding the classification criterion for the *li*-ANC, I propose that it arises as a phonologically default class within the inanimate super-ANC. I argue that this default class is a product of the *Elsewhere Condition*, a widely cited fundamental principle of linguistic organization, which offers a plausible account of the interaction of rules and exceptions in a variety of linguistic phenomena (see i.a. Anderson, 1969; Kiparsky, 1973; Halle, 1997; Stump, 2001; Brown & Hippisley, 2012; Yang, 2016, chap. 3). That is, where the noun-initial phoneme of an inanimate noun does not fall into the class of phonemes defined by the phonological cues proposed here, the noun in question will be assigned to the *li*-ANC. The full proposal of the phonologically determined noun class system is schematised in (10):

(10)



As shown in (10), when the phonological cue set is not active, the noun belongs to the *li*-ANC, whereas when it is active, the noun is evaluated on the basis of a series of parameters.

Stellenbosch University https://scholar.sun.ac.za

Take note that the parameter structuring shown in (10) holds for grammatical structuring more generally (see i.a. Biberauer, 2017: 48, 2018a: 103). If the cue set is active, the first evaluation relates to place of articulation features, whether the noun-initial phoneme is [±cor] or [±back]. Two classes are obtained from this evaluation, one containing nouns with word-initial phonemes that are [-cor, +back] and the other [+cor, -back]. On the one hand, the [-cor, +back] set is further evaluated for round or labial features, the [+round], [+labial] set gives rise to the *u*-ANC and the [-round], [-labial] set obtains the *a*-ANC. The [+cor, -back], on the other hand, is further evaluated as to whether or not the word-initial cue is [±ant], which distinguishes between the *i*-ANC ([+ant]) and the *chi*-ANC ([-ant]). Take note that the AM for the elsewhere *li*-ANC is also [+cor] which would bring difficulties in distinguishing it from the *i*-ANC, specifically *l*-initial nouns, which make up 18% (153/843) of all nouns in this class. As it will be shown by the end of the present chapter, there are other factors that govern the classification of nouns that appear to satisfy criteria for more than one class (see Section 5.7).

My characterisation of the *li*-ANC in (10) is also contra Corbett & Mtenje's (1987:5) views, who claim that this noun class is formally characterized by plosive-initial nouns. I propose, however, that the presence of word-initial plosives in many, but not all of the nouns in the *li*-ANC is in fact a consequence of the fact that the other classes take only nouns that are nasal- and vowel-initial, with plosive-initial nouns thus belonging to either the *m(u)-a*-ANC or the *li*-ANC.

So far, I have now considered the basic classification criteria for nouns that I have analysed as belonging to ANCs that are phonologically determined. Although these classes are subsumed under the inanimate class, the various subclasses under this super-ANC are determined by word-initial characteristics, meaning that further semantic attributes of the nouns in these sub-classes are no longer relevant.

The observations made here regarding the word-initial phonological cue sets also agree with crosslinguistic observations that phonological cues are either edge-aligned or prominent segment-related (see i.a. Sekerina & Brooks, 2007: 21; Benavides-Varela & Mehler, 2015; Sande, 2016: 345, 2017: 73). For example, all the phonological cues are either syllable-initial or are placed in the syllable nucleus (see Chapter 7, Section 7.2.3 for further details). With regard to phonologically determined nominal classification, Kelly (1996: 258) also reports that studies have consistently shown that phonological cues play a role in nominal

Stellenbosch University https://scholar.sun.ac.za

gender assignment cross-linguistically, for example in German, French, Russian,[5] Hebrew, etc. (see i.a. Maratsos & Chalkley, 1980; Maratsos, 1983; Levy, 1988; Kelly, 1992, 1996; Cassidy & Kelly, 2001; Durieux & Gillis, 2001; Bird, Ralph, Seidenberg, McClelland & Patterson, 2003; Déchaine *et al.*, 2014).

One more aspect about the proposed classification that has to be considered is whether the proposed rules are productive enough to be considered as rules. To do this I employ the rule productivity metric proposed by Yang (2016). In the next section, I provide a computation of the distribution of the noun classification rules proposed in sections 5.3 and 5.4.

5.5 Productivity of the ANC rules

As introduced in Chapter 1, Section 1.3.1, rule productivity can be calculated by considering the following: (i) number of attested cases (N) and (ii) number of exceptions (*e*) to the rule. A productive rule is one which has 'a sufficiently small number of exceptions', where both productivity and exceptions are numerical values calculated by the following equation (Yang, 2016: 8–9).

(11) $e \leq \theta_N$, where $\theta_N := \frac{N}{lnN}$ and *lnN* stands for 'logarithm of *N*'.

According to (11), for a rule to be productive, *e* must be equal to or 'below a critical threshold,' θ_N.

To determine whether the semantic basis of the classification of nouns into the semantic super-ANC forms a productive rule, I employ the formula in (11). As introduced in Chapter 4, there are 1222 nouns forming the *m(u)-a*-ANC. Therefore, N = 1222 and the number of acceptable exceptions for the rule to be productive is calculated as follows.

(12) $e \leq \theta_N$
$\theta_N \leq \frac{1222}{ln1222}$
$e \leq 172$

[5] For example, Smoczyńska (1985) claims that gender and agreement are acquired earlier in Polish compared with Russian partly because Polish exhibits a phonologically salient gender marking system on diminutive expressions (cf. Demuth & Weschler, 2012).

Stellenbosch University https://scholar.sun.ac.za

According to (12), the animacy rule in Chichewa can only be productive if the number of *e* is not more than 172. For argument's sake, although the clear-cut inanimate nouns in the *m(u)-a*-ANC are assumed to form a subset of the Chichewa-specific animacy hierarchy, I calculate whether this number would be above the threshold value to offset the animacy rule. The set of these inanimate nouns in the *m(u)-a*-ANC was summarised in Table 4.3, repeated here as Table 5.7.

Table 5.7: Sets of inanimate nouns in the m(u)-a-ANC

Category	Statistical distribution
Letters of the alphabet	24
Instruments	33
Plants	46
Non-count miscellaneous nouns	29
Inherently ***a***-initial nouns	11
Total	**143**

According to Table 5.7, the total clear-cut non-agentive and inanimate nouns, which are the actual attested exceptions in the dataset is 143. As computed in (12), for a dataset of 1222 elements, a rule will be productive if the number of exceptions does not exceed 14% (172/1222). Strikingly, the attested exceptions in Table 5.7 is 12% (143/1222). Therefore, even if the sets of nouns in Table 5.7 are to be treated as genuine exceptions, the semantic rule of animacy is still productive in Chichewa.

Next, I consider the productivity of the phonologically determined noun class rules. Employing the Tolerance Principle formula (11), Table 5.8 summarises the productivity of the various phonological rules introduced in Section 5.4.

Table 5.8: The productivity of phonological rules in phonologically determined ANCs

ANC	Rule	Number of tokens	Θ_N	Attested exceptions	Productivity status
***i*-ANC**	[+cor], [+ant]	1453	13.7% (199.5)	6.7% (98)	productive
***u*-ANC**	[-cor],[+labial], [+round], [+back]	839	14.8% (124.6)	1.1% (10)	productive
***a*-ANC**	[-cor], [-labial], [+back], [-round]	357	17% (60)	0% (0/357)	productive
***chi*-ANC**	[+cor], [-ant]	1075	14.2% (154)	0% (0/1075)	productive

Stellenbosch University https://scholar.sun.ac.za

As shown in Table 5.8, all the phonological rules are productive, that is, the number of exceptions is below the computed critical threshold values (θ_N). With these results from the Tolerance Principle, I can confidently argue that the proposed rules are valid for the various ANCs. However, note that I could not calculate the rule productivity for the *Elsewhere*-based *li*-ANC, as there is no precise rule to be tested.

While the *li*-ANC is the phonologically default class, in Chichewa there is also what I can describe as a syntactically default agreement class. I introduce this next.

5.6 The default *zi*-ANC

As stated in Chapters 1 and 3, the noun classes arise due to the obligatory syntactic agreement that must be encoded on various dependent words such as modifiers and predicates. However, in some cases, the agreement triggers may have conflicting features or may be underspecified such that the agreement system fails to identify the corresponding AM. When such gaps arise, the AM *zi-* is used by default. It is worth noting also that the AM *zi-* is one of the most versatile affixes in Chichewa as it is used in at least six different environments, for example as (i) a plural marker for singular nouns in the *chi*-ANC (13a), (ii) a plural augmentative marker (13b), (iii) a default plural marker for nominal expressions with ambiguous number marking (13c)[6], (iv) an AM for nouns whose form is ambiguous between singular and plural, especially those in the *i*-ANC (14a), (v) an AM for agreement triggers with conflicting features (14b), and (vi) a default AM for underspecified triggers (14c) (see also Chapter 4, Section 4.9 and Chapter 6, Section 6.3).

(13)

	Singular	**Plural**	
a.	chi-lombo	**zi**-lombo	
	SG-N.STEM	PL-N.STEM	
	“beast”	“beasts”	
b.	chi-galu	**zi**-a-galu	
	SG.AUG-dog	SG.AUG-PL-dog	
	“big dog”	“big dogs”	
c.	bambo	a-bambo	a-**zi**-bambo
	father	HON/PL-father	HON-PL-father
	“father/man”	“([HON]) fathers/men”	“[HON] fathers/men”

[6] Although the expression *azibambo* in (13c) is regarded as ungrammatical by grammar purists, this strategy for disambiguating plural honorific from ordinary plural is common in both spoken and written Chichewa (see Chapter 6, Section 6.3.3).

Stellenbosch University https://scholar.sun.ac.za

(14) a. Mbalame **zi**-ku-uluka.
bird (SG/PL) AM-T/A-fly
"Birds are flying."

b. Galu ndi mbuzi **zi**-na-thawa.
dog (*m(u)-a*-ANC) and goat (*i*-ANC) AM-T/A-run
"The dog and the goat run away."

c. **Zi**-ku-on-ek-a kuti si-ti-dya.
AM-T/A-see-STAT-VFV that NEG-AM-eat
"It appears as if we will not eat."

With respect to the agreement and noun class system, there are two important properties of the affix *zi-*. First, it functions as a plural marker and the AM for the respective nouns. In this regard, it can be analysed as forming part of the ANC system, although there are only 16 lemmatised nouns that are *zi*-initial. Second, the affix *zi-* arises as a default AM which cannot be associated with any ANC but rather acts as part of the GAC that I have proposed in Chapter 4. In this regard, the affix *zi-* participates in both the ANC and GAC subsystems, such that both sub-systems would be incomplete without mention of the AM *zi-*. Therefore, the AM *zi-* is both in the ANC and GAC systems. As I will show in Chapter 6, this multifunctionality of the affix *zi-* is a case of underspecification, specifically that the prefix is underspecified such that it is 're-used' by various grammatical functions. Therefore, I do not consider the AM *zi-* as representing two homophonous affixes.

In the next section, I show how the broader Chichewa ANC system can be assumed to involve a mixed noun class system, where the semantic and phonological criteria and the default AM *zi*-identified in the present chapter work together to allocate various nouns to their respective classes.

5.7 The mixed noun class system

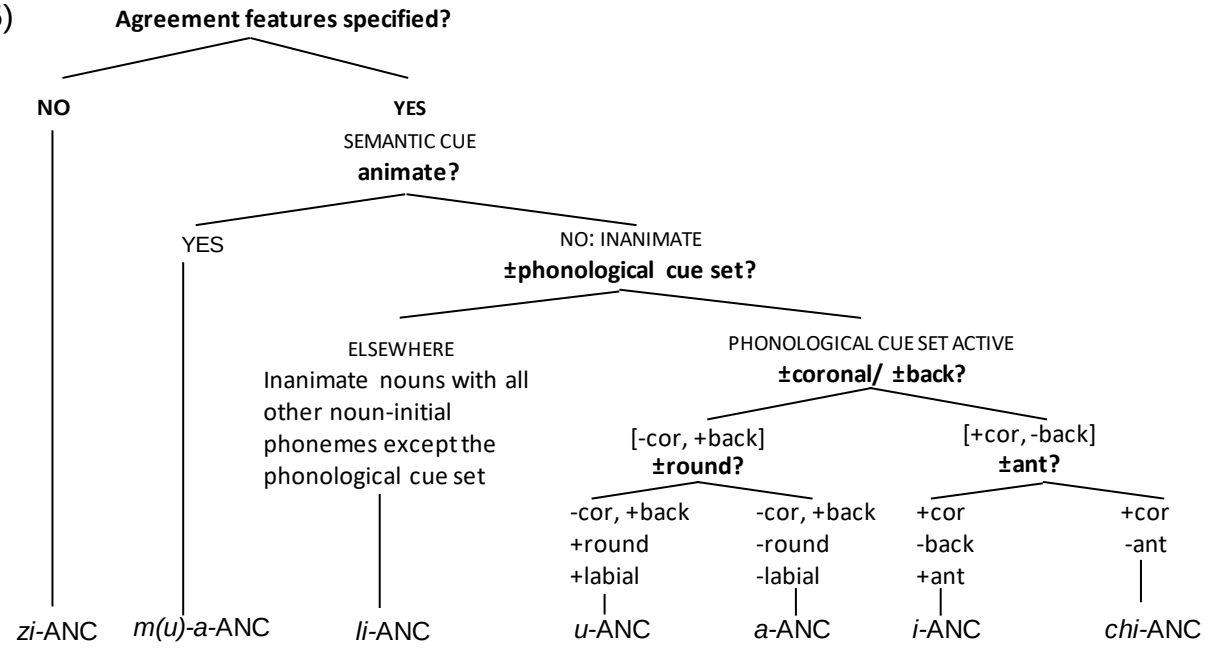
In the preceding discussion, I have shown that Chichewa employs a mixed system of nominal classification; i.e. one entailing both conceptual/semantic and formal/phonological subsystems. In the present section, I show how these subsystems could be taken to interact with each other.

Stellenbosch University https://scholar.sun.ac.za

5.7.1 Hierarchical organisation of the Chichewa ANC system

According to the evidence presented above, the seven ANCs in Chichewa are classified according to sets of semantic and phonological cues. In broad terms, the classification takes the following form:

(15)



In (15), the agreement system first checks if controller's agreement features are specified, if not the default AM *zi-* is triggered. However, nouns that are prefixed with the plural *zi-* marker, due to their alliterative congruence with the prefix *zi-* also trigger AM *zi-*. If the relevant features are specified in the controller, the nouns are firstly classified into two semantically determined super-ANCs, one accepting nouns that are higher on the animacy hierarchy (animate) and another accepting nouns that are lower on the animacy hierarchy (inanimate). The inanimate group is further subclassified based on word-initial characteristics. If a noun does not begin with any of the phonemes in the *phonological cue set*, the noun will belong to the *li-*ANC, while nouns that begin with any of the phonemes in the phonological cue set will belong to one of the subclasses following the path described in (10) above.

The hierarchical organisation shown in (15) is, however, only one possibility as other paths of classification could also be argued for. For example, the phonological criterion could be the primary step that identifies phonologically determined nouns as belonging to one super-ANC and the rest to a non-phonologically based noun class. This non-phonologically based

Stellenbosch University https://scholar.sun.ac.za

class could then further subclassify based on either semantic or phonological criteria. This alternative classification would not, however, explain why agentive and animate nouns in Chichewa are rare within the phonologically determined noun classes. This fact thus indicates that the phonological system is subsumed within a larger semantic division. A further challenge to the alternative is that an analysis along those lines does not fit into patterns that have been observed crosslinguistically. For example, in many languages that employ both semantic and phonological classification criteria, the phonological subclasses are observed to be subsumed under the inanimate or non-human classes (Corbett, 1991, chaps 2–3; Dahl & Fraurud, 1996: 55; Dahl, 2000: 101). This is, for example, the case in Yimas (Lower Sepik family of languages of Papua New Guinea) and Guébie (a Niger-Congo language spoken in Ivory Coast), respectively (see i.a. Foley, 1986: 85–91, 1991: 119; Corbett, 1991: 51–56; Sande, 2016: 340, 2017: 50).

I would like to argue that the patterns illustrated in the foregoing discussion form the basis of the Chichewa noun class system from which the more complex structure arises. In other words, a child learner perceives these distinctions in the process of acquiring the Chichewa noun class system, specific details of which are outlined in Chapter 7.

Next, I present a brief crosslinguistic picture of similar mixed-noun class systems in languages across various language families.

5.7.2 The crosslinguistic distribution of mixed noun class systems

Having proposed the noun class system summarised in (15), there are two relevant observations from crosslinguistic research on gender/noun class research. Firstly, many studies agree that all attested gender systems have a semantic core which may include a formal subsystem (see i.a. Corbett, 1991, chaps 2–3, 2013; Dahl & Fraurud, 1996: 55; Dahl, 2000: 101). Secondly, where the noun class system has a formal criterion, it is the animate or human nouns that are assigned by semantic rules while the inanimate or non-human nouns are assigned by formal rules (Corbett, 1991, 2013; Foley, 1991: 119; Dahl & Fraurud, 1996: 55; Dahl, 2000: 101; Sande, 2017, chap. 3). In this regard, Dahl (2000: 113) concludes that gender systems are organised around three major dimensions which tend to go together as summarised below.

(16)

animate	inanimate
referential	lexical
semantic	formal

Stellenbosch University https://scholar.sun.ac.za

The pairing in (16) implies that animate nouns are usually classified on a referential basis; hence the classification system is semantically based. Inanimate nouns, by contrast, tend to be classified on the basis of lexical form; hence the classification system is formal. This pattern is also borne out in the noun class system being proposed here. However, note that even though the formal system is seemingly not semantically based, in actual fact it is primarily classified on the basis of semantic categorisation. Only the NPs that do not have strong animacy features are further grouped on basis of their formal lexical characteristics (e.g. phonological and morphosyntactic).

The kind of semantic and phonological subsystems proposed in the present chapter are also observed in other languages. A sample of this is summarised in Table 5.7 below.

Table 5.9: Some languages with semantically- and phonologically-determined agreement systems

Language	Language family and geographical location	
Yimas	Lower Sepik family- Papua New Guinea	(Foley, 1986, 1991; Corbett, 1991: 55–56; Aronoff, 1994; Aikhenvald, 2000: 367)
Arapesh	Torricelli family - Papua New Guinea	
Bowili	Niger-Congo - Ghana	(Heine, 1982: 199ff; Aikhenvald, 2004: 1036; Lewis, Gary & Femig, 2015)
Qafar	East Cushitic language – Ethiopia, Djibouti	(Corbett, 1991: 51–52; Dimitriadis, 1997: 7)
Guébie, Godié, Krahn, etc.	Niger-Congo, Kru family - Ivory Coast	(Corbett, 1991: 51; Sande, 2016, 2017)
Hausa	Chadic, Afro-Asiatic – Nigeria, Niger	(Corbett, 1991: 52–53; Lewis *et al.*, 2015)
Bainuk	Niger-Congo, Atlantic – Guinea-Bissau	(Dimitriadis, 1997; Lewis *et al.*, 2015; Sande, 2017)
French	Indo-European – France, Belgium, etc.	(Karmiloff-Smith, 1981; Corbett, 1991: 57–62; Lyster, 2006; Lewis *et al.*, 2015; Culbertson, Gagliardi & Smith, 2017)
Tsez	North Caucasian – Dagestan Republic	(Gagliardi & Lidz, 2014; Lewis *et al.*, 2015; Culbertson *et al.*, 2017)

In many of the languages listed in Table 5.7, the noun class (gender) system is hierarchically organised on the basis of more than one consideration. For example, Foley (1991: 119) identified two broad types of noun classes in Yimas: those with a semantic basis and those where the assignment is strictly phonological. Just as I have proposed for Chichewa here, the phonological subclassification in Yimas is also observed to operate within the inanimate

Stellenbosch University https://scholar.sun.ac.za

subclass. In total, Yimas has 10 distinct noun classes of which four are semantically based, while the other six are phonologically determined. Table 5.7 represents this situation (Foley, 1986, 1991: 119–163; Corbett, 1991: 55–56).[7]

Table 5.10: The mixed noun class system of Yimas

Main systems	Noun class	Class-specific criteria
Semantic subsystem	**I**	male humans and humans whose sex is not highlighted
	II	female humans
	III	higher animals e.g. pigs, dogs
	IV	plants and trees
Phonological subsystem	**V** (50% of all nouns)	Residue class (nouns ending in /p, t, k, m, n, ŋ, r, l/)
	VI	nouns with the cluster /ŋk/ in word-final position simplified to /ŋ/
	VII	nouns marked by final /mp/ simplified to /m/ in word-final positions
	VIII	nouns ending in /i/ or /y/
	IX	nouns ending in what is referred to as the underlying /aw/
	X	nouns with final segments /k/ or /ŋk/

As shown in Table 5.8, the phonological cues in Yimas are word-final. This is also the case in Guébie, Qafar and Hausa (Corbett, 1991; Foley, 1991; Sande, 2017). However, in Bainuk, as in Chichewa, the phonological cues are word-initial (Dimitriadis, 1997: 3; Sande, 2017: 79–81). Nevertheless, the word-final NFVs in Chichewa also play a role in distinguishing the semantic classification of some deverbal nouns, specifically by contributing to the semantic interpretation of the nouns as either agentive or non-agentive (see Chapters 3 and 4). The word-initial and word-final cue patterns support Sande's (2017: 73) prediction that phonological cues are either edge-aligned or prominent segment-related. There is a rich body of research more generally that highlights the importance of edges - initial and final - in acquisition and in language structure (see Chapter 7, Section 7.2.3).

[7] Although Foley (1991: 119–163) identifies ten classes, Corbett (1991: 55–56) reports that there are in fact eleven classes in Yimas. This discrepancy is not significant here as the purpose of this example is simply to show how semantic and phonological noun classes are organised in Yimas.

Stellenbosch University https://scholar.sun.ac.za

Another crosslinguistic pattern, in addition to the broader semantic and phonological noun classes introduced in the foregoing sections, is what appear to be phonological subclasses within the semantic classes. I consider these in the next subsection.

5.7.2.1 Further subsystems

Another property of the Chichewa noun class system, also observed in Yimas, is the presence of what Foley (1986: 85) observed to be a combination of both phonological and semantic criteria in some classes. In Chichewa, the animate noun class system has a large number of *ka-* and *na-* initial nouns that seem to also form a phonological subclass. This conclusion seems plausible considering that not all *ka*-and *na-* initial nouns have an agentive or animate denotation. Even some loan words that are *ka-* or *na*-initial seem to follow this rule. Consider examples in (17).

(17) a kapinga (“type of grass”) Eng.LW
b. kabichi (“cabbage”) Eng.LW
c. nandolo (“pideon peas”)
d. naliti (“needle”) Afrikaans.LW

The consistent occurrence of some *ka-* and *na*-initial nouns in these sub-classes was also observed by Matiki’s (2001: 76) nonce word experiment, where it is reported that nonce words were also observed to follow this rule. The assignment of *ka*-initial nonce words to this predominantly semantically determined noun class implies that the nonce words and the non-agentive nouns are assigned here purely on a phonological basis.

In a similar manner, the *li*-ANC, one associated with non-agentive and inanimate nouns, contains animate/agentive nouns that seem to be assigned to this class on a phonological basis. Consider the following.

(18) a. ligondo (“slender mongoose”)
b. likonyani (“calf”)
c. lipedwa (“bagworm moth”)
d. lipumi (“type of flying termite”)

Although there are only 21% (9/42) *li*-initial animate nouns in my dataset, Matiki’s (2001: 77) nonce word experiment also showed that *l*-initial nonce nouns were assigned to the traditional NC5 (≈*li*-ANC). This implies that this could indeed be a phonologically-determined classification within the Elsewhere ANC.

The manner in which the *ka-, na-* and *l*-initial words appear to form a phonologically determined subclass can be understood in terms of what is known as the *cohort effect* (Sekerina & Brooks, 2007). According to Sekerina & Brooks (2007: 21), the term *cohort*

Stellenbosch University https://scholar.sun.ac.za

effect refers to the "competition among lexical candidates occurring when the initial phonemes of spoken words activate other words sharing the same phonemes" (cf. Swingley, Pinto & Fernald, 1999). The cohort effect here explains why the inanimate nouns that are *ka-, na*-initial and the animate *l*-initial nouns systematically follow a phonological rather than the expected semantic template.

As I have shown in the preceding discussion, phonological and semantic subsystems feed into each other. This relationship between the semantic and phonological subsystems is observed crosslinguistically. For example, Corbett (1991: 38–41) reports that in Russian, semantic rules take precedence. However, in the Ethiopian language, Qafar, phonological rules are observed to override semantic rules (Corbett, 1991). Similarly, according to Alcock & Ngorosho (2004: 28), in Kiswahili, phonological criteria are observed to be stronger than semantic ones (cf. Demuth, 2000). However, cases of semantic assignment overriding the phonological rules are also reported within the same language (cf. Corbett, 1991: 48; Aikhenvald, 2004: 1033). As shown in Chapter 4, each of the phonologically-determined noun class has a few animate nouns whose word-initial properties satisfy the relevant phonological rules. The presence of these animate nouns in the phonologically determined classes implies that the phonological rules can override the semantic rules (cf. Mchombo, 1978: 104; Matiki, 2001: 76). The opposite pattern is, however, also observed in Chichewa, with semantic rules overriding phonological rules: we find all proper names, and personified objects in the semantically determined classes, regardless of their word-initial characteristics. As discussed in Chapter 4, nouns referring to animate entities that are perceived (e.g. culturally) as less animate are also assigned to the inanimate classes irrespective of their animate meanings (cf. Fortune, 1970: 94–96). Semantic and phonological rules may therefore override each other in various ways. Given the fact that exceptions are known to be tolerated, with the tolerance threshold recently having become calculable (cf. again Yang, 2016), we can now understand why there are a minority of animate nouns in the generally inanimate classes and vice versa (cf. Section 5.5).

In the next section, I outline some further supporting evidence for the noun class system that has been proposed.

Stellenbosch University https://scholar.sun.ac.za

5.8 Some relevant observations from previous studies

The noun class system outlined here, does not just have intuitive appeal, but also receives support from previous studies on Bantu languages. Studies have consistently observed that Bantu languages reference both semantic and morpho-phonological properties in their noun class systems. As noted above, what was missing however, was a systematic approach to the nature and interaction of these elements. Mchombo (1978: 104) and Matiki (2001: 76–77) both draw strong conclusions regarding the role of phonology and semantics in the Chichewa noun class system. For example, Mchombo (1978: 104) says the following:

> It is naturally to be expected that semantic features would predominate, i.e. such semantically relevant features as [±human], [±animate], [±liquid], [±abstract], etc., may have the greatest relevance in determining the gender classification. Nonetheless, it seems to be the case that purely phonological aspects, e.g. the phonetic (sic - PKM) features of the initial sounds or syllables do have a role in the determination of gender. This is demonstrably so in Chichewa where one noun class is largely characterized by the occurrence of an underlying nasal prefix N which is nonsyllabic and is homorganic with the following consonants, rather than through any obvious semantic notions. … Yet another class is characterized in part by the non-animateness of the objects and in part by the phonetic feature $[+\ ^{-\text{continuant}}_{-\text{nasal}}]$ in the initial sound.

Similar observations regarding phonologically-determined classification are also made in other related Bantu languages (see i.a. Alcock & Ngorosho, 2004; van der Spuy, 2009: 201).

The second piece of support for the noun class system being proposed here comes from Bantu-wide language acquisition studies focusing on the noun class system. Demuth (2000) summarises findings from different Bantu languages such as Sesotho (Connelly, 1984; Demuth, 1988); Setswana (Tsonope, 1987); Siswati (Kunene, 1979); Zulu (Suzman, 1980) and Isangu (Idiata, 1998). According to Demuth (2000: 283), all these studies report that the morphological realisation of noun class prefixes proceeds via a three-step process. This is summarised in (19)(cf. Demuth, 1988: 73; Herbert, 1991: 110–111).

(19) Stages in the acquisition of Bantu noun class prefixes (Demuth, 2000: 283)

a. No prefixes (full or partial noun stems)
b. ‘Shadow’ vowel and nasal prefixes
c. Full morphophonologically appropriate noun class prefixes.

Although the major objective in these studies was to investigate the acquisition of the NCPs and not necessarily the noun class system per se, the findings are also insightful in the

Stellenbosch University https://scholar.sun.ac.za

context of the present study. For example, the first stage (19a) implies that children have not yet acquired number marking rules, so they use what I have referred to as dependent and independent noun stems in Chapters 3 and 4. At the second stage in (19b), the nasal- and vowel-initial "archi-" class markings (referred to as *prefixes* in (19b)) arise. The fact that the earliest class markings are precisely the nasal and vowel ones may well be highly significant: recall that the identity of the nucleus plays a particularly key role in the determination of the Chichewa agreement system (see Section 5.4). The fact that vowels and nasals, which are the canonical nuclear elements - arise earlier than other noun-initial markings could thus be interpreted as support for the analysis presented above: acquirers appear to have an early sensitivity to these elements, which, in terms of the analysis argued for in the present study, distinguish among various ANCs and GACs in Chichewa (see Chapter 7). While this may be naturally accommodated within my proposed system, it is not something that receives an analysis of any kind within the traditional Bleek-Meinhof schema.

As I stated in Section 5.2, I expect the relevant noun class features to be salient in the Chichewa child language corpus. First, I expect nouns signalling the human vs. non-human, animate vs. non-animate and agentive vs. non-agentive distinctions to be frequent. Second, I expect nouns with the word-initial phonological features ±coronal, ±labial, ±back, ±round to feature highly in the corpus. Although such a corpus is not available yet, in Chapter 7, I show that most of the mentioned features are already observed to be the most salient ones in child language corpora crosslinguistically. In this regard, the Chichewa child language corpus would be necessary to confirm these assumptions and crosslinguistic observations.

5.9 Summary and conclusion

In the present chapter, I set out to look in more detail at the noun-internal distribution properties that qualify the various nouns to belong to the ANCs proposed in Chapter 4. To achieve the stated objective, I adopted a child language acquisition perspective as my point of departure, so as to propose underlying criteria that is plausibly acquirable. Two types of cues were observed in the dataset, namely semantic and phonological. Based on these two cue-types, I showed how they give rise to a two-way super-ANC system, subdivided along the animacy hierarchy, which broadly distinguishes between nouns higher on the animacy hierarchy and those lower on the hierarchy. Unlike in the traditional account, I have argued that in Chichewa animacy effects derive a super-ANC which differentiates the *m(u)-a*-ANC from the rest. On the one hand, the *m(u)-a*-ANC, takes nouns that have animate reference,

Stellenbosch University https://scholar.sun.ac.za

specifically in the following hierarchy; human > animate > agentive > connotatively animate > culturally animate. The inanimate super-ANC, on the other hand, takes all inanimate nouns including animate nouns that are culturally deemed not worth to be ascribed human or animate features.

I then showed that the inanimate super-ANC further classifies nouns based on their word-initial phonological characteristics. I argued that the phonologically determined classification is based on a set of word-initial phonological features which together make up a phonological cue set. The relevant phonemes may be either derivational/inflectional prefixes or idiosyncratic syllable-initial units. The elements in the phonological cue set are distinguished along the following phonological features: [±coronal], [±round], [±labial], [±anterior] and [±back]. Most of these phonological features have not been considered before as underlying the Chichewa noun class system.

To determine whether the proposed criteria are rule-worthy, I employed Yang's (2016) Tolerance Principle metric, which showed that both the semantic and phonological rules are highly productive in Chichewa and that they cannot be offset by the observed exceptions. This confirms that the semantic and phonological regularities are indeed plausible underlying properties of the Chichewa ANC system. In addition to the rule productivity, I also argued that the symmetry observed between nasals and vowels is another strong justification for treating what were traditionally understood as separate noun classes as forming one ANC, namely *u*-ANC (≈NC3+NC14), *a*-ANC (≈NC2 + NC6) and *i*-ANC (≈NC4 + NC9).

In addition to the super-ANC system, I also proposed that the *zi*-ANC arises as the default AC for nominal expressions that have conflicting or underspecified agreement features. I then showed that the overall picture of the Chichewa noun class system is a mixed noun class system which exhibits a hierarchical organisation observed in other grammatical systems in general. There is also support from previous studies in Bantu languages regarding the phonological subsystem and language acquisition. Although, the previous observations were not consistent with the traditional approach in which the relevant studies were conducted, they confirm the findings made in the present study. In this regard, the analysis proposed here thus in many respects represents a regularisation and elucidation of the various observations made in the previous studies on Bantu noun classes.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 6

Properties of morphosyntactically predictable expressions in Chichewa

6.1 Introduction

In the present chapter, I discuss properties of the morphosyntactically predictable expressions introduced in Chapter 4 in more detail. Recall that I distinguish typical nominal expressions that create *agreement-based noun classes* (ANCs), which were the focus of Chapter 5, from morphosyntactically predictable expressions which will be my focus here (see again Chapter 4, Section 4.15). In terms of agreement, the former govern agreement in a morphologically less predictable manner while the latter govern agreement in a morphologically more predictable manner. However, as I will show here some predictable nominal expressions also belong to the ANCs discussed in Chapter 5. In the present Chapter, I analyse the morphologically predictable expressions as belonging to three broad types, namely evaluative expressions, locative phrases (LOCPs) and the clearly non-nominal complex phrases. The evaluative category includes four different subtypes, namely diminutive (DIM), augmentative (AUG), honorific (HON) and attitude (ATT) nominal expressions. Second, the LOCPs also belong to three different types, depending on the marker that derives them, namely *pa* ("on/at"), *ku* ("to/at") and *mu* ("in"). All of the LOCPs convey meanings to do with location or direction. Third, the complex phrases category includes agreement triggers such as CPs, non-locative PPs and instances where the AMs serve expletive-like function.

The rest of the chapter is structured as follows: In Section 6.2, I provide a broad background to the traditional distinction between *primary* and *secondary* noun prefixes. Then, in Section 6.3, I discuss the morpho-semantic properties of the evaluative expressions in Chichewa. The section is further subdivided into six parts, which focus, respectively, on DIMs, AUGs, HONs and ATTs and a discussion of the formal properties of these evaluative expressions, with the sixth subsection providing an interim summary. In Section 6.4, I deal with locative expressions as follows: In Subsection 6.4.1, I introduce the traditional view of LOCPs; in Section 6.4.2 I examine the properties of locative markers (LOCs), whereas in Section 6.4.3, I focus on the syntactic properties of LOCPs. Next, in Section 6.4.4, I consider the properties of LOCPs on the basis of their discourse and grammatical functions. This is followed by Section 6.5, where I focus on agreement triggers that involve complex phrases such as, CPs, non-locative PPs and structures where AMs play expletive functions. Then in Section

Stellenbosch University https://scholar.sun.ac.za

6.6, I consolidate the proposal made in Chapter 4 that the various non-nominal expressions that are seen to trigger agreement should be considered as forming *GACs* rather than *ANCs*. Finally, in section 6.7, I summarise and conclude the chapter.

6.2 The primary-secondary NCP distinction

As noted in Chapters 2 and 4, some works in the previous literature draw a distinction between what is variously referred to as a *basic/primary* versus *secondary* classification (see Maho, 1999: 88) or a distinction between *primary* and *secondary noun prefixes* (see Fortune, 1955: 54, 1970: 87ff; Harding, 1966: 24) or what Givón (1972: 11–12) designates *inherent and derived noun gender* versus *prepositional (locative) gender* (cf. Guthrie, 1948: 856). Other works, however, do not (overtly) acknowledge these distinctions (see i.a. Mufwene, 1980; Sproat, 1985; Myers, 1987; Bresnan & Mchombo, 1995; Ferrari, 2005). The primary-secondary noun class distinction is largely based on the observation that, on the one hand, some nouns in the Bleek-Meinhof system involve bound noun stems that obligatorily require certain sets of prefixes. In Chichewa, these primary classes are taken to comprise the traditional NCs 1/2, 3/4, 5/6, 7/8, 9/10 and 14. On the other hand, secondary noun classes are derived by attaching a different set of prefixes to the nouns identified as forming primary classes. In Chichewa, these include the traditional NCs 12/13, 16, 17 and 18. It is also worth noting that the terms *primary* and *secondary prefix/class* come from the traditional assumption that nouns in Bantu consistently consist of the stem and the prefix, as shown in (1) and (2). Examples (1a) and (2a) illustrate primary noun classes while (1b) and (2b) represent secondary noun classes. The superscript numbers, (^{1}) and (^{2}), in the glosses are meant to represent primary and secondary, respectively. Note that the Bleek-Meinhof modelled glossing convention shown here is for illustrative purposes only.[1]

(1) Chichewa (Harding, 1966: 42)

(a) mu-dzi
NCP3^1-village
“village”

(b) ku-mu-dzi
NCP17^2-NCP3^1-village
“at the village”

(2) Zezuru (Fortune, 1970: 89)

(a) mu-nhu
NCP1^1-person
“person”

(b) chi-mu-nhu
NCP7^2-NCP1^1-person
“a cheeky, ill-tempered person”

[1] For consistency’s sake, the glosses in some examples taken from other sources have been modified.

Stellenbosch University https://scholar.sun.ac.za

In (1a) and (2a), primary prefixes are shown to attach to bound nominal stems while in (1b) and (2b) secondary prefixes are shown to attach to full NPs. However, in Chichewa, there is a dependent nominal stem *-nthu* (“being”) that is observed to take secondary prefixes such as DIM and AUG prefixes as in *ka-nthu* (“small thing”), *chi-nthu* (“thing/big thing”). To my knowledge, this is the only dependent nominal stem that combines with both evaluative and locative secondary prefixes. The set of secondary classes proposed varies from one language to another and also from one study to another. In Chichewa and other related languages, some evidently secondary noun classes share the prefix and AMs of certain primary classes such that the distinction between the two has not been clearly stated. However, as will be shown in Section 6.3, **the morpho-semantic properties of *secondary noun prefixes* are quite regular as they attach to full NPs and introduce an additional regular meaning to the noun** (cf. Chapter 4). This is illustrated in (3).

(3) a. chi[1]-wala
SG-N.STEM
“grasshopper”

b. chi[2]-chi[1]-wala
AUG.SG-SG-N.STEM
“a big grasshopper”

The prefixes *chi*[1]- and *chi*[2]- in (3a) and (3b) are evidently morpho-semantically different, but are given the same status of NCP in the traditional literature (see i.a. Matiki, 2001: 81). Note that the secondary prefix illustrated in (3b) is different from what is variously referred to as the *pre-prefix*, *augment*, or *initial vowel* found in some Bantu languages, such as isiZulu, Luganda, Kirundi (see i.a. de Blois, 1970; Mould, 1974; Hyman & Katamba, 1993; Maho, 1999: 61–63; Ndayiragije, Nikiema & Bhatt, 2012). Here, I will argue that it is consistently justified to analyse the so-called secondary prefixes (such as that in 3b) as exponents of Chichewa’s functional projections above the number-encoding domain in the nominal structure (see Section 6.3.4).

The full set of these functional elements and their semantic properties in Chichewa is given in Table 6.1. Where the traditional primary and secondary NCPs are homophonous, the traditional class number is appended with a plus (+) sign to indicate that the prefix serves both primary and secondary functions. Following the nomenclature introduced in Chapter 4, I will continue to identify the various ACs either as ANCs or GACs.

Stellenbosch University https://scholar.sun.ac.za

Table 6.1: Complex expressions and their ACs in Chichewa

Functional element	NC	Semantic property	Prefix	Example
***ka*-GAC**	12	Singular DIMs and pejorative/ complimentary	*ka-*	***ka****-mwana* (“small/adorable child”)
***ti*-GAC**	13	Plural DIMs and pejorative/ complimentary	*ti-*	***ti****-a-na* (“small/adorable children”)
***chi*-ANC**	7+	Singular AUGs and pejorative	*chi-*	***chi****-mwana* (“big/bad/adorable child”)
***zi*-ANC (*vi-/dzi-*)**	8/10+	Plural AUGs and pejorative /complimentary/ CPs/ conjoined NPs/ expletive	*zi-* *vi-/ dzi-*	***zi-****a-na* (“big/bad/adorable children”) ***vi****-a-galu*/***v****-a-galu* (“big/ ugly/ dangerous dogs/ big good-looking dogs”)
***a*-ANC**	2+	Honorific	*a-*	***a****-bambo* (“HON father”)[2]
***u*-ANC**	14+	Attitude/judgment	*u-*	***u****-galu* (“stupidity”) Lit. “dog like”
***pa*-GAC**	16	Locatives, ‘near’ or ‘explicit’ / CPs/expletives	*pa-*	***pa*** *phiri* (“on the mountain”)
***ku*-GAC**	17	Locatives, ‘remote’ or ‘general’ / CPs/expletives/ infinitives	*ku-*	***ku*** *phiri* (“there about in the mountain”)
***mu*-GAC**	18	Locatives, ‘inside’	*mu-*	***mu*** *phiri* (“in the mountain”)

Note that the *zi*-ANC class in Table 6.1 include the prefixes *vi-/dzi-* that were not found in the nominal dataset used in Chapter 4. These will not form distinct ACs as they are taken to be ‘non-standard' variants of the AM *zi-*. Also note that the prefixes *ka-*, *a-, chi-* and *u-* in Table 6.1, which were considered as deriving subsets of nouns in the context of different ANCs in Chapter 5, additionally perform the indicated secondary functions. However, I will argue that the secondary function being considered in the present chapter involves a structurally different encoding from the primary function discussed in Chapter 5. This implies that the respective affixes are underspecified and hence ‘recycled’ by the so-called secondary functions.

[2] There is no English equivalent to the honorific form; as such we will append the [HON] to all glosses of this nature.

Stellenbosch University https://scholar.sun.ac.za

From the examples in Table 6.1, three different types of expressions can be identified. The first group includes nouns with prefixes that mark augmentation, diminutivisation, etc., giving us what has cross-linguistically been referred to as *evaluative, expressive* or *affective morphology* (see i.a. Scalise, 1986: 131–133; Prieto, 2005, 2015; Steriopolo, 2009; Grandi, 2011, 2015; Contini-Morava & Kilarski, 2013: 277; Déchaine *et al.*, 2014; Körtvélyessy, 2015; Kramer, 2015: 158). According to these studies, *evaluative morphology* generally refers to the phenomenon whereby **descriptive** and **connotative** diminutive, augmentative, pejorative, affective, endearing and other meanings are realised by means of some predictable morphological means. The second group comprises LOCPs – these are expressions that have been grammatically marked to indicate location. As will be shown in Section 6.4.1, locative markers differ significantly from the evaluative prefixes.[3] The third group includes complex phrases such as CPs, conjoined NPs, PPs, and instances where the AMs serve expletive-like functions.

With this distinction in place, I discuss evaluatives, locatives and non-nominal complex expressions in separate sections, namely Section 6.3; 6.4 and 6.5, respectively.

6.3 Evaluative and affective expressions in Chichewa

Like many other Bantu languages, nouns in Chichewa can be morphologically marked to indicate denotative evaluative meanings (e.g. literal size) or various speaker-hearer connotative meanings (e.g. endearment) regarding a particular referent. For example, as introduced above, a noun can be marked to denote that something is small/tiny by means of a singular DIM marker *ka-* or a plural DIM marker *ti-*. Similarly, a noun can be marked to express the notion 'big/huge' by means of a singular AUG marker *chi- or* the plural AUG markers *zi-, dzi-, or vi-*. In addition to functioning as DIM and AUG affixes, these prefixes are also 're-used' to mark more subjective speaker-hearer perspectives such as pejoration, endearment, amelioration, etc. This kind of multiplicity of semantic functions of affixes is also observed in the prefixes *a-* (NC2) and *u-* (NC14). Therefore, in addition to serving as the plural prefix, *a-* is also used to mark nouns of human reference as honorific, whereas the prefix *u-* is further 're-used' to convey speaker attitudes about a particular referent, for

[3] Maho (1999: 97) argues that "secondary classification usually does not entail a change of syntactic status; that is, secondarily classified nouns are still nouns, and as such they govern concords as they did before, albeit new ones." While Maho's (1999) views are correct for the evaluative nouns, Chichewa locative phrases do not seem to be in the same nominal category. I motivate this position in Section 6.4.

Stellenbosch University https://scholar.sun.ac.za

example referring to someone's behaviour as stupid, foolish, desirable, etc. In Steriopolo's (2009: 149) study of Russian, similar affixes (as *u*-) have been referred to as "*attitude"* affixes. The multiplicity of semantic functions associated with evaluative forms, such as the DIM and AUG markers, has been observed to be a crosslinguistic phenomenon (see i.a. Haas, 1972: 148; Jurafsky, 1993: 423, 1996: 534; Steriopolo, 2008, 2014; Körtvélyessy, 2015). Usually, the speaker-hearer perspective interpretations of these morphemes are context-dependent.

It should be stated that evaluative and affective expressions in Chichewa and related Bantu languages have generally been approached from a noun class perspective (see i.a. Carstens, 1991: 32; Matiki, 2001: 81; Contini-Morava, 2002: 25–28; Déchaine *et al.*, 2014). With the exception of the analysis provided by Déchaine *et al.* (2014), the noun class approach to evaluative and affective affixes in Bantu has not revealed the real properties of evaluative morphology in Bantu languages in general and Chichewa in particular. However, crosslinguistic studies of evaluative and affective morphology have revealed properties that appear to be similar to the Chichewa case. The literature on evaluative morphology points to Scalise (1986: 131–133), who observes that evaluative morphology differs from the typical derivational and inflectional morphology; it has consequently sometimes been designated *the third morphology* (see Körtvélyessy, 2015: 1). Research into evaluative morphology has since Scalise (1986) illuminated many aspects of the nature of evaluative morphology: for example, its grammatical properties (see i.a. Scalise, 1986: 131–133; Melissaropoulou & Ralli, 2008; Steriopolo, 2008, 2014; Grandi, 2011), its semantic properties (see i.a. Jurafsky, 1996; Prieto, 2005, 2015), its pragmatic aspects (Dressler & Merlini Barbaresi, 1994), and its sociolinguistic aspects (Daltas, 1987; Gaeta, 2015). From this rich body of literature, comes a wide range of insightful observations that are also relevant to the way in which I analyse evaluative and affective morphology in Chichewa. Some of the most relevant observations to the present discussion are summarised in (4).

(4) a. Evaluative affixes (e.g. AUG, DIM) have multiple functions. For example, they (i) may convey denotative meanings of the referent as small, big, etc. (ii) convey connotative meanings largely influenced by the speaker's subjective perception of the referent and (iii) perform speech act functions such as managing 'politeness' (see i.a. Dressler & Merlini Barbaresi, 1994; Jurafsky, 1996; Steriopolo, 2009, 2014; Körtvélyessy, 2015; Prieto, 2015; Ponsonnet, 2018).

b. Evaluative affixes may attach to various types of bases, e.g. nouns, adjectives, pronouns, adverb-verbs, etc. (Bauer, 1997; Prieto, 2005; Wiltschko & Steriopolo, 2007).

Stellenbosch University https://scholar.sun.ac.za

c. According to Prieto (2005: 12), evaluative morphemes can be analysed as having properties of 'degree words' such as *much, very,* which are accounted for as functional heads in some approaches (Abney, 1987; Cinque, 1999). It may thus be that evaluative morphemes can also been considered as functional heads (Prieto, 2005: 12).

In the next sections, I consider each of the affixes that convey evaluative and affective meanings in Chichewa. I will show that most of the cross-linguistic observations regarding evaluative morphemes given in (4) apply to Chichewa.

6.3.1 Diminutive, ameliorative, and pejorative expressions

The previous literature identified diminutive expressions in Chichewa as forming two independent noun classes, NC12 for the singular DIM and NC13 for plural DIM.[4] Note that the singular DIM prefix *ka-* is similar to the *ka*-prefix found in the *m(u)-a*-NC (≈NC1) and the *ka-* used for deriving manner nominals, as discussed in Chapter 4, Sections 4.3.3 and 4.10, respectively. Although the singular DIM prefix *ka-* is accidentally homophonous with the other *ka*-prefixes discussed in Chapter 4, I will show in Section 6.3.5 below that there is a common pattern with evaluative morphology to 'recycle' affixes from other domains of the nominal structure. The singular and plural diminutives are illustrated in (5a) and (5b), respectively.

(5) a. **ka**-chi-wala
SG.DIM-SG-grasshopper
"The/a tiny grasshopper"

b. **ti**-zi-wala
PL.DIM-PL-grasshoppers
"(The) tiny grasshoppers"

As was pointed out already, the DIM prefixes *ka-* and *ti-* are also used to mark a wide range of speaker-hearer perspectives. Firstly, the diminutive import may sometimes be used to mark loathing or lack of respect, such as shown in (6).

(6) a. ka-m-tsogoleri ka-nu
SG.DIM-SG-leader AM-2ND PRS.SG.POSS
"Your miniscule leader"

b. ti-a-tsogoleri ta-nu
PL.DIM-PL-leader AM-2ND PRS.PL.POSS
"Your miniscule leaders"

[4] Some studies have claimed that some Bantu languages have developed innovative diminutive prefixes, e.g. that a word for 'child' has grammaticalized into a DIM marker (Bresnan & Mchombo, 1995: 216; Maho, 1999: 91). However, if we take into account Jurafsky's (cf. 1993: 425, 1996: 542) universal structure of the semantics of the diminutive, this seems to be an inaccurate application of the diachronic account. In many languages, the core senses of the diminutive are 'child', 'affection' and 'small' (cf. Heine, Claudi & Hunnemeyer, 1991; Matisoff, 1991). This means that the use of the word *child* as a diminutive in Bantu may well reflect the original source form, and not a specialised innovation.

Stellenbosch University https://scholar.sun.ac.za

In (6), the leader(s) may not necessarily be small in stature, but, according to the speaker and the prevailing context, the referents may have an undesirable leadership style, or other properties triggering a negative evaluative perspective.

The third interpretation of the DIM prefix is almost the opposite of (6): the prefixes may be used to convey meanings related to something good. For example, in addition to meaning "small" in (7a) and (7b) below, *ka-* and *ti-* may also be used to convey endearing connotations - "cute child" and "cute twins", respectively.

(7) a. **Ka**-mwana **ka**-ku-mwetulira.
SG.DIM-SG.child AM-T/A-smile
"The cute baby is smiling."

b. **Ti**-mapasa **ti**-ku-gona.
PL.DIM-PL.twins AM-T/A-sleep
"The cute twins are sleeping."

All these meanings are superimposed on the default literal size diminutive meaning, such that the various connotations are dependent on the speaker's or hearer's perspective. Similar behaviour is also observed with AUG morphemes, to which I turn next.

6.3.2 Augmentative, ameliorative and pejorative expressions

Expressions under consideration here are those observed to belong to the *chi-*ANC and *zi-*ANC. To begin with, it should be stated that the grammatical processes involving AUG affixes have not received much attention in the previous literature on Chichewa. One of the reasons for this lack of attention could be because the AUG expressions triggers the AMs *chi-* and *zi-*, which are 'homophonous' with the traditional NC7 and NC8/10 AMs, respectively. As such, the augmented nouns are not identified as forming a separate noun class (cf. Carstens, 1991: 32), as is the case for the DIMs just discussed. Some studies have even confused the non-AUG prefix *chi-* with the AUG prefix. For example, Matiki (2001: 81) describes the Chichewanized English loanword *chi-goli* (SG-goal -> "score") as being derived by the AUG prefix. Although Matiki makes this claim, if we look beyond the homophony, it is clear that the meanings and grammatical means of deriving the AUG and the non-AUG *chi*-initial nominal expressions are quite different. For example, as discussed in Chapter 4, Section 4.8, prefix *chi-* is mainly the singular number-cum-derivational prefix that attaches to nominal stems, whereas the AUG *chi-* only adjoins to full NPs (and not to

Stellenbosch University https://scholar.sun.ac.za

nominal stems) to convey meanings related to the size of a particular referent as being big/huge.[5] Consider example (8) in this regard.

(8) a. **chi**-galu
SG.AUG-dog
"the/a big dog"

b. **zi**-a-galu / **z**-a-galu
PL.AUG-PL-dogs
"(the) big dogs"

In (8b), the plural augmentation form can also be shortened by silencing the vowel on the AUG prefix; hence the form *zagalu*. As with the DIM prefix, it is not only the augmentative meaning that is conveyed by the AUG prefixes; sometimes the augmentative expression may connote two contradictory meanings relating to amelioration/derogation, depending on the speaker's or hearer's perspective. A well-articulated distinction between these morphemes is also presented in Déchaine *et al.* (2014: 20–26), where similar prefixes in Shona are analysed as playing both descriptive and expressive functions. In the present study, the non-evaluative forms are the ones discussed in Chapter 4, Section 4.8 above, i.e. those that are observed to mark number and/or derive non-augmentative nouns from nominal stems and other nouns.

In some varieties of Chichewa, the plural form of the AUG *chi*-initial nouns is marked by the prefixes *dzi-* and *vi-* carrying both denotative and connotative evaluative meanings. However, in varieties where these 'non-standard' forms are not prevalent, they are mainly used to mark speaker-hearer perspectives. These 'non-standard' forms are usually not accepted in prescriptive grammars in Chichewa. Both the standard and the 'non-standard' prefixes are represented in (9).

(9) a. **chi**-m-nyamata
SG.AUG-SG-N.STEM
"(big), good looking boy"

b. i. **zi**-a-nyamata
ii. **dzi**-a-nyamata
iii. **vi**-a-nyamata
PL.AUG-PL-N.STEM
"(big), ugly boys"

c. i. **zi**-a-nyamata
ii. **dzi**-a-nyamata
iii. **vi**-a-nyamata
PL.AUG-PL-N.STEM
"(big), good looking boys"

Example (9a) illustrates the standard AUG morpheme, which also carries complimentary connotations while examples (9b, ii-iii) and (9c, ii-iii) depict nouns with the two 'non-standard' augmentative morphemes, *vi-* and *dzi-,* both of which necessarily produce ameliorative or pejorative connotations. Note that the AUG prefixes always attach to a noun with number

[5] The distinction between number and evaluative affixes will become clear in Section 6.3.5, where I propose the structures that each take.

Stellenbosch University https://scholar.sun.ac.za

prefix already attached, effectively showing that the AUG and the non-AUG prefix *chi-* play different functions (cf. example (3)).

Next, I look at honorific expressions in Chichewa, another type of evaluative morphology which has not received systematic attention in the traditional literature.

6.3.3 Honorific expressions

As introduced in the discussion of the *a*-ANC in Chapter 4, Section 4.5.1, in Chichewa and many other Bantu languages, a distinction as to whether one is talking to or about someone younger/familiar or one who is older/respected/someone one is not familiar with is **grammatically** signalled. This distinction is reflected in personal pronouns and also in nouns of human or personified referents. For example, consider the personal pronouns in (10).

(10) a. iwe
SNG 2ND PRS
"you"

b. inu
HON.SNG 2ND PRS /HON.PL 2ND PRS/ PL 2ND PRS
"you [PL or HON]"

c. iye
SNG 3RD PRS
"him/her"

d. iwo
HON.SNG 3RD PRS / HON.PL3RD PRS /PL 3RD PRS"
"him/her/them [HON]"

In Chichewa, one cannot refer to someone respected as *him/her* using the regular 3rd person pronoun form; instead, the form that must be used is the Chichewa equivalent of the English plural pronoun *them*. The same holds in contexts of direct address, where a respected person must be addressed using a plural second-person form.

This grammatical feature is also reflected on full NPs via the affix *a-* marked on the noun itself. Compare the examples in (11) and (12), where (11) are honorific, while (12) are non-honorific nominal expressions.

(11) a. A-Chipala a-bwera.
HON-NAME OF PERSON AM.T/A-come
"[HON] Mr/Mrs Chipala has come."

b. A-phunzits-i a-bwera.
HON-teach-NV AM.T/A-come
"[HON]The teacher has come"
"(The) teachers have come."

Stellenbosch University https://scholar.sun.ac.za

(12) a. Chipala wa-bwera.
NAME OF PERSON AM.T/A-come
"Chipala has come."

b. Mphunzitsi wa-bwera.
teacher AM.T/A-come
"The/A teacher has come."

Honorific marking is such an important grammatical feature of Chichewa such that, in imperatives, the verb must also encode whether one is being honorific or not. Consider examples in (13).

(13) a. bwera
"come"

b. bwera-ni
come-HON/PL
"You [PL/HON], come."

As with the other evaluative affixes, the HON prefix is also used to mark speaker-oriented connotative meanings. For example, in a context where honorific marking is not necessary, the prefix *a-* may be used to ameliorate a criticism, ridicule, displeasure, etc. Consider (14).

(14) a. *Context: Someone talking to a close friend, John.*
A-Joni mwa-ledzera-nso.
HON-John SM.HON.T/A-drunk-again
"Good John, you are drunk again."

b. *Context: Mother talking to her baby.*
A-bebe mwa-phwanya botolo.
HON-baby SM.HON.T/A-break-APP-FV bottle
"Oh! Dear baby, you have broken the bottle."

As illustrated in (14a), the speaker has the option of not using the HON prefix without offending the addressee; however, in this case, it is used to express a somewhat ameliorated criticism. Equally, in (14b), there is no logical need for the mother to use the honorific prefix apart from expressing some kind of ameliorated worry or displeasure towards the baby. The honorific morpheme is therefore a strategy that may be used to ameliorate any kind of criticism or disapproval.

Except where the HON is attached to personal names, the *a*-affix takes the form of a plural marker on some nouns in the *m(u)-a-*ANC (≈NC1). As introduced in Chapter 1, Section 1.3, this multifunctionality of the prefix *a-* shows that the honorific feature is exponed by means of underspecified affixes in Chichewa; hence the syncretism. Like the AUG morpheme, the morpho-semantic features of the HON have not been well understood in Chichewa. Although Matiki (2001: 66) claims that the NC2 prefix is used to denote singular honorific

Stellenbosch University https://scholar.sun.ac.za

forms for NC1 in general, more careful investigation shows that the HON prefix is used to refer to human or personified entities only. In other words, the HON feature only applies to entities belonging to the highest category in the animacy hierarchy. Therefore, non-human nouns in NC1 may not have the honorific reading (cf. Bresnan & Kanerva, 1989: 39). For example, the noun *a-galu* ("dogs") cannot be interpreted as honorific unless a particular dog has been personified in a poetic sense; otherwise the form is interpreted unambiguously as plural on such nouns.

One may argue that it is the actual plural function that is used to mark HON. However, there are two pieces of evidence that indicate that the HON prefix is structurally not the plural prefix *a-* per se. Firstly, we find the HON prefix with people's names, structures that are not usually associated with traditional NC2 plural number prefixes in Chichewa. Consider the following examples.

(15) a. a-Msaka
HON-Msaka

b. a-Maria
HON-Maria

Secondly, in what is described as 'non-standard' Chichewa, it is common to use the default plural marker *zi-* with nouns of the *m(u)*-a-ANC as in (16c) and (17c).

(16) a. m'busa
priest

b. a-busa
PL/HON-priest
"[HON]priest(s)"

c. a-zi-busa
HON-PL-priest
"[HON]priests"

(17) a. mayi
mother

b. a-mayi
PL/HON-mother
"[HON]mother(s)"

c. a-zi-mayi
HON-PL-mother
"[HON]mothers"

Note that examples (16b) and (17b) are ambiguous in the sense that they can be interpreted as honorific-singular or honorific-plural. However, the counterpart examples (16c) and (17c) are unambiguously interpreted as honorific-plural. It should be noted that examples (16c) and (17c) are usually criticised by grammar purists as ungrammatical, but these expressions are commonly attested in everyday spoken Chichewa – e.g. when one is addressing a group of people as in *azimayi ndi azibambo* "ladies and gentlemen". In addition to this, a simple corpus search of the words *azibusa* and *azimayi* also reveals that this is a common strategy for disambiguating between the honorific-singular and honorific-plural reading, even in some

Stellenbosch University https://scholar.sun.ac.za

written texts (cf. Fortune, 1955: 64 endnote 7).[6] The choice of the affix *zi-* in this context is not surprising as this prefix is usually used as a default plural marker and AM (see again Chapter 5, Section 5.6).[7]

As regards the hierarchical relations between the HON affix and the evaluatives AUG and DIM, it appears the HON is structurally higher. Evidence for this proposal comes from the following:

(18)	a.	munthu	b.	chi-munthu	c. a-chi-munthu
		person		AUG-person	HON-AUG-person
		"person"		"a big person"	"[HON] the very big person"

The example in (18) is taken to be in a context where someone is referring to a person whose name they do not know, but whom they simultaneously want to show respect towards. The order of affixes shown in (18c) is acceptable but not the other way around as in **chi-a-munthu*. What examples (17) and (18) show, however, is that the honorific use of the affix *a-* is not just a subtle notion embedded within plural marking; it is an independent grammatical function that must be identified separately in a similar manner to the DIM and AUG affixes. Crucially, as illustrated in (16c), (17c) and (18c) the HON affix attaches above the number and denotative evaluative AUG projections (see Section 6.3.5 for further structural analysis).

In addition to the above, I also observe that the prefix *u-* (the so-called NC14 prefix) is also used for evaluative purposes in Chichewa. I briefly introduce this function in the next section.

6.3.4 Attitude-conveying expressions

Having considered the less controversial evaluative expressions in the preceding sections, I would now like to extend my observations to some nouns derived via the prefix *u-*. As introduced in Chapter 4, Section 4.6.2, the *u*-ANC (≈NC14) includes nouns exhibiting a

---

[6] I did a word search on the following Chichewa-based website, https://wol.jw.org/ny/wol/d/r107/lp-cn/2008092. There is also a name of a business entity called *azimayi ndi azibambo* ("ladies and gentlemen") in the southern district of Chikhwawa in Malawi. The business is listed on the following website: https://www.africanadvice.com/1366789/General_Dealers/Malawi/Nansolola_Ig_(azimayi_Ndi_Azibambo_Store)_(1966)__Ltd/

[7] For example, while traditional NC9 is usually not morphologically marked for number, in certain contexts where one wants to avoid the number ambiguity, the prefix *zi-* is usually used. Consider *nyumba* ("house/s") which is realised as *zi-nyumba* ("houses"). This example serves to illustrate that the affix *zi-* is sometimes used as the default number marker.

Stellenbosch University https://scholar.sun.ac.za

range of morpho-semantic characteristics. For example, some nouns are inherently *u*-initial while in other nouns, the initial *u*- is a prefix that attaches to adjectival, nominal and verbal stems. Among this latter type, there are cases where the prefix *u*- only attaches to full NPs. In such cases, the prefix is also observed to convey both denotative and connotative affective meanings. The denotative meaning is usually related to nouns conveying what can be described as '-i*sm*-like' (ISM) meanings.[8] These are illustrated in (19).

(19) a. u-chidakwa
ISM-drunkard
"alcoholism"

b. u-kazembe
ISM-diplomat
"diplomatic relations"

c. u-batiz-o
ISM-baptise-NFV
"baptism"

The morphological pattern illustrated in (19), where the prefix *u*- states particular qualities or features of an entity, is also observed in examples that express more subjective, speaker-oriented evaluative meanings. More specifically, the evaluative meanings expressed by what Steriopolo (2009: 149, 2013: 35) refers to as *attitude* (ATT) affixes, which convey the speaker's attitude towards a particular referent. Consider the examples in (20).

(20) a. u-galu
ATT-dog
"stupidity"

b. u-mbuzi
ATT-goat
"stupidity"

c. u-chi-nyama
ATT-AUG-animal
"savagery/heartlessness"

In (20), the prefix *u*- is used to convey the speaker's perceptions of the referent, for example, somebody who is perceived as having characteristics of a dog, a goat, etc. This connotative use of the prefix *u*- clearly differs from the denotative function observed with expressions such as those in (19), which does not convey the subjective construal. Although all the ATT expressions belong to the *u*-ANC, it is necessary that a distinction be drawn between the denotative use of the prefix *u*- and the connotative use. The evaluative use of the prefix *u*- is different from the denotative version of the prefix *u*-: as shown in (20c), the evaluative function linearly precedes - and, thus, apparently structurally dominates - the denotative AUG prefix. As stated in Chapter 1, the multiple semantic interpretations of the same affix are assumed to originate from different syntactic positions, the idea that constitutes our chief focus in the next section.

[8] The word *-ism*, is used as cover term for expressions that refer to the state of being, such as having a particular quality (e.g. *magnetism, Darwinism*) or an attitude, style, practice (e.g. *racism, modernism*), etc.

Stellenbosch University https://scholar.sun.ac.za

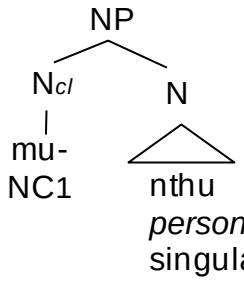
6.3.5 The formal structure of evaluative and affective expressions

In the preceding sections, I have shown that Chichewa employs at least four different types of prefixes to convey both denotative and connotative meanings, namely DIM, AUG, HON and ATT. Each of these evaluative affixes is observed to play more than one semantic function. As Déchaine *et al.* (2014: 26) state, this multifunctionality of evaluative affixes could be taken as a case of accidental homophony (i.e. the same form with distinct lexical entries) or as a case of underspecification (i.e. the same form occupying distinct syntactic positions) (cf. Steriopolo, 2014: 48). However, considering the regularity of the multifunctionality of affixes of this type in Chichewa, the accidental homophony approach may not be a plausible assumption (cf. Déchaine *et al.*, 2014: 26). Therefore, following Déchaine *et al.* (2014), and Steriopolo (2014) among others, I would like to argue that the multifunctionality of the evaluative and affective prefixes in Chichewa is also a case of underspecification, as introduced in Chapter 1, Section 1.3. Specifically, I propose that the evaluative and affective functions in Chichewa are realised by any of the four different underspecified prefixes discussed in the preceding sections.

In order to properly understand the syntactic structure of these expressions in Chichewa, I first recapitulate how the traditional analyses have viewed the formal properties of both the primary and secondary prefixes in Bantu. As discussed in Chapter 2, there are two popular views regarding the formal structure of Bantu noun classes. The first view assumes that both primary and secondary prefixes carry noun class features (see i.a. Mufwene, 1980; Sproat, 1985; Myers, 1987; Bresnan & Mchombo, 1995; Maho, 1999: 97; Ferrari, 2005). Under this view, it does not seem to matter whether the prefix attaches to a root/stem or to a full NP. These alternatives are schematised in (21).

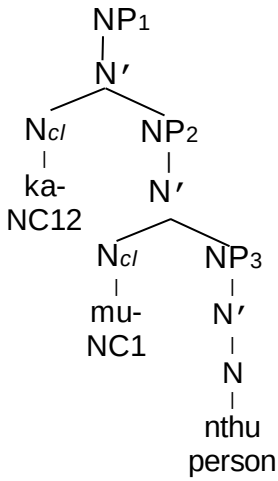
Stellenbosch University https://scholar.sun.ac.za

(21) a. *mu-nthu* (“person”)



(see i.a. Trithart, 1977: 19; Bresnan, 1995: 29; Carstens, 2008: 138)

b. *ka-mu-nthu* (“small person”)[9]



(Bresnan & Mchombo, 1995: 214)

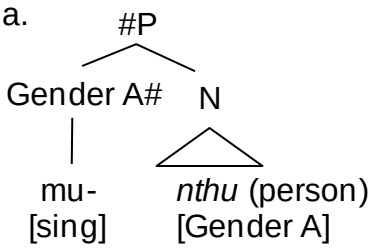
The diagram in (21a) represents the structure of a noun with the NC1 primary NCP, whereas that in (21b) depicts the complex structure, one comprising both the primary and secondary NCPs. As can be seen in (21b), the primary prefix *mu-* and the secondary prefix *ka-* are both indicated to carry noun class features. This implies that the primary and secondary prefixes are functionally the same. Although this idea of secondary classification has been assumed in many studies (see i.a. Fortune, 1955: 54–59; Harding, 1966: 24; Mugane, 1997: 174; Maho, 1999: 97), none of these works provide any justification for this problematic ‘double classification’. The view illustrated in (21b) also contravenes the widely attested constraint known as the *Obligatory Contour Principle* (OCP), which prohibits the juxtaposition of two identical phonological, morphological or syntactic elements (see i.a. McCarthy, 1988: 88; Golston, 1995: 354–356; Anttila & Fong, 2000; Mohanan & Mohanan, 2003: 302–305; Tseng, 2008).

The second view is proposed by Carstens (1991: 32–40, 2008: 138–144), who argues that the noun class features are actually inherently located on the noun stem and that the primary prefixes are gender-specific number spell outs. Similarly, the secondary prefixes are gender-specific number spell outs of the phonologically null DIM or AUG projections. These views can be summarised via simplified schemas in (22a) and (22b), respectively.

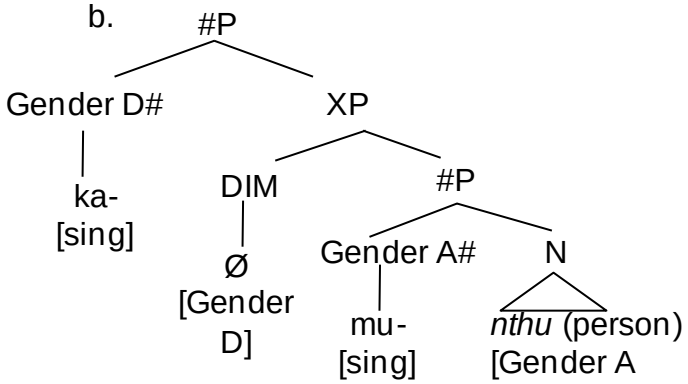
[9] The tree diagram in (21a) has been adapted from Carstens (2008: 138), while that in (21b) has been taken from Bresnan & Mchombo (1995: 214). However, it should be noted that the structures does not reflect Carstens’ (2008) and Bresnan & Mchombo’s (1995) but are depictions of analyses that they argue should be rejected..

Stellenbosch University https://scholar.sun.ac.za

(22) a.



b.



As shown in (22a) the prefix is only a Gender A-specific singular number spell out, whereas in (22b) the diminutive notion is introduced by a phonologically null affix, which is also taken to be a gender-marked projection, with the DIM prefix merely being a number spell out of that specific gender, designated as Gender D.

Despite some differences between the traditional (i.a. Bresnan & Mchombo, 1995; Mugane, 1997; Ferrari, 2005) and Carstens' (1991, 2008) views, the primary and secondary prefixes are analysed as having similar formal properties in both accounts. In the former, the prefixes carry noun class features whereas in the latter, they are analysed as carrying only gender-specific number features. Although the homophony between some of the secondary prefixes and the primary prefixes could be taken as evidence that they are one and the same thing, the different interpretations they give is a primary clue that they are functionally different. In fact, Carstens (1991: 221) already observes that the primary (non-DIM) and the secondary (DIM) affixes do not yield ambiguous interpretations.

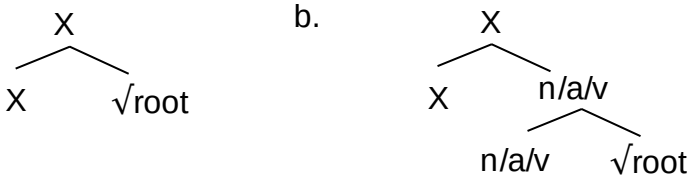
As the preceding discussion has shown, the morphosyntactic structure of evaluative and affective morphosyntax has not been as carefully studied in Chichewa as it might have been. To further illuminate the properties of secondary and evaluative prefixes in this language, I follow insights from relevant crosslinguistic studies, such as Scalise (1986: 131–133), Jurafsky (1993, 1996); Wiltschko & Steriopolo (2007); Steriopolo (2008, 2014); Déchaine *et al.* (2014); Štekauer (2015) and Biberauer (2018c).

I classify these studies into three broad approaches to evaluative and affective morphosyntax: first is a Distributed Morphology (DM) analysis (i.a. Wiltschko & Steriopolo, 2007; Steriopolo, 2008, 2014), second is an Interface Syntax model analysis proposed by Déchaine *et al.* (2014) and third, is The Peripheral Speaker-Hearer Hypothesis (PSHH) as outlined in Biberauer (2018c) (cf. Heim & Wiltschko, 2017). All the three of the analyses offer relevant insights into the formal properties of Chichewa evaluative morphology.

Stellenbosch University https://scholar.sun.ac.za

Focusing on the DM analysis first, Steriopolo (2013: 39 ff, 2014: 48–50) identifies two crucial parameters of variation in respect of how diminutives are manifested, namely whether the affixes are syntactic heads or syntactic modifiers. On the one hand, syntactic heads are taken to project, so they can change the syntactic category or grammatical features of the base, e.g. gender/noun class (Steriopolo, 2014: 51; Kramer, 2015: 228). Such affixes are observed to attach to both roots and syntactically specified categories, as schematised in (23a) and (23b), respectively.

(23) a. b.



Syntactic modifiers, on the other hand, "do not project, they do not determine a syntactic category or grammatical features of the output" (Steriopolo, 2014: 53). Consider the diagrammatic schema in (24a).

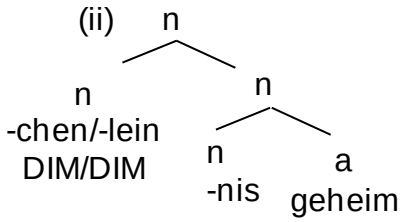
(24) a. DIM as a syntactic modifier b. DIM as a syntactic head

The DIM affix illustrated in (24a) does not change the expression's syntactic category of the base, implying that it is a syntactic modifier whereas in (24b), the DIM affix morpheme changes the syntactic category of the base, from Y to X. According to Steriopolo (2014), evaluative morphemes are found to vary crosslinguistically as to whether they are syntactic heads or syntactic modifiers. For example, DIM affixes in German are heads whereas in Halkomelem they are modifiers (Wiltschko & Steriopolo, 2007: 5; Steriopolo, 2013: 35–37, 2014: 50–51).

(25) a. German DIM (head) (Wiltschko & Steriopolo, 2007: 5–6)

(i) geheim-nis-chen
secret-NOM-DIM
"(cute) little secret"

(ii)

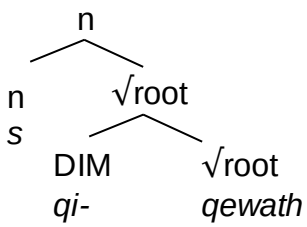


Stellenbosch University https://scholar.sun.ac.za

b. Halkomelem DIM (modifier) (Wiltschko & Steriopolo, 2007: 5)

(i) s-qi-qewath
NOM-DIM-rabbit
“little rabbit”

(ii)



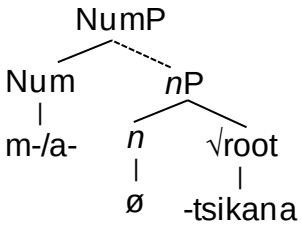
What is crucial for the present study is Steriopolo’s (2014) analysis that languages vary regarding whether the evaluative morphemes are located higher or lower in the nominal structure.

To apply the foregoing DM analysis to Chichewa, I first need to clarify the structure of the so-called primary prefixes before I propose the structure of evaluative and affective expressions. As argued in Chapter 3, the primary prefixes in Chichewa appear to play two key functions: they function predictably as a number prefix and, less predictably, as derivational affixes as in (26) and (27), respectively. The number prefixes are observed to attach to nominal stems (cf. Sections 3.3 and 3.4 of Chapter 3); that is, they do not function as nominalisers that attach to category-neutral roots. These are illustrated in (26a-b) and schematised in (26c).

(26) a. m-tsikana
SG-nominal stem
“the/a girl”

b. a-tsikana
PL-nominal stem
“girls”

c.



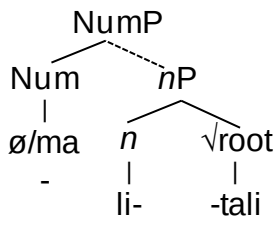
In (26), the primary prefixes are shown to attach to nominal stems (*n*P) to mark number, with the singular and plural prefixes substituting each other. However, in nouns where number marking is additive, the plural number prefix attaches after the first set of primary prefixes, which I have analysed as derivational prefixes (DPFs) in (27c) (see also Chapters 3 and 4).

Stellenbosch University https://scholar.sun.ac.za

(27) a. li-tali
DPF-long
“length”

b. ma-li-tali
PL-DPF-length
“lengths”

c.



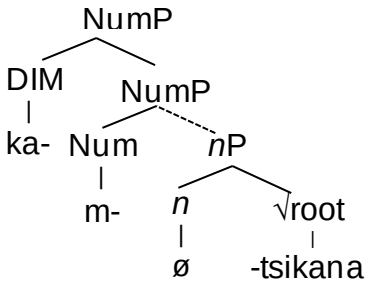
Since the prefix *li-* is below the number projection and its function differs from that of plural marker *ma-*, I assume that *li-* only serves the derivational function, as shown in (27c). Therefore, the so-called primary prefixes also fall into two categories, namely those that mark number (e.g. (26a-b)) and those that strictly play the derivational functions (e.g. (27a)).

The morphosyntactic properties of primary prefixes like those illustrated in (26-27) differ from those of evaluative prefixes, which are observed to only attach to NPs that have already undergone the derivation and number marking morphological processes. Consider the following examples.

(28) a. ka-mtsikana
SG.DIM-SG.girl
“the/a small girl”

b. ti-a-tsikana
PL.DIM-PL.girl
“small girls”

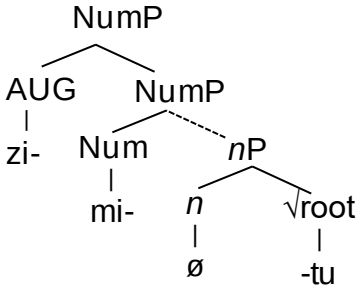
c.



(29) a. chi-mutu
SG.AUG-SG.head
“big head”

b. zi-mitu
PL.AUG-PL.head
“big heads”

c.



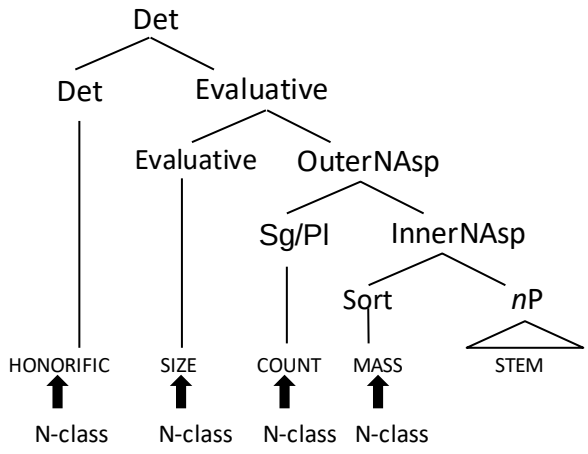
As shown in (28) and (29), the DIM and AUG prefixes attach to complex nominal phrases (NumPs or NPs) and not to nominal stems (*n*P). If Steriopolo’s (2014) analysis is on the right track, evaluative prefixes in Chichewa can be taken to be modifiers of nominals (cf. Carstens, 2008: 145). Crucially, the evaluative and affective prefixes are observed to consistently attach above the number projection such that they cannot be implicated in the nominalising morphology. However, since the evaluative function is observed to recycle the affixes from the number and nominalising morphology, one needs to be careful not to confuse the two functions. What is consistent in Chichewa is that when the affixes *u-, chi-,*

Stellenbosch University https://scholar.sun.ac.za

a- and *zi-* are found on the number projection or below, they do not serve any of the evaluative or affective functions. However, when these affixes are above the number projection, they turn to convey evaluative and affective meanings. Given this background, the relevant question would be how to explain this distribution.

Dechaine *et al*. (2014:26) argue that the multi-functionality associated with evaluative elements in Shona and other Bantu languages is structurally determined; that is, it falls out from "the same morpheme occupying distinct syntactic positions" (cf. Embick & Noyer, 2007: 299–300). Regarding syntactic positions, it is hypothesised that core functional (F) categories are universal, but languages differ as to how prolific their F-category inventories can be. These core F-categories form a universal abstract syntactic spine which is realised as a verbal spine and a nominal spine (cf. Ritter & Wiltschko, 2009). Drawing parallels with the verbal spine, Déchaine *et al.* (2014: 27) argue that in Shona and Bantu languages in general, "the CLASS domain subdivides into the nominal equivalent of inner and outer aspect." The INNER NOMINAL ASPECT (InnerNAsp) corresponds to the classifying function applicable to mass nouns, namely sorting. The OUTER NOMINAL ASPECT (OuterNAsp) corresponds to the classifying function applicable to count nouns, namely the singular/plural contrast. The projection above OuterNAsp includes evaluatives, which are argued to be the extension of the classifying function. The topmost projection of the spine is associated with the D position. According to Déchaine *et al.* (2014: 28), the D position is occupied by the so-called honorific NCP in Shona. The full structure of this nominal spine is schematised in (30) (cf. Déchaine *et al.*, 2014: 28).

(30)



Take note that Déchaine *et al.* (2014) identify the traditional NCPs as belonging to the mass and count projections, agreeing with Carstens' (1991) and my observation that the so-called NCPs are predictably number affixes. However, the proposed structure in (30) is not

Stellenbosch University https://scholar.sun.ac.za

compatible with certain crosslinguistic and Chichewa-specific aspects of evaluative morphology. First, the hypothesis is based on the traditional Bantu noun class system, which I have shown in chapters 3, 4 and 5 that does not hold for the facts in Chichewa. Therefore, although the structure reasonably captures the order of Fs, the analysis of prefixes as representing noun classes cannot be extended to Chichewa, especially considering the data types and the ANC system proposed in Chapter 5. I therefore adopt the various F projections in (30) but modify the status of the affixes involved (see 33 below).

Second, the suggested structure in (30) fails to account for the pattern observed in Halkomelem, Russian and Kiswahili, where evaluative affixes are also observed to attach to roots, which are positions lower than InnerNAsp in the nominal spine (see i.a. Carstens, 1991: 34; Wiltschko & Steriopolo, 2007: 5; Steriopolo, 2013: 35–37, 2014: 53). Therefore, it could not generally be the case that evaluative morphology is an extension of the classifying function, as analysed by Déchaine *et al.* (2014: 28). Equally, this also concerns how evaluative morphemes interact with number. In some languages, the evaluative morphemes attach before the number affixes while in other languages they attach after the number affixes (Steriopolo, 2013: 33). The variations regarding the structural position of evaluative and affective prefixes in different languages, however, certainly take a systematic pattern. According to Biberauer (2018c), the fractal make-up of phrase structure leads us to expect evaluative morphology to surface at a number of 'edges', with languages, in principle, having the choice as to which edges to exploit. In this regard, Déchaine *et al.'s* (2014) proposed nominal spine still makes the correct prediction but only needs to be modified regarding the observation that languages will differ as to which edges to exploit for evaluative markers. For example, languages like Russian can be said to place their evaluative morphemes at the edges located lower than the OuterNAsp, while languages like Chichewa and German place their evaluative affixes above OuterNAsp.

Third, although Dechaine *et al*. (2014:28) argue that the HON prefix is a determiner in Shona, in Chichewa the HON prefix exhibits similar characteristics to the other evaluative and affective morphemes. As with the other evaluative affixes, the HON is exponed via the underspecified affix *a-,* which is also used as a plural marker for some nouns in the *a*-ANC. As discussed in Section 6.3.3, in Chichewa, the HON only applies to nouns of human reference and does not show any properties related to determiners. What appears to be the case is that languages - even closely related ones like Shona and Chichewa - may "recycle" different elements to serve evaluative and affective functions. However, both in Shona and

Stellenbosch University https://scholar.sun.ac.za

Chichewa, the HON function is located higher than the descriptive evaluative morphemes, AUG and DIM.

Fourth, according to Dechaine et al. (2014:28), the traditional NC14 prefix *u-* only participates in nominal aspect typing, specifically as a mass or count marker. However, this is not the case in Chichewa where the prefix *u-* appears in two different positions. Firstly, it appears as a derivational prefix below the number-hosting position as shown in (31).

(31) a. ma-u-fulu
PL-DPF-N.STEM
"freedoms"

b. ma-u-dindo
PL-DPF-N.STEM
"responsibilities"

The structural position shown in (31) fits the analysis offered by Dechaine et al. (2014:28). Secondly, however, when the prefix *u-* also appears above the number-hosting projection, it conveys evaluative functions. Consider the examples in (32).

(32) a. u-chi-nyama
ATT-SG.AUG-animal
"heartlessness/ savagery"

b. u-zi-bambo
ATT-PL-father
"manliness/masculinity"

As shown in (32), the evaluative affix *u-* comes before and, thus, structurally above the OuterNAsp projection. The crucial aspect of the prefix *u-* in (32) is that the interpretation of such meanings is generally dependent on the speaker-hearer's evaluation, in this case, identifying someone's behaviour as typifying savagery or masculinity. Further note that the attitude-marking affix *u-* in (32) is located above the evaluative AUG *chi-,* implying that the more subjective evaluative function is also further above the descriptive evaluative projection. Therefore, although the non-evaluative function shown in (31) is similar to the Shona facts as analysed by Dechaine et al. (2014:28), the evaluative function in Chichewa demonstrates another variation in the way the two languages recycle elements for evaluative functions.

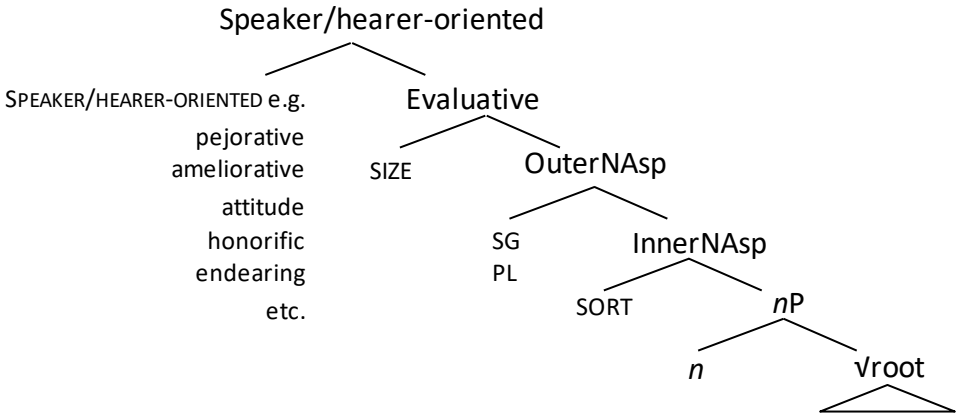
Given the foregoing background, the picture that emerges is as follows. In Chichewa, there is a set of affixes that are used for nominal derivation and number marking, namely *chi-, zi-, a-* and *u-* which are also 're-used' to express additional meanings related to descriptive evaluation and also affective and speaker-hearer perspectives. In their former function, these affixes are found lower in the nominal spine, while in their evaluative and affective functions the affixes are found higher on the nominal spine. Specifically, the descriptive evaluative functions (i.e. DIM and AUG) are located immediately above the number-hosting projection while the more subjective evaluative functions and the other speaker-hearer

Stellenbosch University https://scholar.sun.ac.za

perspective and their related speech act attenuation or intensification functions are observed to occur above the basic evaluative projections. In this regard, the topmost projection for the Chichewa nominal spine agrees with the widely attested constituent-edge-based occurrence of speaker-hearer perspective marking elements (cf. Prieto, 2005; Heim, Keupdjio, Lam Wai-Man, Osa-Gómez, Thoma, *et al.*, 2016; Wiltschko, 2017; Biberauer, 2018c).

Drawing insights from Cinque's (1999) hierarchy of clausal functional projections, Prieto's (2005) structure of evaluative morpho-syntax in Spanish, Déchaine *et al.*'s (2014) nominal spine, Steriopolo's (2014) analysis of Russian and Biberauer's (2018c) Peripheral Speaker-Hearer Hypothesis (PSHH), I propose modifying the schema in (30) as follows.

(33) **The nominal spine for Chichewa** (Adapted from Déchaine *et al.*, 2014: 28).



In (33), the domain ranging from the bottom to OuterNAsp projection represents the region where the so-called primary prefixes attach. Following Déchaine *et al.* (2014) and as introduced in Chapter 3, in Chichewa some of the prefixes are involved in nominalising various roots which derive the *n*P, while some are involved in the sort function which is associated with mass, ISM-denoting nouns. The OuterNAsp represents the projection where number-marking affixes attach, as observed on count nouns. The domain above OuterNAsp represents projections that carry descriptive evaluative affixes such as DIM and AUG. As shown in (18) and (32) above, the honorific and attitude prefixes precede the DIM and AUG and not vice versa. Equally, as shown in Sections 6.3.1 to 6.3.4 all the other speaker-hearer evaluative construals are all observed to precede the descriptive evaluative projections (DIM, AUG). Based on this evidence, I propose that the honorific, attitude and other subjective speaker-hearer functions could be above the descriptive evaluative projection as schematised in (33). Therefore, the topmost projection represents a wide range of speaker-hearer and speech act attenuation or intensification functions. As has been shown in the

Stellenbosch University https://scholar.sun.ac.za

foregoing discussion, the affixes that attach to the region above the OuterNAsp are largely recycled from the lower domains.

Although the affixes *chi-*, *zi-*, *a-* and *u-* fit very well into the proposed structure, the DIM affixes exhibit a different pattern. First, although the singular DIM affix *ka-* is similar in form to many other *ka*-affixes, such as the derivational affix for *m(u)-a*-ANC and manner nominal, these appear to be accidentally homophonous because these *ka*-s do not appear in similar nominal environments. The prefix *ka-* that derives a subset of nouns in the *m(u)-a*-ANC and the DIM *ka-* do not control similar AMs. Although the manner nominal *ka-* and DIM *ka-* control the same AM, manner nominal *ka-* only attaches to verbal expressions while the DIM *ka-* attaches to concrete nouns, which are two different syntactic environments; this suggests that the prefix is not recycled as observed with the other evaluative affixes. Second, the plural DIM affix *ti-* is not homophonous with any of the identified AMs in Chichewa. For these reasons, the denotative DIM affixes do not show properties suggesting that they are recycled from the lower domains of the nominal spine as is the case with the other evaluative affixes. However, the speaker-hearer use of the DIM affixes is a result of recycling the denotative DIM affix. As I will show in Section 6.6, the AMs controlled by DIM NPs do not coincide with any AM in the ANC system; hence the DIM and manner nominals form their own GACs.

Summing up this section, I have shown that the multiplicity of meanings associated with the set of prefixes *ka-, chi-, zi-, a-* and *u-* is mainly structure-dependent. When these affixes are found lower in the nominal spine, they are observed to convey denotative meanings mainly associated with derivation and number marking. However, when these affixes are found above the number-hosting projection of the nominal spine, they are observed to convey descriptive evaluative meanings. These affixes are further 'reused' to express speaker-hearer perspectives; in this latter case, the affixes are found at the topmost left periphery of the nominal spine. The observation that speaker-hearer marking elements occupy the peripheral positions which then determine the structure of extended projections of different types has been widely attested - cf. the work started by Wiltschko (2014), which is also developed in Dechaine et al. (2014) and Biberauer (2018c).

The next section is an interim summary of the discussion focusing on evaluative expressions and their structure.

Stellenbosch University https://scholar.sun.ac.za

6.3.6 Interim summary and conclusion

So far, I have considered the morpho-semantic properties of the first group of what were traditionally identified as secondary prefixes, namely AUG, DIM, HON and ATT. I have shown that the secondary prefixes are recycled from a subset of underspecified prefixes which are also involved in the derivational and number marking morphological processes. The expressions that have been referred to as evaluative and affective are in fact a result of the so-called "primary prefixes" being reused in the upper projections of the nominal spine, specifically the domain just above the number-hosting projection. All the four types of evaluative affixes are also observed to convey more subjective speaker-hearer perspectives such as pejorative, ameliorative, endearing, complimentary, and many others. To achieve these subjective construals, the affixes are attached to the topmost projection, the domain that has been widely associated with such speech act functions (see i.a. Déchaine *et al.*, 2014; Heim & Wiltschko, 2017; Biberauer, 2018c).

In the next section, I consider the second type of predictable complex phrases, namely LOCPs.

6.4 Locative phrases in Chichewa

Having looked at expressions with evaluative morphology in the preceding section, in the current section, I discuss the second group of morphosyntactically predictable complex phrases, popularly known as *locative nouns* in the traditional literature (see i.a. Givón, 1972: 12–13, 29; Maho, 1999: 98; Buell, 2007: 105–106).

The categorial status of locative expressions remains a contentious topic in Bantu. To avoid prejudicing the matter, I thus continue to identify these expressions by the relatively neutral term of *locative phrase* (LOCP).[10] The phenomenon of locative marking in Chichewa has received significantly more scholarly attention than that of evaluative morphology. Some of the works focusing on the morpho-syntax of LOCPs in Chichewa include Trithart (1977); Bresnan & Kanerva (1989); Bresnan (1991, 1994, 1995); Carstens (1994, 1997); Bresnan & Mchombo (1995), among others. As already introduced in Chapters 2 and 4, LOCPs in Chichewa are traditionally assumed to be formed by 'prefixing' one of the three locative

[10] The term, *locative phrase* is also used in some previous studies; see i.a. Bresnan (1995:27ff); Bresnan & Mchombo (1995:199ff); and Carstens (1997:361ff).

Stellenbosch University https://scholar.sun.ac.za

markers (LOCs) *pa(-)*, *ku(-)* and *mu(-)* to nouns, as in (32a-c), or to adverbial stems, as in (32d-f).[11] However, as will become clear by the end of the present chapter, LOCs are not necessarily noun prefixes. Some studies agree that, unlike typical noun prefixes, LOCs are not morphologically bound (see i.a. Bresnan & Mchombo, 1995: 195–208; Carstens, 1997: 368; Salzmann, 2004: 15–17). However, I observe that LOCs are only bound when attached to adverbial or adjectival roots and not NPs. In this regard, I do not gloss them in (34) and subsequent examples as prefixed to NPs (a similar convention is also adopted in Bresnan (1991: 58ff), Mchombo (2004: 5ff), among others), but only to adverbial and adjectival stems.

(34) a. London — ku London — LOC London — "to/in London"

b. mwala — pa mwala — LOC rock — "on (the) rock"

c. mphika — mu m'phika — LOC pot — "in (the) pot"

d. -seli — ku-seli — LOC-ADV.STEM — "behind"
behind

e. -tali — pa-tali — LOC-ADJ.STEM — "far"
long

f. -kati — m-kati — LOC-ADV.STEM — "inside"
mid

In the next section, I provide a summary of the traditional views regarding the syntactic status of LOCPs in Bantu.

6.4.1 The traditional view of LocPs

The putative (traditional) view is implied in many studies focusing on Bantu noun classes. For example, the locative marking demonstrated in (34) above has been analysed as a type of nominal derivation; on this view, the so-called *locative nouns* thus form part of the Bleek-Meinhof noun class system (see i.a. Doke, 1927, 1965; Ziervogel & Mabuza, 1976; Bresnan & Kanerva, 1989; Bresnan, 1991, 1994; Carstens, 1997; Poulos & Msimang, 1998; Marten, 2010; Marten & Kula, 2012; van der Spuy, 2014; Caha & Pantcheva, 2015; Zeller, 2017). As is also more generally assumed for the traditional noun class system, LOCPs are then also taken to have existed as noun classes in Proto-Bantu (PB) (see i.a. Bresnan & Mchombo, 1995: 212; Salzmann, 2004: 18). In this regard, NC16 is associated with the reconstructed NCP **ku-* (referring to distal or non-specific location), NC17 with the prefix **pa-* (referring to proximate or specific location) and NC18 with the prefix **mu-* (referring to

[11] LOCPs rarely attach to nominal stems. The only example that I have come across is the nominal stem *-nthu* ("being"), as in *pa-nthu* which conveys some generic meaning ("on something").

Stellenbosch University https://scholar.sun.ac.za

interiority) (Marten, 2010: 250–251). Additional so-called locative NCPs have also been reconstructed, for example prefixes *k'a-* and *n(l)-*, identified as NC24 and NC25, respectively (Gauton, 1999: 220).[12] The locative prefixes *k'a-* and *n(l)* have only been attested in a few languages, for example in Kamba, Swahili, Tsonga, etc., where they are also reported not to be productive anymore (Gauton, 1999: 220–221). As to why locative markers are not bound like the other prefixes, it is assumed that the NCPs in general were originally elements of NPs such as determiners or articles which became bound. In this regard, LOCS are assumed not to have completed this historical process (see i.a. Greenberg, 1978: 70; Bresnan & Mchombo, 1995: 212; Salzmann, 2004: 18).[13] However, there are also contradictory claims on this matter. For example, Marten (2010) claims that siSwati LOCPs are evolving from DPs to PPs while Greenberg (1978: 70) assumes that LOCS evolved from prepositions to NCPs. Such contradictory claims obviously do not provide strong support for these diachronic assumptions.[14]

In addition to these morphological properties, there are also claims about semantic correlations between the primary noun classes and LOCPs. For example, Bresnan (1991: 62) states that, in Chichewa, locatives are "gender classes; that is, they are part of a system that signals contrasts between grammatical categorizations of people, things, locations, qualities, and the like – kinds of things (genera) designated by NPs" (cf. Bresnan & Mchombo, 1995: 183). In fact, some studies consider locatives to be one of the most semantically stable noun classes in Bantu (cf. Chapter 2, Section 2.4.3). However, this semantic stability is not really of much interest, given the range of syntactic categories that can realise locatives, for example, nouns, prepositions, adverbs, etc.

Despite the strong NP view being popular and assumed in many studies (see i.a. Loogman, 1965; Polomé, 1967; Bresnan, 1991, 1995; Du Plessis & Visser, 1992; Bresnan & Mchombo,

---

[12] Marten (2010: 252) only mentions noun class 24, which is associated with the reconstructed noun prefix **i-*, suspected to be the antecedent of the siSwati locative prefix *e-*. Gauton (1999: 220) also mentions the locative prefixes *e-* and *ke-* found in Nguni languages, but does not associate them with any noun class number in the Bleek-Meinhof schema.

[13] I consider the prefixes as free morphemes on different grounds, provided in section 6.4.2.

[14] It is also widely reported that the current grammatical status of locative phrases is very different in many Bantu languages (cf. i.a. Gauton, 1999; Thwala, 2006; Marten, 2010). For example, in Kiswahili the locative prefixes are no longer attested on nouns; locative marking is done by means of the suffix *-ni*. However, the three locative prefixes do surface in Kiswahili agreement morphology – i.e. on the verb and modifiers. In siSwati, only the prefix *ku-* is attested, but other locative marking strategies are also reported (Thwala, 2006; Marten, 2010). Chichewa, however, is observed to exhibit the typical PB system, in terms of the noun and agreement marking.

Stellenbosch University https://scholar.sun.ac.za

1995; Diercks, 2011; Zeller, 2017), there are four opposing views to this popular position: the first is that expressed in Trithart (1977) and Givón (1972: 28, 35), who argue that LOCPs may in some cases be nominal and in other instances manifest as PPs. The second view is that of Caha & Pantcheva (2015: 35), who conclude that locative prefixes in Bantu do not belong to any of the familiar category labels. A similar view is also espoused by Taylor (2007) in the analysis of Zulu locatives. The third view, as in Thwala (2006) and Marten (2010), takes LOCPs to be PPs in some Eastern Bantu languages and in siSwati, respectively. The fourth view is that of Ngoboka (2017), who argues that locatives in Kinyarwanda are determiners (Ds) and not noun prefixes or prepositions, although they have the semantics of prepositions. Having shown in Chapters 3, 4, and 5 that the Chichewa noun class system is not structured as assumed in these traditional analyses, I will not review all these contrasting views, but instead I first show the unique properties of Chichewa locatives. Thereafter, it will be easier to determine whether the earlier proposals can offer a proper account of these expressions.

Given the foregoing overview, it is worth noting that the categorial status of Bantu LOCPs is not a clear-cut issue. However, the categorial ambiguity of these 'adposition-like expressions' is also not a Bantu-specific phenomenon. In Chomsky's (1970) *Remarks on Nominalisation*, the preposition is designated as [-N,-V], which implies that the P shares the property of [-V] with the noun. For example, they can both be clefted (cf. Rauh, 2010: 93–96; Panagiotidis, 2015: 14). Svenonius (2006) also presents a crosslinguistic perspective on this phenomenon where similar expressions do not fit neatly into the nominal or the adposition category (cf. Kayne, 2004). Interestingly, even those who argue for the nominal analysis of Bantu LOCPs implicitly or explicitly cite examples that show that Bantu LOCs differ significantly from evaluative and other noun-related affixes (see i.a. Bresnan & Mchombo, 1995: 195–208; Carstens, 1997: 366–368; Salzmann, 2004: 15–17).

What appears to be generally accepted is that Bantu LOCPs are relatively complex expressions composed of the LOC and an NP, an adverbial stem or an adjectival stem. Although I have so far considered the properties of the noun-related affixes (Chapters 4 and 5) and evaluative affixes (Section 6.3), the properties of LOCs have not been properly examined. This forms the objective for the next section.

Stellenbosch University https://scholar.sun.ac.za

6.4.2 Properties of LOCs

The major objective in this section is to evaluate the properties of LOCs in Chichewa. Specifically, I consider whether the properties of LOCs are similar to those of noun-related prefixes such as derivational, number and evaluative markers. The evaluation will centre on a set of four diagnostic tests, namely (i) the ability to take a complex nominal complement test, (ii) the inbound anaphoric test (Bresnan & Mchombo, 1995: 195–208; Salzmann, 2004: 15–17), (iii) the conjoinability test and (iv) the substitution test (Fromkin, 2000). In all these tests, I will show that LOCs do not share morphological properties with noun related prefixes.

6.4.2.1 LOCs take complex NP and PP complements

Unlike prefixes associated with nominal expressions, LOCs in Chichewa and other Bantu languages have been observed to take complex complements. There are four types of complex expressions that are observed to serve as complements of LOCs, namely (i) conjoined NPs, (ii) NPs with demonstrative and associative markers, (iii) NPs with evaluative morphology and (iv) non-locative PPs. Each of these structures is discussed below.

First, LOCs in Chichewa are observed to take scope over conjoined elements, as shown in (35a). Although arguing for the NP view of LOCPs, Bresnan & Mchombo (1995: 205–206) discuss the examples in (35) (see also Carstens, 1997: 365–366; van der Spuy, 2006: 322). All LOCPs in (35) and subsequent examples will be enclosed in rectangular brackets.

(35) Scope over *conjoined elements test* (Bresnan & Mchombo, 1995: 205–206)

a. Mu-ku-pit-a [ku {m-sika kapena m-zinda}]?
SM-T/A-go-FV LOC 3-market or 3-city
"Are you going to the market or the city?"

b. *A-na-b-a ka- {m-pando kapena m-tondo}?
SM-T/A-steal-FV DIM 3-chair or 3-mortar
Intended: "Did s/he steal a little chair or a little mortar?"

In (35a), the LOC *ku-* is shown taking scope over the conjoined nouns *msika* and *mzinda*, in contrast to the DIM prefix *ka-* in (35b) (Bresnan & Mchombo, 1995: 205–206).

In addition to taking conjoined complements, Marten (2010: 258) also shows that in siSwati, LOCs, unlike other noun-related prefixes, behave like prepositions by taking complex nominal complements - for example, ones with a demonstrative prefix, as in (36).

Stellenbosch University https://scholar.sun.ac.za

(36) ku-le-si-kolwa
LOC-DEM-7-school
“at the school”

In addition to selecting conjoined elements, locative markers in Chichewa can also select non-locative PPs and adverbs. Consider the examples in (37).

(37) a. Sing’anga a-na-zama [mu/pa {PP za ufiti}].
traditional doctor SM-T/A-deep LOC about witchcraft
“The traditional doctor is steeped in witchraft.”

b. Mtsogoleri wa-yankhula-po [pa {PP za uchigawenga}].
leader SM.T/A-speak-CLITIC LOC about terrorism
“The leader has commented on issues about terrorism.”

c. Zofufuza zi-ku-loza [ku {PPza katangale}].
research SM-T/A-point LOC about corruption
“Findings are pointing towards corrupt practices.”

The LOCPS in (37) are complex expressions made up of at least two elements, namely (i) PPs, in curly brackets and (ii) LOCS. None of the noun-related derivational or inflectional affixes discussed in the preceding chapters is observed to affix to PPs to derive a noun. In addition to that, as I will show in (44) below, the LOCP complements in (37a-b) cannot be substituted for by a nominal expression but only by another PP.

Further to taking complex NPs, PPs and ADVs, LOCs can also select nouns with evaluative morphology such as AUG, DIM, HON, etc. Consider the examples in (38).[15]

(38) a. mu ka-nyumba
LOC DIM-house
“in the small house”

b. pa chi-mwala
LOC AUG-rock
“on the big stone”

c. pa a-ka
LOC AM-DIM.DEM
“on this small thing”

If the structures in (38) are to be analysed as NPs, then there is a need to revise my earlier analysis of evaluative expressions in Section 6.3.5, where I hypothesised that evaluative affixes occur towards the edge of the nominal spine in Chichewa. Therefore, if LOCs form part of the nominal structure, it has the following three implications: (i) LOCs are either speaker-oriented themselves, or (ii) evaluative affixes (in both their denotative and connotative use) are not edge-related in Chichewa, or (iii) LOCs do indeed not form part of the nominal spine. The present section intends to show that the latter implication holds for

[15] This insightful observation was first pointed out to me by Eunice Mumba, who was at the time one of the 2018 final-year undergraduate students at Chancellor College, University of Malawi.

Stellenbosch University https://scholar.sun.ac.za

Chichewa – i.e. LOCs are independent of the NP spine just as observed by many others (see i.a. Bresnan & Mchombo, 1995: 195–208; Carstens, 1997: 368; Salzmann, 2004: 15–17).

Summing up this subsection, in all these examples (35-38), the LOCs are shown to take some complex phrasal complements or to attach to non-nominal stems, just like the non-locative prepositions in Chichewa such as *za /ya* ("of/about") and *ndi* ("with") in (39).

(39) a. A-na-lemba nthano {PP zá [akalulu osati agalu]}.
SM-T/A-write folktale about rabbits not dogs
"S/he wrote folktales about rabbits not dogs."

b. Wa-ona nyumba {PP yá [mfumu ndi mphunzitsi]}.
SM.T/A-see house of chief and teacher
"S/he has seen the house of the chief and the teacher."

c. Ti-na-guza chigodi {PPndí [manja osati zingwe]}.
1ST PRS.PL-T/A-drag log with hands not ropes
"We dragged the log with bare hands not ropes."

In all three of the cases where LOCs select complex phrasal expressions, the LOCs thus appear to share morpho-syntactic properties with prepositions and not with noun-related prefixes (cf. Bresnan & Mchombo, 1995: 205).

In the next section I consider further evidence that shows that LOCs are distinct from nominal prefixes.

6.4.2.2 LOCs pass the inbound anaphoric island test

Another closely related exceptional property of LOCs is revealed by what is referred to as the *inbound anaphoric island test* (Bresnan & Mchombo, 1995: 189–192). The inbound anaphoric island constraint states that phrases can contain anaphoric and deictic uses of syntactically independent pronouns whereas derived words and compounds cannot (Postal, 1969; Ward, Sproat & McKoon, 1991; Bresnan & Mchombo, 1995: 189–192; van der Spuy, 2006: 315). According to Bresnan & Mchombo (1995: 201–202), this test shows that LOCs are phrasal while the noun-related prefixes are not, as illustrated in (40a) and (40b), respectively (cf. Salzmann, 2004: 16).[16]

[16] Similary, Carstens (1997: 366) analyses LOCs as independent words, that is, as not bound morphemes.

Stellenbosch University https://scholar.sun.ac.za

(40) Inbound anaphoric island test in Chichewa (Bresnan & Mchombo, 1995: 201)

a.		b.	
mu	iyi	*m-	uyu
LOC	9.this	1.PREFIX	1.this
“in this (e.g. house)”		(intended: “this one”)	

Again, as shown in (40a), the locative marker *mu* accepts the deictic pronoun *iyi* – i.e. passes the inbound anaphoric test, but the NC1 prefix in (40b) and all other noun “class markers fail this test” (Bresnan & Mchombo, 1995: 201). The implication of this result is that LOCs are phrases and also different from nominal prefixes. Next, I consider yet another property that further supports the conclusion that LOCs are different from nominal prefixes in Chichewa.

6.4.2.3 LOCs of the same noun can be conjoined

Locative markers are also observed to behave independently from the noun they select as they can conjoin before the noun (Bresnan & Mchombo, 1995: 206; Carstens, 1997: 366). Consider the example in (41).

(41) Conjoined LOCs (Bresnan & Mchombo, 1995: 206)

Ndi-na-jambula	zithunzi	[(mu	ndi	pa)	madengu].
1ST PRS SG-T/A-draw	picture	18	and	16	baskets

“I drew pictures in and on baskets.”

However, it should be pointed out that the example in (41) is relatively odd; importantly, though, it is better than a similar structure involving the typical noun prefixes, such as that in (42).

(42)

*[ti- ndi zi-]	a-galu
[DIM.PL and AUG]	PL-dogs

Intended: “small and big dogs”

The fact that LOCs can conjoin separately, while evaluative morphemes cannot, supports the position taken by Bresnan & Mchombo (1995: 195–208) and Carstens (1997: 368) that LOCs are independent phrases. The question that arises now is as follows: if LOCs are independent phrases, to which category do they belong? Many Bantuists continue to argue that the LOC and the noun still form a complex NP or DP (see i.a. Bresnan & Mchombo, 1995: 208–212; Carstens, 1997: 381–393; Ngoboka, 2017). For example, Ngoboka (2017: 293) argues that although LOCs “have semantic properties of prepositions, syntactically they are determiners on a par with augments and demonstratives.” However, I have not come across any account that explains how the union of a LOC and an adverb/PP obtains an NP or DP. For argument’s sake, if LOCPs are NPs or DPs, I expect them to behave like any

Stellenbosch University https://scholar.sun.ac.za

other NP in Chichewa. I test the distributional properties of LOCPs in the next section (6.4.3). Before that, however, I provide an interim summary of the unique properties of LOCs.

To sum up this section, I have so far shown that LOCs differ from nominal prefixes in at least the following respects: (i) the LOC can take complex phrasal complements or non-nominal complements such as PPs and Advs, (ii) the LOC may pass the anaphoric island test and (iii) LOCs associated with the same noun may conjoin. In addition to the tests discussed above, all Bresnan & Mchombo's (1995: 195–208) other tests show that LOCs differ from the regular nominal prefixes in the sense that LOCs appear to be morphologically free when attached to NPs (cf. Carstens, 1997: 366; Zeller, 2017: 12). These results have consistently shown that LOCs have some unique properties not shared with the regular noun prefix, but with typical prepositions in Chichewa. Therefore, the empirical evidence from Chichewa appears to favour the PP view over the NP and DP views.

Next, I consider the syntactic properties of LOCPs.

6.4.3 Syntactic properties of LOCPs

I have shown that LOCs exhibit features that are uncharacteristic of all noun-related prefixes. Although the PP view is compatible with the empirical evidence considered in Section 6.4.2, by far the most generally accepted view in Bantu literature remains the nominal one. If indeed LOCPs are NPs or DPs as is traditionally assumed, I expect LOCPs to show at least some distributional patterns of typical NPs in Chichewa; in the absence of such patterns, it is not clear how a child acquirer would learn that LOCS are in fact nominal, and not prepositional in nature. To determine this, I consider three syntactic properties of LOCPs. First, I run two classic constituency tests, namely the *coordination* and *substitution* tests (Fromkin, 2000), and then I compare the properties of object NPs and object LOCPs.

First to be considered is the coordination test. The idea behind the coordination test is that expressions of the same category can be conjoined using conjunction morphemes while those belonging to different syntactic categories cannot (Fromkin, 2000: 160–162). Our expectation is thus that if LOCPs are nominal, they should accept coordination with any other regular NP. Consider examples in (43).

(43) a. [LOCP ku Lilongwe] ndi [LOCP ku phiri]
LOC PLACE NAME and LOC mountain
"to/in Lilongwe and to/at the mountain"

Stellenbosch University https://scholar.sun.ac.za

b. *[LOCP ku mtsinje] ndi [NP phiri]
LOC river and mountain

c. *[LOCP pa mwala] ndi [DIM NP ka-mtengo]
LOC stone and DIM-tree

d. [DIM NP ka-mwana] ndi [NP mbuzi]
DIM-child and goat
"The small child and the goat"

In (43a), two locative expressions accept coordination, however in (43b-c) the locative does not pass coordination tests with NPs of various types. The ungrammaticality of the examples (43b-c) can be accounted for as a violation of the *Law of Coordination of Likes*, "which allows that only like constituents may be coordinated" (Fromkin, 2000: 162). Interestingly, the diminutive expression, *kamwana* accepts coordination with the non-diminutive NP, *mbuzi* in (43d), which is also further evidence that evaluative NPs are structurally different from LOCPs, and that they remain nominal in nature.

In addition to failure to coordinate with NPs, LOCPs also fail to pass a simple substitution test with NPs of various kinds. The expectation from this test is that constituents of the same category may be substituted for one another or by their pro-forms without affecting the grammaticality of the expressions (Fromkin, 2000: 155–158, 379–380). Consider examples in (44).

(44) a. Ana a-na-khala pa-mpando
children AM-T/A-sit LOC-chair
"The children sat on the chair."

b. Ana a-na-khala pa-tali/apa/dzulo
children AM-T/A-sit LOC-long/LOC.DEM/yesterday
"The children sat far/ there /yesterday."

c. *Ana a-na-khala mpando
children AM-T/A-sit chair

d. *Ana a-na-khala uwu
children AM-T/A-sit this

As shown in (44a-b), the LOCP *pa mpando* ("on the chair") can only be substituted for by the locative pronoun *apa* ("here"), the adverbs *patali* ("far") or *dzulo* ("yesterday"). However, as shown in (44c-d), the LOCP cannot be replaced by any ordinary NP or non-locative demonstrative pronoun. Another way to account for the ungrammaticality in (44c-d) would be to appeal to the categorial selection properties of the verb *khala* ("sit") in (44): it requires a LOCP, which must be distinct from an ordinary NP, given the ungrammaticality that

Stellenbosch University https://scholar.sun.ac.za

combination with an NP of this type produces. Strikingly, we do not find any cases where the verb's categorial selection requirements distinguish between nouns on the basis of their various prefixes, such as DIM, AUG, HON, etc.

The third unique syntactic feature of LocPs is observed in the distribution properties of *ordinary object NPs* and *object LOCPs*. The canonical VP word order in Chichewa requires that the object NP must immediately follow the verb (see i.a. Bresnan & Mchombo, 1987: 744–745; Mchombo, 2004: 19–20). If the verb is to move out of the VP, the OM corresponding to the noun class of the object NP must be prefixed to the verb stem. However, this rule is not obligatory when we are dealing with LOCPs. Consider the following examples.

(45) a. Mphaka wa-swa mbale.
cat SM-break plate
"The cat has broken a plate."

b. Mbale, wa-**i**-swa mphaka.
plate SM-**OM**-break cat
"The plate, the cat has broken it"

c. ??Mbale **ya**-swa mphaka.
plate SM-break cat
NOT INTENDED: "The plate has broken a cat."

(46) a. Mphaka a-ma-gona pa mphasa.
cat SM-T/A-sleep LOC mat
"The cat sleeps on the mat."

b. Pa mphasa, a-ma-**pa**-gona mphaka.
LOC mat SM-T/A-'OM'[17]-sleep cat
"On the mat, the cat sleeps on it."

c. Pa mphasa, **pa**-ma-gona mphaka.[18]
LOC mat 'SM'-T/A-sleep cat
"On the mat, the cat sleeps on it."

The examples in (45) illustrate the object NP in its default position without the OM (45a) and in reversed position with the OM in place (45b). However, moving the object NP without the

[17] The glosses 'OM' and 'SM', in (46b) and (46c), are used for the purpose of convenience. In Section 6.4.4, I show that these morphemes in fact serve the discourse functions of Topic and Focus and not the grammatical function of subject and object.

[18] Locative inversions of the type in (46) are also allowed in transitives as in (i)

(i) Pa-mpando pa-ma-lemba mwana
LOC-chair LOC-T/A-write child
"On the chair, the child was writing on it."

Stellenbosch University https://scholar.sun.ac.za

OM and letting the object control the SM in (45c) yields a totally different sentence all together. This is not the case in (46) where the LOCP can be moved freely with or without the OM in the verb morphology and still give the same interpretation. What the examples in (45) and (46) show is that object NP and object LOCP exhibit distinctively different agreement properties: i.e. object NP requires obligatory OM while object LOCP does not (see also (56) below).

The failure of LOCPs to pass coordination and substitution tests with NPs and their being unconstrained by the OM obligatory rule are further evidence that LOCPs and NPs do not share categorial properties. However, the other often cited piece of evidence for a nominal analysis of LOCPs in Chichewa is based on grammatical functions (see i.a. Bresnan, 1991, 1994; Bresnan & Mchombo, 1995; Caha & Pantcheva, 2015). Specifically, the argument is that LOCPs are NPs because they occupy subject and object positions, where they also control subject and object verb agreement, respectively (see i.a. Bresnan, 1991; Carstens, 1994; Bresnan & Mchombo, 1995). Consider the following Chichewa examples from Bresnan (1991: 58).

(47) a. [**Ku** San Jose] **ku**-ma-ndi-sangalatsa.
LOC NAME AM-T/A-OM-please
"It pleases me in San Jose, (being in) San Jose pleases me."

b. Ndi-ma-konda [ku San Jose].
SM-T/A-love LOC NAME
"I like it in San Jose."

c. Ndi-ma-**ku**-konda [**ku** San Jose].
SM-T/A-OM-love LOC NAME
"I like it in San Jose."

d. Ndi-ma-sangalats-idw-a ndi [ku San Jose].
SM-T/A-please-PASS-FV by LOC NAME
"I'm pleased by (being in) San Jose."

The examples in (47) are argued to illustrate that Chichewa LOCPs can be in subject position (47a), object position (47b-c) and that they can also function as the object of the preposition *ndi* ("by") (47d). Bresnan (1991: 58–59) also claims that the locative objects illustrated in (47b-c) pass Hyman & Duranti's (1982) three classical locative object tests for Bantu, namely object marking on the verb, passivation and word order. However, as shown in (44), object NP and object LocP do not follow the same word order restrictions in Chichewa.

Stellenbosch University https://scholar.sun.ac.za

Bresnan (1991: 53) states that the constructions in (47) are not grammatical in English where LOCPs are actually PPs that appear in non-subject and non-object positions, as shown in (48).

(48) a. [(*In) San Jose] makes me happy.
b. I like [(*in) San Jose].

Following Thwala (2006), I argue that although the grammatical functions diagnostic has been a reliable tool for languages similar to the one it was initially designed for, English, it may not yield similar results when applied to typologically very different languages, such as Chichewa. In the next section, I evaluate the merits of the grammatical function diagnostic.

6.4.4 Grammatical and discourse functions of LOCPs

As is apparent in the works of Bresnan (1991, 1994, 1995) and others, the syntactic arguments for the NP analysis of LOCPs are largely expressed with reference to the structure of English. However, it should be pointed out that Chichewa and Bantu languages in general are typologically different from Indo-European languages such as English. There are three key differences that are relevant to how one analyses the grammatical properties of these languages. Firstly, Bantu languages and English differ in the way they encode relationships between words in a phrase, with many Bantu languages being analysed as head-marking and English as a dependent-marking language (Nichols, 1986: 56–57, 1992: 46–64; Bresnan & Mchombo, 1995: 743).[19] Secondly, Bantu languages and English also differ in terms of the way in which they encode the grammatical and discourse functions of subject and topic, as outlined in Li & Thompson (1976: 459), Morimoto (2006), and van der Wal (2015), among others. Specifically, it is observed that English is a subject-prominent language whereas many Bantu languages are observed to be topic-prominent (Morimoto, 2006; van der Wal, 2015). Thirdly, where the grammatical functions of subject and object

[19] The head vs. dependent-marking distinction arises from cross-linguistic observations which have picked up on the fact that in some languages syntactic relations are morphologically marked on the head of a constituent - the so-called *head-marking* languages - while in other languages the relationships are marked on the dependent constituent - the so-called *dependent-marking* languages (Nichols, 1986: 56–57, 1992: 46–64). A head, in this case is a constituent that determines the categorial status of a phrase. For example, the verb is the head of a verb phrase whereas the dependent can be the complement or modifier of the head word.

Stellenbosch University https://scholar.sun.ac.za

are relevant, these languages also differ in how the subject and object grammatical functions are encoded.[20]

For the purposes of the present discussion, the grammatical and discourse functions difference is the most relevant and it is therefore the one that I will pursue in some detail.

6.4.4.1 Background to subject-object and topic-focus functions in Bantu

Although the discussion about the categorial status of Bantu LOCPs has generally been made with a strong appeal to the notions of subject and object, it has also been argued that some Bantu languages are topic-prominent, with the consequence that the grammatical functions of subject and object may not have the same structural bearing as they do in English (see i.a. Givón, 1976; Li & Thompson, 1976: 476; Morimoto, 2006; van der Wal, 2015). This distinction originates from observed variations in how different languages appear to modify the linear structure of the sentence in response to linguistic and extra-linguistic contexts – the phenomenon that is currently often designated *information structure* (Lambrecht, 1994; Erteschik-Shir, 2007; van der Wal, 2015). According to van der Wal (2015: 3), information structure relates to "the packaging of information to facilitate the hearer's processing of the information". This usually involves categorising information as *old* (given) or *new* to the conversation. On the one hand, 'old information' is taken to be active in the hearer's mind and can easily be referred to by the speaker. 'New information', on the other hand, needs to be activated in the hearer's mind (Lambrecht, 1994: 46; Erteschik-Shir, 2007: 7; van der Wal, 2015: 3). Old information is designated as the *topic* of the sentence and the rest of the sentence adds information about this topic and is hence referred to as the *comment* (Bresnan & Mchombo, 1987: 746; van der Wal, 2015: 3). Consider the following English examples.

---

[20] Some works also recognise the differences in terms of how the subject and object functions are encoded. There are two types of languages in this regard, namely *configurational* and *non-configurational* languages (Hale, 1981, 1982, 1983; Bresnan & Mchombo, 1987: 742; Baker, 2001; Salzmann, 2004: 179–182). In configurational languages, the subject and object functions are "distinctively encoded by phrase structure", whereas in non-configurational languages, these functions are "not distinctively encoded by phrase structure" (Bresnan & Mchombo, 1987: 742). For example, English is a configurational language where the grammatical functions of subject and object consistently appear in specific phrase structural positions (Baker, 2001: 407). Although Bresnan & Mchombo (1987: 743) state that Chichewa is a configurational language, it can be argued that Chichewa is more of a non-configurational language as per the criteria outlined in Hale (1981, 1982, 1983) and Jelinek (1984).

Stellenbosch University https://scholar.sun.ac.za

(49) The subject- and topic-prominence distinction in English (Li & Thompson, 1976: 459)

a. John hit Mary.
SUBJECT PREDICATE

b. As for education, John prefers Bertrand Russell's ideas.
TOPIC COMMENT

In (49a) the NP *John* is the sentence-initial subject. However, in (49b) the sentence begins with the topic, *as for education.* Example (49b) does not reflect the canonical structure of English sentences, but in some languages, this is the standard way of structuring sentences. For instance, consider the following examples from a Lolo-Burmese language called Lisu (Li & Thompson, 1976: 461).

(50) Topic difference in Lisu (Li & Thompson, 1976: 472)

a. làthyu nya ánà khù-a
people TOPIC MARKER dog bite-DECLARATIVE MARKER
"People (TOPIC) { they bite dogs." / dogs bite them." }

b. ánà nya làthyu khù-a
dog TOPIC MARKER people bite-DECLARATIVE MARKER
"Dogs (TOPIC) { they bite people." / people bite them." }

Take note that the sentence-initial NPs in (50a) and (50b) only differ in terms of expressing the sentence topic and not the sentence's agentive subject, as is the case in English (Li & Thompson, 1976: 472). In this language, the grammar does not encode the grammatical functions of subject, leaving the ambiguity over agency to be resolved via other means – context, for instance. Here, then, the sentence-initial positions are associated with topic information. On the basis of evidence of this nature, Li & Thompson (1976: 460) argue that "some languages can be more insightfully described by taking the concept of *topic* to be basic, while others can be more insightfully described by taking the notion of *subject* as basic" (cf. Morimoto, 2006: 166).

In addition to coding sentence topics, information structure also involves the highlighting of 'new information' or information that the speaker wants to come back to, or contrast with other information. This is designated as the *focus* of the sentence (Bresnan & Mchombo, 1987: 746; van der Wal, 2015: 3). The dedicated focus position in Bantu has been hypothesised to be either the so-called *Immediate After the Verb* (IAV) position or the sentence-final position (see van der Wal, 2015: 15). Consider the Kinyarwanda examples in (51).

Stellenbosch University https://scholar.sun.ac.za

(51) Topic and Focus marking in Kinyarwanda (Morimoto, 2006: 163; Salzmann, 2011: 10)

a. Umuhuûngu **a**-ra-som-a igitabo. SVO
1boy 1-t/a-read-fv 7book
"The boy is reading the book."

b. Igitabo **ki**-soma umuhuûngu. OVS
7book 7-read 1boy
"The boy (FOC) is reading the book (TOP)."

Take note of the reversed word orders in (51a) and (51b). Although *word order reversal* is a common phenomenon in Bantu languages, it is not common for the sentence-initial logical object (e.g. *igitabo*) to agree with the SM, as shown in (51b) (Morimoto, 2006: 163). If agreement is taken to be subject-oriented, then the sentence in (51b) would be interpreted as the book is the agent; *it is the book that is reading the boy*. However, as shown in the idiomatic glossing, this is not the relevant interpretation; rather, the noun *igitabo* in the preverbal position encodes the discourse function of topic while the noun *umuhuûngu* is the focus of the sentence (Morimoto, 2006: 164). Therefore, it should be plausible to associate the sentence-initial position with topic elements and the post-verbal position with focus elements in Kinyarwanda and related languages.

Just to anticipate the important problems under consideration in the present section, the grammatical process taking place in (51), the so-called *subject-object (S-O) reversal,* is very common in Chichewa. However, the postposed logical subject must continue to control subject agreement on the verb. Consider the Chichewa version of (51) in (52).

(52) a. Mnyamata **a**-ku-werenga buku.
boy SM-T/A-read book
"The boy is reading a book." (Lit.: "The boy, he is reading the book.")

b. Buku **a**-ku-werenga mnyamata.
book SM-T/A-read boy
"The boy is reading the book." (Lit.: "The book, the boy is reading.")

c. ??Buku **li**-ku-werenga mnyamata.
book SM-T/A-read boy
??"The book, it is reading the boy."

d. Buku, **lií**-kú-wérenga mnyamata.
book SM-T/A-read boy
"The book, which is being read by the boy."

Note that in Chichewa, when the logical object controls subject-verb agreement as in (52c), the outcome is the distinctively odd interpretation. The only felicitous related structure is (52d), where *buku* is not a subject, but the head of a relative clause as evidenced by high

Stellenbosch University https://scholar.sun.ac.za

tone on the SM and the two following syllables.[21] However, as also shown in (46), the fronted LOCP controls SM agreement in exactly the same way as the Kinyarwandan topic agreement in (51b). Consider the following example:

(53) a. Mnyamata wa-khala [pa mpando].
boy SM.T/A-sit on chair
"The boy has sat on the chair." (Lit.: "The boy, he has sat on the chair.")

b. [Pa mpando], wa-khala mnyamata.
on chair SM.T/A-sit boy
"On the chair, the boy has sat."

c. [**Pa** mpando] **pa**-khala mnyamata.
on chair AM.T/A-sit boy
"On the chair, the boy has sat on it." (Lit.: "On the chair, on it the boy has sat.")
"On the chair sat a boy."

d. [**Pa** mpando], **páá**-khálá mnyamata.
on chair AM.T/A-sit boy
"On the chair, where the boy has sat."

The structure in (53c) was ungrammatical in (52c), which shows that, unlike fronted object NPs, the fronted LOCPs are allowed to control the *subject agreement* on the verb. The LOCP can also control SM agreement on the relative clause construction in (53d). Note, however, that structures like those in (53c) usually have two potential readings: one reading is where the LOCP is given the grammatical function of subject, with the odd interpretation as if the LOCP is the agent and the other one is where the LOCP is interpreted as the topic of the sentence ("On the chair, the boy has sat on it."). The traditional literature does not point out this possibility as the glossing convention is to simply capture the felicitous interpretation. Although Bresnan & Mchombo (1987: 745) also observe that the SM involving ordinary NPs ambiguously marks topic and subject in Chichewa, they however claim that LOCPs only function as subjects and not as topics. However, the interpretation of sentences such as that in (51c) clearly shows that the sentence-initial position ambiguously encodes subject and topic functions even in locative inversion constructions (cf. Schachter, 1992: 109). Nevertheless, the debate on this matter has produced diverging views.

Next, I provide a brief outline of some diagnostics that show that Chichewa grammaticalizes the topic and focus functions of subject NPs and LOCPs.

[21] High tone marking is a well-known strategy for introducing relative clauses in Chichewa and other Bantu languages (see i.a. Mchombo, 2004: 42; Simango, 2006: 279; Downing & Mtenje, 2011: 74–75).

Stellenbosch University https://scholar.sun.ac.za

6.4.4.2 The topic and focus functions debate in Chichewa

Some studies allege that Bantu languages are evolving from topic-prominent to subject-prominent systems, claiming that what we refer to as *subject agreement* was originally topic agreement in PB (Givón, 1976: 149; Morimoto, 2006: 161, 177). As regards their synchronic status, there are claims that some Bantu languages are topic-prominent while others are subject-prominent (see i.a. Bresnan & Mchombo, 1987; Bresnan & Kanerva, 1992; Bresnan, 1994; Salzmann, 2004: 296; Morimoto, 2006). Crucially, Chichewa is analysed as belonging to the latter type (see Bresnan & Mchombo, 1987).

Following Bresnan & Mchombo (1987), many studies have treated Chichewa as different from other Bantu languages in the sense that Chichewa, like English, is a subject-prominent language (see i.a. Bresnan & Kanerva, 1992; Bresnan, 1994; Salzmann, 2004: 296; Morimoto, 2006). One diagnostic proposed for testing this claim is what is referred to as *subject with in situ question* (Bresnan & Mchombo, 1987: 759–760). Bresnan & Mchombo argue that since wh-elements are taken to be inherently focused because they ask for new information, topic-prominent languages should not allow subjects to be questioned in situ. The argument is that questioning the subject in situ would entail a focus element being located in a position (the clause-initial one) reserved for topical elements. Indeed, the way to ask *wh*-subject questions in Chichewa is via a cleft, which circumvents the problem of having the focus element in the topic position. Consider (54).

(54) a. Mwana wa-tola mpira.
child AM-pick ball
"The child has picked up the ball."

b. Ndi ndani amene wá-tóla mpira?
COP who AM-REL AM-pick ball
"Who has picked up the ball?"

c. N'ndani wátóla mpira?
COP.who AM-pick ball
"Who has picked up the ball?"

d. ??Ndani watola mpira?
COP.who AM-pick ball

The questions in (54b) and (54c) are based on the declarative sentence in (54a), where the subject *mwana* is also the topic. Although Chichewa has an in situ question rule, in object questions for example, the subject with in situ question in (54d) is not successful as the expression is interpreted as if *Ndani* is a name of a person. If interpreted as a question, it

Stellenbosch University https://scholar.sun.ac.za

implies that one has read it as a contraction similar to (54c). Bresnan & Mchombo (1987: 762) argue that clefting (like that shown in (54)) is a strategy that splits the focus and topic functions into two different clauses where the question word (*-ani*) is the focus and the AM- of the relative pronoun is the topic. However, they (1987: 759–760, 778) do not analyse constructions like that in (54d) as clefts contractions but as a typical grammatical subject with in situ question. Since clefting is the only incontrovertible way for asking questions of this type in Chichewa, this means that clefting is a strategy to avoid the topic-focus functions clash. In this regard, it can be concluded that the SM also marks topic and not just subject agreement, as Bresnan & Mchombo (1987) argue.

Against this backdrop, I suggest that a more useful way of understanding information structure in Chichewa would be to adopt proposals regarding topic agreement made for Dzamba, Kirundi, Kinyarwanda and many other Bantu languages discussed in Morimoto (2006) and van der Wal (2015). Research focusing on these languages has generally observed that the so-called *subject agreement* is in fact determined by *preverbal elements* and not strictly by subject elements (see i.a. Collins, 2004; Carstens, 2005; Baker, 2008; van der Wal, 2015). Therefore, since the elements that may surface in preverbal positions can be determined by their topicality (see van der Wal, 2015: 7), it should be plausible to argue that what we have in these Bantu languages is also topic and focus agreement and not strictly subject and object agreement (cf. Bokamba, 1979; Kinyalolo, 1986; Bresnan, 1994: 85–87; Morimoto, 2006, 2008).

6.4.4.3 LocPs as topics in Chichewa

Having argued that Chichewa exhibits some indications of topic grammaticalization, I propose the following to be the case: in Chichewa, as in other Bantu languages, the sentence-initial constituent can serve as both the subject and also the topic of the sentence. However, in some circumstances, the subject reading may be prominent and, in others, the prominent reading will be the topic reading. In locative inversion constructions, it is the case that the subject function is not encoded by the grammar, which gives rise to the topic function only. Consider the locative inversion example below:

(55) a. Ana a-khala pa bedi.
children SM.T/A-sit LOC bed
"Children are sitting on the bed."

Stellenbosch University https://scholar.sun.ac.za

b. Pa bedi pa-khala ana.
LOC bed SM.T/A-sit children
"On the bed, on it there are sitting children."

In (55a), the NP *ana* is serving the function of thematic subject and at the same time it may be the topic of the sentence, i.e. it is given that the sentence is about children. In (55b), it is still understood that the logical thematic subject is the postposed NP *ana*. However, the LocP *pa bedi* is what is given as old information, thus its topical position. Since inverted objects do not trigger SM agreement it appears that the only topics allowed to control topic agreement are subject NPs and LocPs. As shown in (46) above, one reason inverted objects do not trigger SM agreement is because Chichewa enforces a rule that if an object is to be moved from within the VP, it must leave behind an OM in the VP (see i.a. Bresnan & Mchombo, 1987: 744–745; Mchombo, 2004: 19–20). For this reason, the object NP cannot at the same time control another AM within the same phrase as shown in (56)

(56) a. *Chilombo$_k$, cha$_k$-chi$_k$-menya ana.
beast SM.T/A-OM-hit children
INTENDED: "Children, they have hit the beast."

As argued above, the fact that the obligatory OM for moved objects does not apply to LocPs is another piece of evidence to show that the LocP is not within the VP or is not subject to NP object syntactic constraints.

Since most default topics (subjects and LocPs) are grammaticalized in Chichewa, the LocP controls agreement on the verb. What is ultimately at issue here, as Erteschik-Shir (2007: 1) states, is whether one wants to account for this phenomenon in purely syntactic terms or whether one assumes a pragmatic explanation or mixed views. Nevertheless, it appears plausible to accept that the linear order of constituents is to some extent "determined by notions having to do with what is contextually known or what is not" (Erteschik-Shir, 2007: 1).

I therefore conclude that in Chichewa, LocPs can occur in sentence-initial positions as topics. If they are topics, the argument that LocPs must be nominal because of their subject function automatically falls away. Now, considering the evidence presented in Section 6.4.2 about Locs and the various ways in which their behaviour resembles that of prepositions, I further conclude that LocPs in Chichewa are locative PPs (cf. Givón, 1972: 12). There is no syntactic constraint on the category of topics, which can be PPs, or CPs, or something else.

Stellenbosch University https://scholar.sun.ac.za

With this in place, I can now easily account for the distribution patterns of CPs and other complex agreement triggers which are clearly non-nominal, which also trigger their own AMs. A brief overview of PPs and CPs in similar positions is provided next.

6.5 Other clearly non-nominal expressions that trigger topic agreement

In the present section, I consider other non-nominal expressions that trigger the AMs *zi-, ku-,* and *pa-*. Note that the AM *zi-* has already been considered in our discussion of evaluative expressions in Section 6.3.2 and the AMs *ku-,* and *pa-* in the foregoing discussion about LocPs. In the present section, these AMs are mainly observed to be triggered by complex expressions and underspecified agreement triggers. First to be considered will be non-locative PPs.

6.5.1 Agreement involving non-locative PPs

Just like LocPs, non-locative PPs are also observed to trigger alliterative topic agreement. Consider the examples in (57).

(57) a. [PP **Zá** ngozi] **za**-ti-khudza.
PREP.ASC. accident AM-OM-touch
"We have been concerned about the accident."
(Lit.: "About the accident, it has concerned us.")

b. **Za**-ti-khudza [PP **zá** ngozi].
AM-OM-touch PREP.ASC. accident
"We have been concerned about the accident."

The expression *za ngozi* ("about the accident") in (57a) is undoubtedly a PP. This PP is observed to trigger obligatory agreement on the verb.[22] Just like LocPs, I argue that this agreement is related to topic encoding and not subject agreement. To my knowledge, I have not come across any arguments in favour of treating PPs such as that in (57) as NPs. Therefore, if the non-locative PP controls agreement, it would be unconvincing to argue for a different status for LocPs. The same pattern is also observed with CPs, to be discussed next.

[22] The observation that non-locative PPs control agreement is counter Bresnan (1994: 11 ff) and Bresnan & Mchombo (Bresnan & Mchombo, 1995: 211), who argue that PPs do not control agreement in Chichewa. Unfortunately, their argument is based on only one type of preposition in Chichewa, namely *ndi* ("with").

Stellenbosch University https://scholar.sun.ac.za

6.5.2 Agreement involving CPs

The fact that CPs trigger agreement and that they occupy the so-called subject and object positions has been observed in the previous literature (cf. Bresnan, 1995: 34ff). In Chichewa, CPs are introduced by *kuti* ("that").[23] As such, Bresnan (1995: 34) designates them *kuti clauses*. They are marked in the curly brackets in (58).[24]

(58) a. A-na-nen-a {CPkuti a-dani a-ku-bwera}.
SM-T/A-say-FV that PL.enemy SM-T/A-come
"S/he said {that enemies are coming}."

b. Zi-ku-onek-a {CPkuti ku-gwa mvula}.
SM-T/A-seem-FV that SM.T/A-fall rain
"It seems {that it is going to rain}."

c. Ku-ku-onek-a {CPkuti ku-gwa mvula}.
SM-T/A-seem-FV that SM.T/A-fall rain
"It seems {that it is going to rain}."

d. {CPKuti ku-gwa mvula} ku-ku-onek-a.
that SM.T/A-fall rain SM-T/A-seem-FV
"It appears {that it is going to rain}."
Lit.: "That it is going to rain, it shows."

According to Bresnan (1995: 34ff), CPs are also observed to exhibit similar properties to those identified for LOCPs in Chichewa. For example, (i) they may function as prepositional complements, predicate 'subjects' and 'objects', (ii) they participate in the so-called 'locative inversion'[25] and (iii) they trigger 'subject and object agreement' with the verb. For a detailed discussion of these syntactic properties, see Bresnan (1995: 34ff).

Although Bresnan (1995: 42–43) associates the Chichewa CP with the so-called *infinitive NC15* because of the homophonous prefix *ku-,* it should also be stated that sometimes the CP may also be headed by the preposition *za-,* which eventually triggers the *zi-* AM. This is shown in (59).

---

[23] Bresnan (1995: 34) describes CPs as "propositional complements which designate reported speech, thoughts and perceptions".

[24] The complementiser *kuti* literally meaning "saying" could have its origins in the verb root *-ti* "say" and the *ku-* could be the infinitive marker or an expletive prefix.

[25] Take note that even though CPs are not locatives, they are analysed as participating in 'locative inversion' in Bresnan (1995: 34ff).

Stellenbosch University https://scholar.sun.ac.za

(59) a. {CP**Zá** kuti mu-na-khoza mayeso} **za-**mv-ek-a kale.
of that SM-T/A-pass exams SM.T/A-hear-STAT-FV past
"That you passed exams, it has already been heard."

b. Anthu a-**zi**-mva {CP**zá**kuti mwa-khoza mayeso}.
people SM.T/A-OM-hear of that SM.T/A-pass exams
"People have heard (it) that you have passed exams."

Even when the associative (ASC) marker is not used in the CP, the default AM *za-* may still be triggered in the verbal agreement.

(60) {CPKuti mu-na-khoza mayeso} **za**-mv-ek-a kale
that SM-T/A-pass exams SM.T/A-hear-STAT-FV already
"That you passed exams, it has already been known."

In (60) the complementiser *kuti* is not controlling alliterative agreement, but the verb takes the default AM *za-*. As introduced in Chapter 4, the AM *zi-* is not always triggered by concordial agreement; rather it is also used as a default AM. The default agreement marking observed with CPs in (60) is also found with other controllers, such as conjoined NPs and underspecified agreement triggers. These are discussed next.

6.5.3 Conjoined NPs and underspecified agreement triggers

As already introduced in Chapter 4, Section 4.9, in Chichewa the AM *zi-* functions as a default AM for conjoined NPs and where the agreement controller features are not specified in the sentence or discourse. Consider the following examples.

(61) a. Njuchi ndi chiwala **zi**-ma-thawa chule
Bees (*i*-ANC) and grasshopper (*chi*-ANC) SM-T/A-run frog
"Bees and a grasshopper were running away from the frog."

b. **Zi**-ku-oneka kuti a-kwiya
SM-T/A-appear that SM.T/A-be angry
"It appears that she/he/them is/are angry."

c. Ngati wa-mangidwa **zi**-dziwika.
If SM.T/A-be arrested SM.T/A- be known
"If s/he is arrested, it will be known."

In (61a), the AM *zi-* is triggered by the conjoined NP with conflicting noun class features whereas, in (61b-c), there is no specific controller and the *zi-* is functioning as the expletive pronoun, which can be taken to be the equivalent of "it" in English. Similar, expletive-like functions of the AMs are also observed with the AMs *ku-* and *pa-* as illustrated in (62-63).

Stellenbosch University https://scholar.sun.ac.za

(62) a. **Ku**-ku-oneka ngati mvula i-chedwa chaka chino.
SM-T/A-appear if rain SM.T/A-be late year this
"It appears as if the rains will be late this year."

b. **Ku**-li-be ku-gona lero.
SM-COP-NEG SM-sleep today
"There is no sleeping today."

(63) **Pa**-libe vuto mu-tha ku-bwera.
SM-COP-NEG problem SM-can INF-come
"There is no problem, you can come."

So, in addition to functioning like locative AMs, *ku-* and *pa-* also function as expletive AMs in Chichewa. The major difference between the default AM *zi-* and the other default AMs *ku-* and *pa-* is that the former is more generally associated with missing NP or conflicting NP agreement controller features while the latter are commonly used as indicating non-specific location as is the case with the expletive *there* in English. Further, the expletive-like AMs *ku-* and *pa-* also convey subtle differences, for example, *ku-* functions to represent more general (unspecific) missing location features while *pa-* stands in for missing locative features that are relatively proximal.

Having considered these examples, it should be clear that AMs are not always controlled by NPs but also by complex phrases, unspecified agreement triggers and they can also take on the function of expletive affixes. Therefore, this is further justification for identifying GACs as being different from the intuitively satisfying ANC system proposed in Chapter 5. In the next section, I propose the Chichewa-specific GAC system.

6.6 The general agreement class system in Chichewa

As shown in the present chapter, specifically Section 6.5 above, there is undisputable evidence that agreement in Chichewa is not only a nominal phenomenon but also affects non-nominal expressions such as CPs and PPs. This agreement could be subject/object- or topic/focus-related. Therefore, the proposed distinction between ANCs and GACs is well motivated. Strikingly, though, what the ANC and the GAC have in common is their organisation around phonological considerations. However, the GAC agreement cue set is not just phonologically determine; it also appears to be morphologically governed, i.e. the morphological status of phrase-initial elements seems to be key. Specifically, these phrase-initial elements are predictable affixes. First, consider the *GAC affix cue set* in Table 6.2.

Stellenbosch University https://scholar.sun.ac.za

Table 6.2: GAC affix cue set

	Prefix cue set	GAC	Expression types
1	***ka-***	***ka-GAC***	(i) Singular DIM ***ka-*,** (ii) ***ka***-initial manner nominals (see Section 4.10)[26]
2	***ti-***	***ti-GAC***	(i) Plural DIM ***ti-*** (ii) some t-initial nouns (*tulo* "sleep", tuvi "faeces")
3	***mu-***	***mu-GAC***	(i) LOC **mu-,** (ii) lexicalised LOC ***mu***-initial nouns
4	***pa-***	***pa-GAC***	(i) LOC **pa-**, (ii) lexicalised LOC ***pa***-initial nouns, (iii) expletive function
5	***ku-***	***ku-GAC***	(i) LOC **ku-**, (ii) ***ku***-initial nominal infinitives, (iii) CP **ku-**, (iv) expletive function

Note that the GAC agreement cue set in Table 6.2, does not include the AUG, HON and ATT NPs because these complex NPs trigger the same AMs as the *chi*-ANC, *a*-ANC and *u*-ANC, respectively. Also recall that the respective prefixes were observed to recycle prefixes from nouns in the relevant ANCs.

Returning to the issue of the phonological and morphological nature of the agreement cue set: there is evidence that the GACs and their AMs are phonologically determined. As shown in Table 6.2, some GACs include more than one type of expression, for example, *ka*-GAC includes DIMs and manner nominals which share the property of the prefix *ka-;* the *ku*-GAC, in turn, includes locatives, infinitives and CPs which share the prefix *ku-*. The morphological status of the LOC *ku-* and infinitive prefix *ku-* are different, which suggests that they only come to trigger similar AMs on phonological grounds, as is the case in ANCs. However, this is not the case with the majority of nouns whose word-initial elements are phonologically similar to the GAC affix cue set. For example, there are several *ka*-initial nouns in the *m(u)-a*-ANC, but they do not trigger the same AMs as expressions found in the *ka*-GAC. This suggests that the criteria behind the GAC subsystem must be sensitive to both the

[26] For example

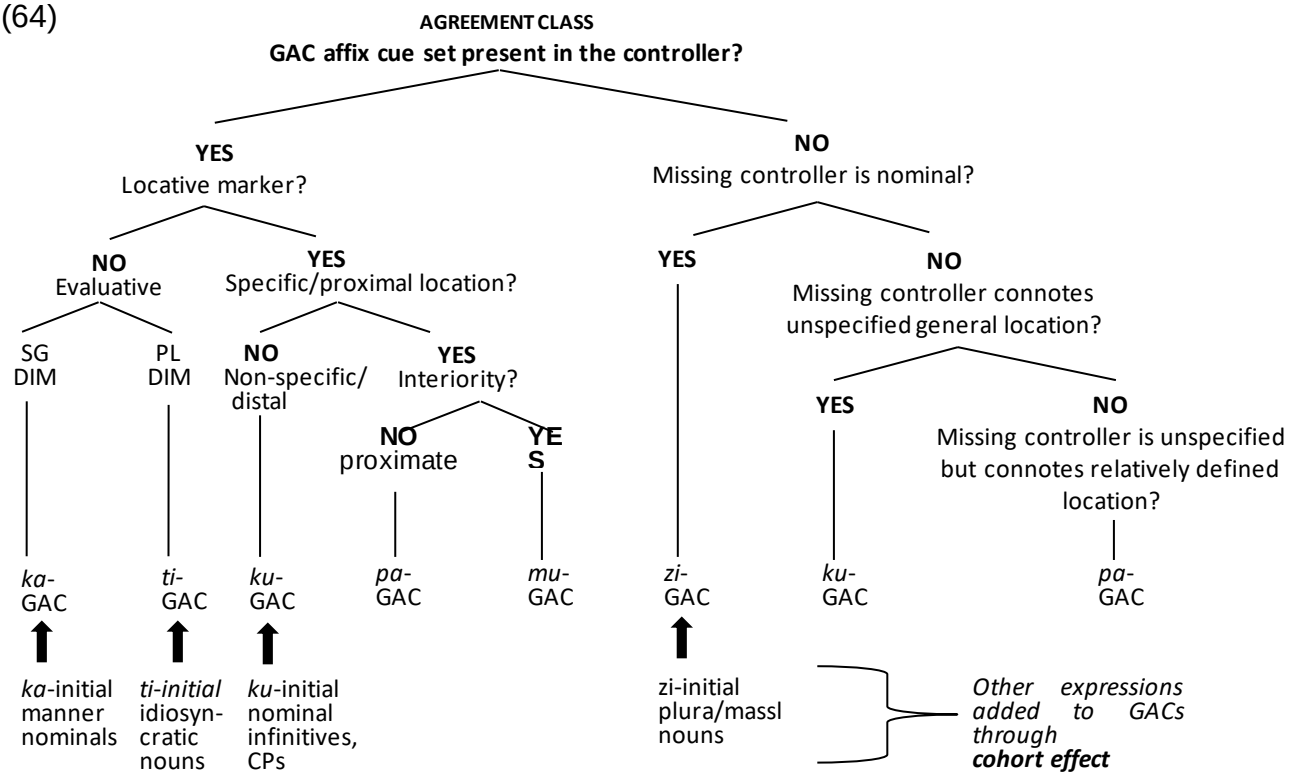
(i) Ka-yend-edw-e ka-tsopano ka-phunzits-idw-e kwa aliyense.
PREFIX-walk-PASS-FV AM-modern AM-teach-PASS-FV to everybody
"The modern manner of walking should be taught to everybody."

Stellenbosch University https://scholar.sun.ac.za

morphological and phonological status of the phrase-initial element. Specifically, the less productive *ka-* of the *m(u)-a-*ANC does not parallel the productive *ka-* prefix for DIM and manner nominals.

Modelling the interplay between morphological and phonological considerations seems to be more elusive compared to the semantic and phonological criteria in the ANC system. I thus propose a highly speculative schema for the GAC morphological and phonological criteria as follows: First, I assume that some expressions making up the various GACs are core to the system while others come to trigger the same AM with the GACs due to phonological identity, so-called *cohort effect* (Sekerina & Brooks, 2007: 21). Therefore, I assume that the morphologically-determined cues are primary and derive the following structure.

(64)



The schema in (64) is modelled on locative and evaluative criteria because they represent one of the highly regular morphological processes outside the derivational processes observed with nouns forming the ANC. Therefore, I do not consider *ka*-initial manner nominal and *ku*-initial nominal infinitives as key in the GAC morphological classification because although the *ka-* and *ku-* are regular they form part of the nominalisation morphology that is characteristic of the ANC system. However, because their word-initial characteristics do not match any of the AMs in the ANC system, the manner and infinitival

Stellenbosch University https://scholar.sun.ac.za

nouns come to trigger the AMs associated with GACs on a phonological basis, i.e. their word-initial syllables are homophonous with the respective GAC AMs. A similar pattern has also been observed with highly regular evaluative expressions, which also trigger AMs associated with ANCs. Therefore, although the GAC system is syntactically based, it attracts nominal expressions on a phonological basis. I suggest that the morpho-phonological interplay arises on the stated grounds.

As shown in (64), where the agreement trigger's initial phoneme matches the GAC cue set, the relevant AM is used on the various agreement targets. Since it is obligatory for dependent words such as adjectives, numerals, and verbs to have an AM, I expect default agreement to arise where there are no relevant features for the AM. In this case, default agreement is a case of underspecification; hence the already existing AMs *zi-*, *ku-* and *pa-* fill the agreement function on the dependent words. Specifically, if the controller is missing, it is evaluated whether the missing expression is nominal in nature or locative. If the missing controller has nominal features, as shown in Chapter 5, the default AM *zi-* is used; otherwise if the missing controller has locative connotations, two different AMs *ku-* and *pa-* may be used, one for more general location and the other for relatively specific location, respectively.

The agreement class system modelled in (64) is descriptively adequate in terms of the type of agreement trigger and their AMs. It is not yet clear how the GAC cue set and the ANC phonological cue set interact. However, what is clear is that the GAC subsystem is cued by morphologically determined affixes. Nevertheless, highly productive nominal elements such as nominal infinitives and manner nominals do not appear to trigger the various AMs due to their derivational prefixes, but due to their accidental phonological identity with the AMs. As stated in Chapter 5, there is a small number of morphologically idiosyncratic nominal expressions whose word-initial phonemes are also homophonous with those of GAC cue set in Table 6.2, such that they also trigger the AM associated with the relevant GACs.

Having stated the above, drawing the distinction between ANC and GAC subsystems helps to understand how different expressions trigger their respective AMs. On the one hand, typical nominal expressions trigger AMs following the semantically driven ANC system with a phonologically determined sub-system. The morphologically predictable expressions (both nominal and non-nominal), on the other hand, select their AMs primarily on GAC morphologically-determined cue features, with the underspecified sub-component also

Stellenbosch University https://scholar.sun.ac.za

following a syntactic category basis, that is, whether the underspecified agreement trigger is nominal or locative.

The next chapter considers the theoretical underpinnings of the ANC and GAC subsystems. However, before that, in the next section, I conclude the present Chapter.

6.7 Conclusion

In the present chapter, I focused on the properties of morphosyntactically predictable expressions that are also observed to trigger agreement in Chichewa. I have analysed these expressions as falling into three subcategories, namely evaluative NPs, locative PPs and other non-nominal phrases such as CPs and non-locative PPs; all of which are observed to trigger AMs on modifiers and predicates. First, I focused on the evaluative NPs and their affixes. I showed that these affixes are multifunctional such that they do not just carry denotative DIM and AUG meanings, for instance, but also other fine-grained speaker-hearer subjective meanings. I argued that the multifunctional nature of these affixes arises from the case of underspecification; that is, there are sets of underspecified affixes which are exploited by evaluative functions such as AUGs (*chi-* and *zi-*), HON (*a-*) and ATT (*u-*). These functional features are analysed as 'recycling' the already existing underspecified affixes within the nominal structure.

I further showed that these evaluative affixes are, in remarkable ways, structurally different from the so-called primary noun prefixes. While the primary prefixes function as number or derivational affixes which attach lower in the nominal spine, evaluative affixes manifest as functional elements that attach higher in the nominal spine. Although the evaluative use of prefixes typifies the productive nature of elements in the GAC, the AUG, ATT and HON expressions do not form distinct GACs as their word-initial characteristics (in the form of prefixes) coincide with the ANC phonological cue set. However, the DIM NPs do not recycle any prefixes from within the proposed ANC subsystem nor do they coincide with any AM within the ANC phonological cue set. As a result, separate GACs are postulated, namely singular DIM *ka-* and plural DIM *ti-*.

In addition to evaluative expressions, I also discussed LocPs. I considered a range of unique properties of Locs and LocPs and concluded that LocPs do not show strong features that support the traditional NP view. I also considered the grammatical and discourse functions of LocPs and argued that, in Chichewa, grammatical and discourse functions can be

Stellenbosch University https://scholar.sun.ac.za

ambiguously encoded. With the exception of object NP arguments, the AMs ambiguously encode the grammatical functions of subject and also the discourse functions of topic. I argued that object NPs fail to mark topic agreement because of the obligatory OM rule for all displaced object NP arguments in Chichewa. I claim that this obligatory OM rule does not apply to LOCPs because they are not of the same category as NP objects. Having shown this, I concluded that LOCPs in sentence-initial positions do not really serve the grammatical function of subject as per the traditional account but rather that they serve the discourse function of topic. The grammatical function arguments for the nominal view of LOCPs are thus undermined; therefore, I concluded that LOCPs should simply be treated as locative PPs, and that this analysis also allows us to account for the agreement relationships involving other clearly non-nominal expressions such as PPs and CPs in Chichewa. Because of their non-nominal nature, I argued that these should be appropriately identified as forming GACs and not noun classes. I proposed that there is a GAC morphological cue set that determines the agreement system. However, due to phonological identity, some nominal expressions also happen to trigger the GAC-related AMs. I also introduced three different types of underspecified agreement triggers, namely nominal, locative general and locative proximal. I showed that these underspecified agreement triggers correspond to the default AMs *zi-, ku-* and *pa-* respectively. The presence of non-nominal elements that trigger agreement was argued to be strong motivation for identifying the GAC subsystem as one that is distinct from the typical ANC subsystem.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 7
The underlying principles of the Chichewa noun and agreement class system

7.1 Introduction

In the present chapter, I discuss the underlying principles of the Chichewa agreement-based noun class (ANC) and general agreement class (GAC) systems. Specifically, I intend to probe why the Chichewa ANC and GAC systems might be organised in the manner proposed in Chapters 5 and 6. To address this objective, I will show how the observed organisation of the Chichewa ANC system correlates with universal tendencies of language structuring and language acquisition. The rest of the discussion is organised as follows: In Section 7.2, I introduce the three key organisational principles underlying the Chichewa ANC system, namely (i) the *semantic primacy principle* (Section 7.2.1), (ii) the *phonological primacy principle* (Section 7.2.2) and (iii) the *constituent-edge principle* (Section 7.2.3). In each of these sections, I show that all the three principles are manifested in many of the world's languages and that they are pervasive in a wide range of grammatical structures. A consolidated account of how the three principles converge in making the acquisition of the Chichewa ANC system logically possible is offered in Section 7.3. Here, I also show that the interaction of the three principles in language acquisition is amenable to three theoretical accounts, namely the Three Factors Model generally (Chomsky, 2005; Biberauer, 2017, 2018a), Maximise Minimal Means specifically (Biberauer, 2017, 2018a) and the Tolerance Principle (Yang, 2016). In Section 7.4, I argue that the Chichewa ANC system is the basic system which the GAC builds on. Overall, I show that the various components making up the ANC and GAC systems are organised in a hierarchical order, a property observed with several other grammatical systems. Finally, I summarise and conclude the chapter in Section 7.4.

7.2 Basic principles of the Chichewa agreement-based noun class system

As introduced in Chapters 3, 4 and 6, the corpus evidence has shown that there are two broad types of expressions that trigger agreement in Chichewa. The first type includes typical nouns, which govern agreement in a morphosyntactically relatively unpredictable way. The second set includes a mix of complex expressions, which govern agreement in a morphosyntactically more predictable manner. Some of the expressions in the latter group are nominal in nature while others are not, such as PPs and CPs. On the basis of this

Stellenbosch University https://scholar.sun.ac.za

distinction, I proposed in Chapter 4, Section 4.16 that the first type be considered as forming the typical *agreement-based noun class system* (see Chapter 5) and the second type as forming the more *general agreement class system* (see Chapter 6). Since acquisition research on Bantu languages has shown that the prefix system arises **after** children have shown the ability to use the agreement system of non-prefixed nouns (cf. i.a. Demuth & Weschler, 2012), my discussion will primarily focus on the properties of the ANC system. I will only turn to the relevant aspects of the GAC system later to show how it builds on the already established ANC system.

As shown in Chapters 4 and 5, there are three important determinants that appear to underlie the Chichewa ANC system, namely (i) the semantic feature of *animacy*; (ii) a noun-specific *phonological cue set* and (iii) the word/phrase-initial domain. I assume that these factors play a key role in structuring the ANC and GAC systems and I will accordingly refer to these factors as (i) the *semantic primacy principle*, (ii) the *phonological primacy principle* and (iii) the *constituent-edge principle*, respectively. I discuss each of these principles separately in the following subsections.

7.2.1 The semantic primacy principle

As discussed in Chapter 5, the Chichewa ANC system is primarily organised around a semantic core, specifically following the categorisation of the *animacy hierarchy*. Some aspects of the animacy hierarchy were introduced in Chapter 4, Section 4.3.10 (cf. Comrie, 1989, chap. 9; Yamamoto, 1999; Croft, 2003: 128–132). In the present section, I mainly focus on some cross-linguistic observations regarding the animacy hierarchy and its relevance in grammatical structure and language acquisition. I discuss the relevance of the animacy hierarchy in grammatical structure in Section 7.2.1.1 and in language acquisition in Section 7.2.1.2.

7.2.1.1 The animacy hierarchy and language structure

Although the capacity to make animate-inanimate distinctions is considered to be domain-general, that is, not intrinsically linguistic (Gelman & Opfer, 2002; Becker, 2014: 63), it is observed to systematically affect a wide range of morphosyntactic and pragmatic properties in a variety of world's languages (see i.a. Comrie, 1989, chap. 9; Dahl & Fraurud, 1996; Yamamoto, 1999: 1). Cross-linguistic research has shown that the animacy hierarchy manifests in several grammatical structures, for example, in word order preferences,

Stellenbosch University https://scholar.sun.ac.za

selection restrictions, morphological marking (e.g. case, object marking, agreement), number, person, gender/noun classification (Bloom, 1970; Lempert, 1989; Comrie, 1989; Pinker, 1996; Dahl & Fraurud, 1996; Yamamoto, 1999; Dahl, 2000; Gelman & Opfer, 2002; Aissen, 2003; Morimoto, 2008; de Swart *et al.*, 2008; Becker, 2009, 2014, chap. 3, 2015). For example, Chamorro (an Austronesian language of the Mariana Islands; Chung (2013: 251)) and Navajo (of the Na-Dené language family; Dryer & Haspelmath (2013)) require that the subject of a transitive sentence must be higher on the animacy hierarchy than objects (Comrie, 1989: 197; Becker, 2014: 66). In Japanese, Jacaltec and Blackfoot, in turn, inanimate subjects are banned in transitive sentences (Kuno, 1973; Comrie, 1989; Chung, 1998; Ritter & Rosen, 2010; Becker, 2014: 66).[27] According to de Swart (2008: 133), corpus studies have also shown that in transitive sentences of many languages (even in languages that do not exhibit an overt animacy distinction), animate NPs occur more often in subject positions and inanimate NPs in object positions (Dahl & Fraurud, 1996; Kempen & Harbusch, 2003; Øvrelid, 2004; Bresnan, Cueni, Nikitina & Baayen, 2007; Becker, 2014).

Language processing studies (e.g. comprehension) have also shown that animacy information plays a significant role in processing the interpretation of different types of constructions (see i.a. Stowe, 1989; Weckerly & Kutas, 1999; Mak, Vonk & Schriefers, 2002; van Nice & Dietrich, 2003; Lamers & de Hoop, 2004; de Hoop & Lamers, 2006; de Swart *et al.*, 2008; Pae, Schanding, Kwon & Lee, 2014). This implies that the animacy distinction can help the hearer/reader to parse sentences (de Swart *et al.*, 2008: 134).

As already stated in Chapter 5, Section 5.5.2, studies in grammatical gender systems have concluded that all attested gender systems have a semantic core: that is, they have a semantic basis for classifying all or a subset of nouns in a particular language (Corbett, 1991: 8, 2013). Of great relevance to the present discussion is that this semantic classification is generally based on the animacy hierarchy, where animate nouns are usually treated on the basis of semantic rules while inanimate nouns are treated by formal rules (Corbett, 1991, chaps 2–3; Dahl & Fraurud, 1996: 55; Dahl, 2000: 101). According to Dahl's (2000: 102) analysis, even in sex-based gender systems, the sex distinction only superficially overshadows the importance of animacy because the sex distinction is only

[27] Jacaltec belongs to the Mayan family of languages, spoken in Guatemala while Blackfoot belongs to the Algic family of languages, spoken in Canada and United States of America (Dryer & Haspelmath, 2013).

Stellenbosch University https://scholar.sun.ac.za

applicable to entities above a certain cut-off point in the animacy hierarchy. Therefore, sex differentiation can be understood as falling within the animacy hierarchy. In sum, Dahl & Fraurud (1996: 47) argue that the animacy distinction is so pervasive in human languages that "… it tends to be taken for granted and become invisible."

If animacy is this pervasive in linguistic structure, we might expect it to be a salient feature in language acquisition in general (Becker, 2014). In the next subsection, I consider the relevance of the animacy feature in the course of language acquisition.

7.2.1.2 Animacy and language acquisition

In addition to observing the pervasiveness of animacy in a variety of grammatical structures, many studies have also observed that animacy plays an important role in language acquisition (Gelman & Opfer, 2002; Childers & Echols, 2004; Becker, 2009, 2014, 2015). Evidence from experimental studies of infant perception and the conception of animacy has shown that babies can distinguish animates from inanimates at a very early stage in their lives (Gelman & Opfer, 2002; Becker, 2014). It is generally observed that there are properties of animates that children specifically pay attention to, for example, featural cues (e.g. face), dynamic cues (e.g. self-propelled motion), (Gelman & Opfer, 2002: 154–155; Studdert-Kennedy & Goldstein, 2003: 247; Becker, 2014: 166). In addition to these findings regarding children's perception of animacy, it has also been argued that animate things play a special role in children's general cognitive apprehension of the world (Slobin, 1981; Legerstee, Pomerleau, Malcuit & Feider, 1987; Legerstee, 1994; Becker, 2014: 166).

It has also been widely argued that the ability to distinguish between animate and inanimate things is also found in non-human animals, suggesting that this capacity is domain-general (Hare, Call, Agnetta & Tomasello, 2000; Gelman & Opfer, 2002: 151; Flombaum & Santos, 2005; Becker, 2014: 287–289). For example, Gelman & Opfer (2002: 151) argue that a creature that is unable to distinguish inanimate from animate can be "severely impaired". Following these observations, it is assumed that the capacity to make animacy distinction may indeed constitute a *domain-general innate knowledge* (Becker, 2014: 166); alternatively, it must at least be a concept that is acquired very early indeed.

The question that follows then is, at what stage do children begin to translate this perceptual capacity into conceptual categories? According to Rakison & Poulin-Dubois (2001: 213), developing conceptual knowledge implies that babies "have disassociated perceptual

Stellenbosch University https://scholar.sun.ac.za

information from its source and converted it into some kind of abstract representational format" (cf. Mandler & McDonough, 1993; Molina, van de Walle, Condry & Spelke, 2004; Becker, 2014: 161, 164).

Children must indeed begin to create conceptual categories at an appropriate age to be able to employ these distinctions in the various lexical and grammatical structures that they are in the process of acquiring (Becker, 2014: 156, 242). As Childers & Echols (2004: 110) observe, "a major development in language acquisition is learning how to encode information about who-did-what-to-whom in language." They (2004: 111) argue that one conceptual cue that might be implicated in making these semantic categories is that of animacy because animate entities tend to be agents and inanimate entities tend to be patients. Therefore, it is assumed that children will pay attention to the animacy of NPs. Becker (2014: 173) claims that by the time robust syntactic acquisition begins (around 18 months), children have the basic expectations for encoding animacy distinction. She (2014: 68) argues as follows:

> … children would be at a distinct advantage if they expect animate NPs to line up with the subject and/or agent role and/or topic, and inanimate NPs to line up with the object and/or patient/theme role and/or non-topic, since this is clearly the unmarked pattern. (cf. Slobin, 1981: 187; Pinker, 1996: 128–130; Childers & Echols, 2004; Becker, 2015)

In this regard, children can exploit this regularity to identify basic sentence structure, for example, "find the most animate noun and it will be the subject" (Becker, 2014: 2). If animacy plays such a significant role in the syntactic development of something as central as argument encoding, it should not be surprising that noun (and agreement) class systems are also based on the animacy hierarchy: noun class systems are precisely manifested in syntactic processes, such as in subject-verb, object-verb and head-modifier agreement.

There are two relevant predictions that can be drawn from Becker's (2014) observations. First, noun class systems and morphosyntactic development are closely related. For example, in Chichewa, all verbs and modifiers in construction with NPs must morphologically mark this syntactic relationship in the form of agreement markers (AMs). Therefore, in Becker's (2014) terms, for example, one can postulate a rule that the agent of the canonical sentence will be animate and trigger a particular AM or that canonical patients will be inanimate and also trigger different AMs. In addition, in the case of some Bantu and other languages, where OMs only track the most animate objects, we see another domain where the animacy hierarchy is straightforwardly harnessed. Since, as discussed above, most

Stellenbosch University https://scholar.sun.ac.za

prototypical sentences involve agent subjects, the animate-inanimate cue for AM selection would be expected to arise much earlier than agreement marking involving other cues. This could serve as a crucial piece of support for the semantically based super-ANCs proposed in Chapter 5, as (4), repeated here in (1).

(1) **The semantically-determined super-ANC in Chichewa**

Animate nouns		**Inanimate nouns**	
m(u)-a-ANC	(≈NC1)	*li*-ANC	(≈NC5)
		i-ANC	(≈NC4 and 9)
		u-ANC	(≈NC3 and 14)
		a-ANC	(≈NC2 and 6)
		chi-ANC	(≈NC7)

Strikingly, it is the animate class only that directly selects its AM based on the semantic basis, while those belonging to the inanimate class are further classified based on word-initial phonological characteristics.

The second prediction following Becker's (2014) views above is that the animacy-based ANC system proposed for Chichewa is not a language-specific property but a more universal grammatical phenomenon (cf. Corbett, 1991, chaps 2–3; Dahl & Fraurud, 1996: 55; Dahl, 2000: 101). This implies that the structure and acquisition of the Chichewa ANC can also be understood in terms of the identified universal tendencies regarding the animacy hierarchy and grammatical structure.

Having provided this overview about animacy and language acquisition, I next summarise the discussion regarding the *semantic primacy principle*.

7.2.1.3 Interim summary

The discussion focusing on the *semantic primacy principle* has highlighted four key points. First, animacy is commonly manifested in many world languages, including in languages where this feature is not commonly identified as being centrally significant. Second, animacy may affect a variety of grammatical structures, such as word order, argument marking, noun class, etc. Third, it seems rather clear that the animacy distinction is a domain-general capacity and that this capacity is significantly active in children at a very early stage of life, importantly, before syntactic development (Becker, 2014). Fourth, the capacity to distinguish animate from inanimate entities has been reported to play a role in bootstrapping the

Stellenbosch University https://scholar.sun.ac.za

acquisition of some aspects of linguistic structure, such as basic word order and argument encoding. Against this backdrop, it is of great interest that animacy emerges as it does (see again Chapter 5, Section 5.3) in my consideration of the properties that appear to determine the structuring of the Chichewa ANC system. Similarly, the fact that animacy is not a core organising principle in the Bleek-Meinhof system constitutes a major shortcoming for this approach (see again Chapter 2, Section 2.4.3).

In the next section, I consider the second core principle of the Chichewa ANC system, namely the *phonological primacy principle*.

7.2.2 The phonological primacy principle

As discussed in Chapter 5, the inanimate nouns in the semantic component of the ANC system are further subcategorised based on word-initial phonological features. These phonological features were analysed as forming the *phonological cue set* for the Chichewa ANC system (see Chapter 5, Section 5.4, repeated in Table 7.1).

Table 7.1 Key phonological features of the Chichewa ANC system

	Noun class	Phonological features	Attested recurring word-initial phonemes
a.	*i*-ANC	[+cor], [+ant], [-back]	/N/, /i/, /s/, /t/
b.	*u*-ANC	[-cor], [+back], [+labial], [+round]	*/m/, /mu/, /mo/, /u/, /o/*, */b/, /bw/*, */w/*
c.	*a*-ANC	[-cor], [+back], [-labial], [-round]	*/ma/, /a/*
d.	*chi*-ANC	[+cor], [-ant], [-back]	/ʧ/
e.	*li*-ANC	Elsewhere	Inanimate nouns whose word-initial phonological features are not found in the *phonological cue set* in (a-d)

From a cross-linguistic perspective, the phonological features in Table 7.1 exhibit three significant properties that are of relevance in the present discussion, namely (i) they are primary and highly frequent phonemes in the phonological inventories of world's languages; (ii) they can be shown to interrelate in a hierarchical structure as predicted by the theory of *feature geometry* and (iii) they are found to be the most salient and the first phonemes to be acquired (see i.a. Clements, 1985; McCarthy, 1988; Stoel-Gammon & Herrington, 1990; Ladefoged & Maddieson, 1996; Studdert-Kennedy & Goldstein, 2003).

I discuss each of these properties in the following subsections.

Stellenbosch University https://scholar.sun.ac.za

7.2.2.1 The phonological cue set and cross-linguistic primary phoneme inventories

Focusing on the first property of the *phonological cue set*, one striking observation about the phonological systems of the world's languages is that they categorise phonemes in rather similar ways (McCarthy, 1988: 84; Rice & Avery, 1995: 25; Ladefoged & Maddieson, 1996: 4; Studdert-Kennedy & Goldstein, 2003: 244). Many studies have shown that the phonological systems of all languages make a primary distinction based on place of articulation contrasts, which involve the following speech organs and places of articulation: the lips (labial phonemes), the tongue tip (coronal phonemes), the tongue body (dorsal phonemes) and, in some languages, the velum (nasal phonemes) (McCarthy, 1988; Lahiri & Evers, 1991; Gierut *et al.*, 1993; Studdert-Kennedy & Goldstein, 2003: 244; Chen & Kent, 2005; Johnson & Reimers, 2010, chap. 1). With specific reference to vowel inventories, since the work of Jakobson (1968), it has been consistently observed that almost all languages include at least the three cardinal vowels (/i/, /a/ and /u/), referred to as the *fundamental vowel triangle* or *corner vowels* (cf. Stoel-Gammon & Herrington, 1990: 148; Stoel-Gammon & Pollock, 2008: 529; Johnson & Reimers, 2010: 53).

Strikingly, as shown in Table 7.1, the Chichewa *phonological cue set* includes these primary phonemes (both the consonant and vowel sets). For example, the *phonological cue set* primarily centres on place of articulation contrasts (e.g. coronal, labial, anterior). Given these patterns, a relevant question would be how these specific phonemes came to be such crosslinguistically basic and frequent elements? From an evolutionary linguistic perspective, Studdert-Kennedy & Goldstein (2003: 253) hypothesise that the speech organs involved in the production of these phonemes "provide the universal biologically determined base for discrete units of phonetic action in every language." Therefore, each place of articulation is assumed to represent a phonetic gesture and the less frequent phonemes arose over time, as different linguistic communities expanded and elaborated the basic phonological gestures into various subcategories of the main gestures produced by a particular organ. If Studdert-Kennedy & Goldstein's hypothesis is to be adopted, the phonemes in question simply fall within the cardinal phonetic gestures; hence their observed tendency to universality. However, irrespective of this hypothesis, the key point here is that phonemes that are attested as forming crosslinguistic basic inventories are also found to underlie the Chichewa ANC phonological cue set.

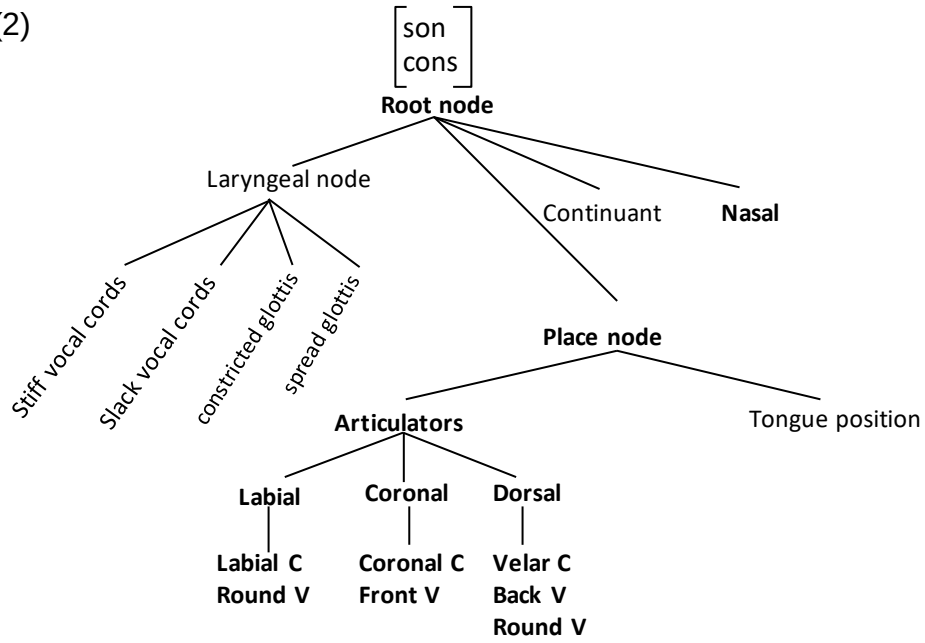
Stellenbosch University https://scholar.sun.ac.za

In the next subsection, I present the second observed cross-linguistic property of the phonemes in the *phonological cue set*.

7.2.2.2 The phonological cue set as predicted by the theory of feature geometry

The second property of the phonological features comprising the *phonological cue set* is predicted by the theory of *feature geometry*, which has shown that these features are hierarchically organised, in a very similar manner in many languages (Clements, 1985; McCarthy, 1988; Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003). Although there are many versions of *feature geometry*, they make rather similar predictions (see i.a. Gierut *et al.*, 1993: 223; Chen & Kent, 2005). For the purposes of the present discussion, I adopt the model proposed by McCarthy (1988: 105), subsequently revised in Lahiri & Evers (1991) and Gierut *et al.* (1993).[28] The adapted model of the relevant version of the *feature geometry* is illustrated in (2).

(2)



In (2), each node represents a *class node*, which may dominate other subordinate nodes or sets of individual features. At the very top is the *root node*, which includes the major class features sonorant ([son]) and consonantal ([cons]), while the terminal nodes represent the actual phonological features (McCarthy 1988:89). All subordinate nodes and terminal nodes are taken to be in a dependency relation with the major class features (McCarthy, 1988: 97).

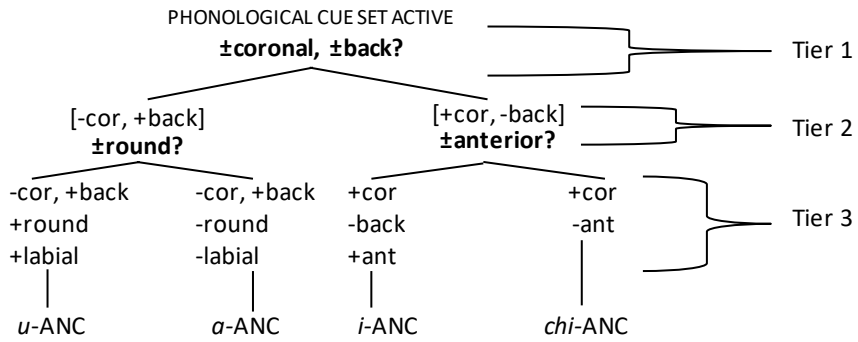
[28] Justification for this choice will be provided below.

Stellenbosch University https://scholar.sun.ac.za

For example, the *Laryngeal node* and the *place node* are subordinates, and therefore dependents, of the *root node*. According to McCarthy (1988: 97), these phonological features form natural classes that predict a variety of phonological processes and categorisations in many languages (see also Clements, 1985; McCarthy, 1988; Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003: 244; Chen & Kent, 2005). For example, the *feature geometry* explains the behaviour of some phonological processes, such as *spreading*, in which the features of the higher node may extend over phonemes in the subordinate nodes, but not to members outside the dominated class (McCarthy, 1988: 86, 97; Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003: 244). The organisation of these phonological features is also similarly predicted by Place of Articulation (POA) Theory (Chomsky & Halle, 1968) and Articulator Theory (AT) (Sagey, 1986; McCarthy, 1988; Mester, 1988).

Strikingly, the feature geometry also predicts the clusters which form the *phonological cue set* that derive the four phonologically determined Chichewa ANC represented in Table 7.1. The specific hierarchical organisation of the Chichewa phonological cue set is illustrated in (3) (see also (10) in Chapter 5).

(3) **The hierarchical order of the Chichewa ANC phonological cue set**[29]



In (3), the phonological cue set in Chichewa is represented as involving at least three tiers of hierarchical organisation. Tier 1 represents the parent node of the system, which is observed to involve phonological features [±cor] and [±back]. Tier 2 has two nodes, each dependent on the parent node's features, namely [-cor, +back] and [+cor, -back]. The first node ([-cor, +back]) in Tier 2 is further subcategorised into two nodes based on the feature

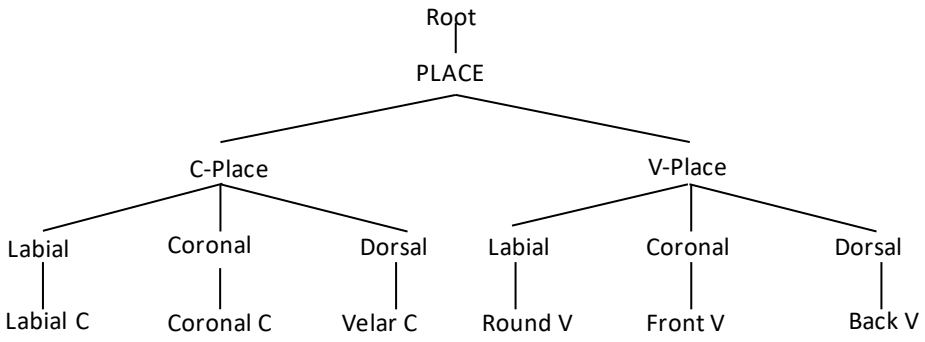
[29] Note that here I have excluded the Elsewhere-based *li*-ANC since it does not, by virtue of its Elsewhere status, directly participate in the hierarchy of features.

Stellenbosch University https://scholar.sun.ac.za

±round and ±labial. Whereas, the second node of Tier 2 ([+cor, -back]) dominates two other nodes in the Tier 3, which are distinguished based on the features [+ant] and [-ant]. As shown in (3), the hierarchical organisation of the Chichewa phonological cue set corresponds to the place node tier of the feature geometry illustrated in (2). Strikingly, it is the same phonological features of [±cor], [±round], [±back], [±labial], etc. that are involved in both featural hierarchies. There are also similarities in consonant and vowel symmetry, which I consider next (see again Chapter 5, Section 5.4).

As stated above, there are different models of *feature geometry*. Some of these variations are of particular relevance to the present discussion, specifically regarding consonant-vowel symmetries. I therefore, briefly introduce these variations. One facet on which various models of *feature geometry* differ is with respect to how consonants and vowels are characterized in the feature hierarchy. According to Gierut, Cho & Dinnsen (1993: 220-222), there are three variations on how consonants and vowels are represented in a feature geometry. In the first, which follows the structure in (2), the place features for consonants and vowels are represented on the same tier - i.e. vowels and consonants are interdependent; this is the so-called *fully integrated feature geometry* (Gierut *et al.*, 1993: 220–1). In the second version, the place features for consonants (C-Place) and vowels (V-Place) are isolated; therefore, they are independent and are represented on non-interacting tiers; this is the so-called *fully segregated feature geometry*, as in (4).

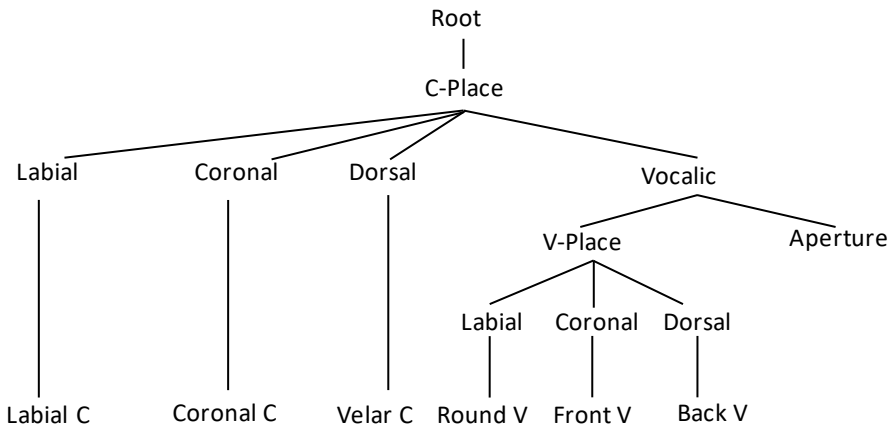
(4) *A fully segregated feature geometry* (Gierut *et al.*, 1993: 221)



In the third model, the so-called *partial segregation feature geometry*, the consonant and vowel place features are represented on different tiers, but overlap depending on the place of articulation for the vowel and the consonant (Gierut *et al.*, 1993: 222). See the model in (5)

Stellenbosch University https://scholar.sun.ac.za

(5) *Partial segregation feature geometry* (Gierut *et al.*, 1993: 222)



Note that in (5) the vowel place of articulation is subsumed by the C-place, but each is represented in non-interacting tiers; hence the designation *partial segregation*.

There is one important point regarding the variations of the feature geometry illustrated in (2), (4) and (5). It is shown that despite the differences in the models of feature geometry, they make quite similar predictions in terms of the possible features that may go together and features that may not go together (cf. Gierut *et al.*, 1993: 223; Chen & Kent, 2005: 513). For example, although the *fully integrated feature geometry* (2) seems to provide the most logical characterisation, all three versions confirm the consonant-vowel symmetry observed in the phonological cue set underlying the Chichewa phonologically-determined super-ANC system. Worth noting here is that in addition to the phonemes occurring crosslinguistically, they are also organised in a similar manner.

Having discussed the hierarchical organisation of the phonological features in the *phonological cue set*, I next present some observations regarding how these phonological features interact during the process of language acquisition.

7.2.2.3 The phonological cue set and language acquisition

The third aspect regarding the properties of the phonemes in the *phonological cue set* is that these phonemes are acquired in accordance with the hierarchical structure predicted by the *feature geometry*. Studies focusing on the development of phonemic inventories have shown that phoneme acquisition takes particular developmental paths (see i.a. Stoel-Gammon, 1985; Gierut, Simmerman & Neumann, 1994; Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003; Stokes, Klee, Carson & Carson, 2005). For example, Stokes *et al.* (2005) proposes an *implicational feature hierarchy* of phonemic development in typically

Stellenbosch University https://scholar.sun.ac.za

developing English speaking children. See Table 7.2, which shows the development of the phoneme inventory in both word-initial and word-final positions during acquisition (Stokes *et al.*, 2005: 825–826).

Table 7.2: Stages depicting phoneme inventory acquisition of English-speaking children

Development stage	Phonological feature acquired
Level I	[consonant], [sonorant], [coronal]
Level II	[voice], [continuant]
Level III	[anterior], [nasal]
Level IV	[lateral], [delayed release], [strident]

As shown in Table 7.2, the levels correspond to the feature hierarchy predicted by the *feature geometry* in (1), where [cons], [son] and [cor] are the major class features, which are also observed to be acquired first. According to Studdert-Kennedy & Goldstein (2003: 253), the development of phonological features is a cumulative process which follows a hierarchical order which starts with simpler and more general features and becomes increasingly more complex and specific (cf. Gierut *et al.*, 1993). This implies that the features in the feature hierarchy are primarily organised with respect to simplicity and generality. Therefore, although there is need for Chichewa-specific evidence of phonemic development, it is reasonable to predict that the topmost features on both the hierarchy of the *feature geometry* and the Chichewa ANC *phonological cue set* are the most general, simple and the first acquired. The key finding here is that nominal classification can be traced further to the natural class features of phoneme categorisation.

Therefore, like the *semantic primacy principle*, the *phonological primacy principle* also accounts for phonemes that are salient at the very early stage of the child's language development. It should be plausible, therefore, to generalise that the semantic and phonological features that underlie the structure of the Chichewa ANC system are features that are, crucially, acquired (very) early, and that they should serve as the basis for the structuring of divergent formal systems, such as the noun class. In what follows, I will show that the salience of the early phonemic distinctions is further enhanced by the third principle of the Chichewa ANC system, namely the *constituent-edge principle*.

I briefly introduce the *constituent-edge principle* in the next section.

Stellenbosch University https://scholar.sun.ac.za

7.2.3 The constituent-edge principle

As shown in Chapter 5, the phonological features that underlie the Chichewa ANC system occur in word-initial positions, which points to another crosslinguistically oft-referenced structural category, namely the so-called *constituent-edge*. Constituent-edge patterns manifest in a variety of linguistic phenomena (see i.a. McCarthy & Prince, 1993: 79; Adger, de Cat & Tsoulas, 2004; Sekerina & Brooks, 2007; Endress, Nespor & Mehler, 2009: 351; Kandybowicz, 2009; Ko, 2014; Benavides-Varela & Mehler, 2015). Firstly, many phonological processes are observed to apply with respect to constituent-edges. In many languages, for example, stress is always assigned relative to word-edges (Hayes, 1995; Marvin, 2002: 43–48; Endress *et al.*, 2009: 351). For example, stress falls on the left edge in Hungarian, on the right edge in French, mostly on the second syllable from the right in Italian (Endress *et al.*, 2009: 351). It was also stated in Chapter 5, Section 5.5.2 that in languages that employ phonologically determined agreement systems, the agreement is usually triggered by either word-initial or word-final phonemes (cf. Foley, 1986, 1991; Corbett, 1991; Sande, 2015, 2017). With specific reference to phonologically determined agreement, Sande (2017: 73) makes the following prediction.

(6) Any edge-aligned or prominent segment or suprasegmental can control agreement.

The Chichewa agreement system in general is also observed to be organised in similar ways, as all phonological features of the agreement trigger are word-initial; this is also the case with the corresponding AM features on the agreeing hosts. Regarding the prominent segment aspect, it was shown in (5-6) of Chapter 5, that the relevant phonological features in many nasal-initial nouns are determined by the syllable nucleus segment and not necessarily the syllable onset. In this case, the relevant cue is in both the syllable-prominent and constituent-edge positions. I also showed in Chapters 3 and 4 that suffixes, i.e. word-final elements play a semantic role in determining whether a particular noun is agentive or not. In this case, the word-final segments may also contribute to the semantic salience of agentive and non-agentive nouns, which is also a rule worthy property for determining the super-ANCs.

Secondly, morphosyntactic processes, such as duplication and affixation, are also cross-linguistically observed to follow *constituent-edges* (word-initial or word-final), whereas medial reduplications or infixes, are very rare (Broselow & McCarthy, 1983; Endress *et al.*, 2009: 351). See also the discussion in Chapter 6, Section 6.4.4 where I reported that

Stellenbosch University https://scholar.sun.ac.za

syntactic processes involving speaker-hearer perspective markers and their related speech act attenuation or intensification functions are cross-linguistically observed to be edge-related (cf. Cinque, 1999; Prieto, 2005; Heim *et al.*, 2016; Wiltschko, 2017; Biberauer, 2018c).

Therefore, as with the semantic and the phonological principles above, the *constituent-edge principle* also manifests in a wide range of grammatical phenomena in many languages. Importantly, constituent-edge generalisations have also been observed to play an important role in the process of language acquisition. In the next subsection, I highlight the relevance of the constituent-edge principle in language acquisition.

7.2.3.1 The edge principle and language acquisition

Edge-based positional cues are also found to be significant in language acquisition. For example, children are observed to be sensitive to word segment categorisation, with word-initial and word-final being the most salient segments (McCarthy & Prince, 1993; Ingram & Ingram, 2001: 276; Endress *et al.*, 2009: 351; Cilibrasi, 2015). For instance, word-initial phonemes are among the most prevalent in children's developing phonological inventories (Ingram & Ingram, 2001: 280). Renfrew (1966) also reports a phonological impairment called *persistence of open syllables*, a situation where the word-initial consonant productions are relatively advanced in relation to the final consonants, which are uniformly missing (cf. Ingram & Ingram, 2001: 276).

Edge-based positional regularities are also assumed to form part of the contrasts that increase perceptual salience (Smith, 2005; Cilibrasi, 2015). For example, prosodic domain-initial consonants in many languages are observed to undergo a process of strengthening relative to their non-initial counterparts (see i.a. Cho & Jun, 2000; Cho & Keating, 2001; Smith, 2005). Therefore, word-initial consonants often realize more contrasts than do consonants elsewhere in the word, implying that further to having the features in the *phonological cue set* being higher in the inventory hierarchy and simpler in articulation, they are also made more salient in the word-initial positions.

Recall that the phonological features in question are found on both the agreement trigger expressions and on the agreeing host; put differently there is a *doubling* pattern. Therefore, in addition to having these ANC phonological features in more prominent word-initial positions, the salience of these features is enhanced by the doubling pattern at the initial

Stellenbosch University https://scholar.sun.ac.za

edges of words in phrases and sentences. Specifically, the AMs consistently manifest as prefixes on stems of various kinds, such as verbs and modifiers. Consider, for example, the segments in boldface on the noun and the agreeing host in (7a-b) and (7b), representing the *i*-ANC and *chi*-ANC, respectively.

(7) a. **Mb**alame **i**-modzi **i**-na-tera pa denga
bird (*i*-ANC) AM-one AM-T/A-petch LOC roof
"One bird perched on the roof."

b. M**i**nga **i**-modzi **i**-na-ndi-baya
thorn (*i*-ANC) AM-one AM-T/A-OM-prick
"One thorn pricked me."

c. **Ch**itsulo **chi**-modzi **cha**-chita dzimbiri
metal (*chi*-ANC)AM-one AM.T/A-do rust
"One metal object has developed rust."

Since constituent-edges increase salience, it means that in Chichewa the ANC phonological features and the doubled AMs provide reliable cues for both language acquisition and language processing.

However, what remains to be investigated further is how the matching of phonological features in agreement relationships is implemented in the syntax (cf. Sande, 2017). There are considerable disagreements about whether phonological features can impact syntactic operations, such as agreement. On the one hand, the majority of scholars argue for what is known as *phonology-free syntax,* where phonological features are taken to have no bearing on how syntactic processes (e.g. agreement) are implemented (see i.a. Pullum & Zwicky, 1988; Pollard & Sag, 1994: 60; Dobrin, 1995, 1998; Dimitriadis, 1997: 1–2; Corbett, 2006, chaps 89, 91; Scheer, 2011; Chung, 2013: 251; Kramer, 2015: 250; Sande, 2016, 2017, chap. 3). Others, however, argue for various modes of phonological influence in syntactic processes (see i.a. Baker & Brame, 1972; Hetzron, 1972; Tegey, 1975; Dobrin, 1995, 1998; Aronoff, 1998; Jackendoff, 2000; Richards, 2016). In the present study, I will not pursue this question as it falls outside the purview of the set aim, which is primarily concerned with the underlying principles of the Chichewa noun and agreement class system.

In this regard, I conclude that the highly salient animacy features, the prominent phonological features and the edge-based positional regularities are all key determinants that aid in the acquisition of the Chichewa ANC system. In the next section, I consider how these three principles can illuminate our understanding of how the proposed Chichewa ANC system could plausibly be acquired.

Stellenbosch University https://scholar.sun.ac.za

7.3 Acquisition of the Chichewa ANC and GAC systems

In the present section, I consider how the three principles converge to derive the type of ANC system proposed in Chapter 5. Specifically, I attempt to answer the following question: at what point do the three principles converge to derive the hierarchical organisation that is observed in the Chichewa ANC system? I will answer this question in two main parts. In the first part, I focus on the principles that are implicated in the initial trigger of linguistic development. Then in the second part, I sketch some assumptions regarding how these principles interact in linguistic development.

7.3.1 The perceptual triggers of language development

Although both sensitivity to animacy and phonological perception are observed to be accomplished very early in life, studies have shown that phonological perception is the first trigger for linguistic development (see Mulford, 1985; Kelly, 1996; Demuth, 2000: 284; Culbertson *et al.*, 2017). Kelly (1996: 253) states that "children in the first year of life achieve a number of perceptual prerequisites to the use of phonological cues to grammatical class" (cf. Stokes *et al.*, 2005; Gervain & Werker, 2008). For example, consider the phonologically biased linguistic development milestones summarised in Table 7.3.

Table 7.3: Infants' perception of phonological prerequisites that cue grammatical class (Kelly, 1996: 252–253; Gervain & Werker, 2008)

Age	Perceptual milestone	
Neonates	Newly born babies show the ability to discriminate their parents' language from a foreign one, based on phonological perception, specifically the basic prosody of the language.	(Mehler, Jusczyk, Lambertz, Halsted, Bertoncini, *et al.*, 1988: 148; Gervain & Werker, 2008; Johnson & Reimers, 2010, chap. 4)
6 months	Infants have organised vowel categories around prototypes relevant to their language.	(Grieser & Kuhl, 1989: 586–587)
9 months	American infants show preference to listen to first syllable stress words over second syllable stress words.	(Jusczyk, Cutler & Redanz, 1993: 682ff)

It has further been argued that many of the phonological achievements (e.g. those in Table 7.3) correlate with divergent aspects of grammatical structure. For example, in English, nouns have more low vowels while verbs have more high vowels, implying that by 6 months an infant has begun to indirectly categorise nouns and verbs – see, for example Nespor, Peña & Mehler (2003). It has also been observed that disyllabic nouns have trochaic stress

Stellenbosch University https://scholar.sun.ac.za

while verbs have iambic stress, implying that the prosodic differences would help differentiate between verbs and nouns (Kelly, 1996: 252). Once phonological perception triggers the development of language, children are attuned to pay special attention to these salient phonological aspects (see i.a. Studdert-Kennedy & Goldstein, 2003; Gervain & Werker, 2008).

The key finding illustrated in Table 7.3 is that phonological perception appears to play a key role in triggering the earliest aspects of language acquisition. In addition to this, children are very sensitive to certain aspects of the sound input from birth. As discussed in Section 7.2.2.2, these phonological aspects make up phonemes that child acquirers produce at the earliest stages of acquisition. These phonemes are also shown to be essential building blocks for all languages. It is therefore telling that these are precisely the phonemes that matter in the structuring of the Chichewa ANC system.

In the next section, I consider how these salient phonological features may interact with the semantic feature of animacy.

7.3.2 The grammatical encoding of semantic features

Having demonstrated the plausibility of assuming that phonological features play a role in triggering linguistic development, I now consider how these phonological features interact with the relevant semantic features. To begin with, I appeal to Slobin's (1981) work, *The origins of grammatical encoding of events*, and specifically his account of how the pairing of form and meaning begins. He (1981: 185) argues that the notions that are grammaticalized first are those which are salient to the child; he refers to these as *prototypical situations*. The prototypical situations are assumed to be encoded in the most basic grammatical forms in the ambient language. Slobin (1981: 185) refers to these highly salient grammatical structures as *canonical* grammatical forms. The beginning of grammar, therefore, involves paying "attention to both prototypical situations in the world of reference and canonical forms in the world of language".

As introduced in Chapter 1, in Biberauer's (2017) MMM terms, this intersection could be understood to produce basic Saussurean arbitrariness. However, the Saussurean arbitrariness assumption only accounts for arbitrary one-to-one pairings of form and meaning; it does not account for the acquisition of grammatical patterns, such as agreement and word order, or for the nature of functional elements like underspecified forms and

Stellenbosch University https://scholar.sun.ac.za

expletives (see again, Chapter 1, Section 1.3.1). In this regard, Biberauer (2017) argues that children pay attention not just to arbitrary form-meaning mappings, but also to the systematic **departure** from this arbitrary pairing of meaning and form. Crucially, the Chichewa agreement system in general typifies a case where there is systematic departure from Saussurean arbitrariness because agreement **systematically** requires the presence of "extra" form which doubles (part of) a meaning that is already expressed by another element. For example, the AMs on modifiers and verbs are consistently required without contributing meaning independently of that which is encoded on the noun with which these elements agree. What we consistently have, then, are multiple forms conveying one meaning. In this regard, as introduced in Chapter 1, acquirers need to establish the nature of the pattern regulating the distribution of doubled elements, which results in the postulation of a formal [F] feature to capture the systematic distribution of the doubled form.

In the next section, I sketch how the Chichewa ANC subsystem could be acquired.

7.3.2.1 Acquisition of the ANC system

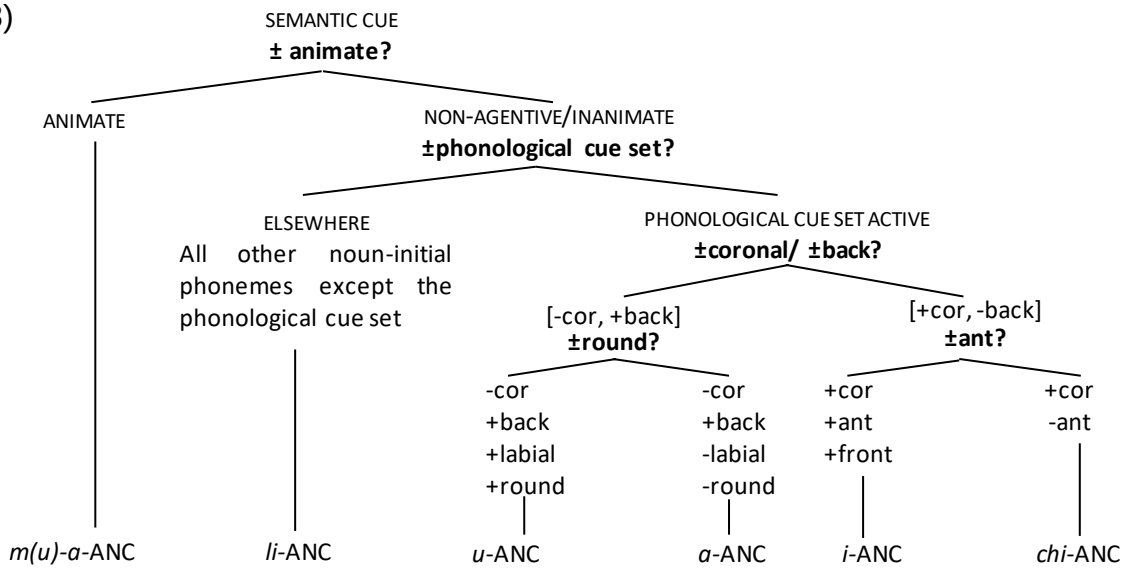
In the context of a child learning Chichewa, I assume that animacy is central to one of language acquirers' most salient *prototypical* situations and that children notice regularities regarding the way reference to animate as opposed to inanimate entities is achieved in phrases and sentences in the target language (cf. Slobin, 1981: 185; Childers & Echols, 2004: 110). As introduced in Section 7.2.1, one early milestone in language acquisition is encoding information about "who-did-what-to-whom" (Childers & Echols, 2004: 110). In this regard, it is worth noting that word order - e.g. SVO, with Subject before the verb and Object immediately after the verb – is more basic than agreement. So acquirers will already know that their language is VO - something that is today quite generally believed to be cued by prosody (see Gervain & Werker, 2008 for overview discussion) - with the result that they can map the Subject vs Object placement facts onto the animacy observation, and, from there, work out what the agreement system marks (see i.a. Dahl & Fraurud, 1996; Kempen & Harbusch, 2003; Øvrelid, 2004; Bresnan et al., 2007; Becker, 2014).

A child learner, therefore, pays attention in the first instance to the arbitrary form and meaning pairs that derive nominal expressions in Chichewa. Second, the child notices the semantic (animacy) effect in the word order patterns, for example that animate NPs are predominantly preverbal and inanimate NPs are postverbal; this leads to the postulation of animacy-oriented rules that regulate word order. Third, the child notices that these NPs are

Stellenbosch University https://scholar.sun.ac.za

in particular relationships with other words that they are in construction with and that this relationship frequently picks up on animacy distinctions which are marked via AMs (e.g. animate nouns controlling the *mu-a* AMs and inanimate nouns another set of AMs). The animacy and AM pairing thus gives rise to a distinction between the two animate and inanimate super-ANCs introduced in Chapter 5. Refer again to the flow diagram ((15) in Chapter 5), repeated here as (8).

(8)



As shown in (8) the animate class does not feature further subdivisions. This implies that the acquirer simply needs to learn a specific AM for the animate nouns. However, within the semantically less salient inanimate group, children further realise, thanks to the observed sub-regularities within this group, that they need to further refine the formally based inanimate system; and they do so drawing on their strong phonological sensitivities. As already discussed above, there is a sequence in which the phonological cues become active, where sub-categories can only be postulated after the higher categories have been identified. Strikingly, the hierarchical order in (8) also coincides with the predicted order of phoneme inventory development, as predicted by the feature geometry and the phoneme inventory acquisition studies in (2) and Table 7.2, respectively.

In this regard, I assume that within the phonologically determined class, children firstly learn the [±cor]- and [±back]-based noun class distinctions, which derive the broad ([+cor], [-back]) and ([-cor], [+back]) classes in (8). Only after the broader [±cor] and [±back] distinction has been achieved may acquirers be able to make further distinctions within the two sub-branches which involve the featural distinctions of ([+round] and [-round]) or ([+ant] and [-

Stellenbosch University https://scholar.sun.ac.za

ant]). The prediction is that children will only be capable of making the relevant distinctions for those features that are acquired first, i.e. those that are higher in the hierarchy of available phonemes than those shown to be acquired later, which are also lower in the hierarchy.

Since inanimate nouns that do not have any of the word-initial features associated with the *phonological cue set* are assigned their own default AM (*li*-), I conjecture that the Elsewhere-based *li*-ANC is also among the early classes to be acquired together with the animate and the [±cor]-based ANCs. Equally, the more general default *zi*-ANC should arise alongside the super-ANC due to its prevalence in the PLD. Take note that the default AMs *li*- and *zi*- are also [+cor], which implies that they are among the first phonemes to be acquired as predicted by the feature geometry and phoneme inventory acquisition studies. In effect, I expect the entire ANC system to be in place before the GAC, to which I turn to shortly. However, the exact order of acquisition of the two default classes can only be ascertained with a carefully set language acquisition study.

Following Yang (2016), I also assume that children will acquire the further distinctions upon encountering *enough* of such patterns in the PLD (cf. Pinker, 1984, 1996; Gagliardi, 2012). I set out the logic of how these patterns are assumed to be postulated and generalised by means of a rule-based computation in Section 7.4 below.

7.3.2.2 Acquisition of the GAC system

The account provided in the foregoing discussion has only focused on the morphologically less predictable nominal expressions that make up the ANC subsystem. However, as I have shown in Chapters 4, 5 and 6, there are GACs that do not include nouns alone but also other complex expressions such as CPs and PPs, which are consistently characterised by predictable morphosyntactic processes. As has been shown by Bantu NCP acquisition studies, agreement classification involving these 'prefix-based forms' arise later, at least after children have shown ability to trigger correct agreement of non-prefixed nouns (see i.a. Kunene, 1979; Suzman, 1980; Connelly, 1984; Tsonope, 1987; Demuth, 1988, 2000: 283; Idiata, 1998; Demuth & Weschler, 2012). To account for the acquisition of GACs, I first assume that the ANC subsystem forms part of the developmental preliminaries of the entire agreement system. Second, I assume that children continue to revise their rules depending on the growing complexity of their grammatical abilities (such as prefixation) and also upon encountering more productive rules that characterise the broader agreement system (Yang 2016, Biberauer 2017, see also Section 7.4). However, when children learn the

Stellenbosch University https://scholar.sun.ac.za

morphologically complex expressions such AUG, HON and ATT (see Chapter 6), whose prefixes have a form similar to the already existing ANC AMs, they do not postulate new AMs, but instead identify them as the already mastered ANC AMs.

For complex expressions that do not take the same AM forms as the already existing ANC AMs, children will revise the set of AMs accordingly by adding the *GAC cue set* in (9).

(9) a. *ka-* singular DIMs and manner nominals
b. *ti-* plural DIMs
c. *ku* locative phrases and nominal infinitives
d. *pa* locative phrases
e. *mu* locative phrases

As stated in Chapter 6, one important difference between the ANC and GAC cue sets is that it appears that the assignment criteria for the GACs are sensitive to the functional status of the affixes in (9). For example, in some cases where the non-prefixed word-initial phoneme is accidentally homophonous with the initial syllable of a particular GAC, the expression does not trigger the agreement associated with the GAC, but, instead, that of the typical ANC. Consider the following contrasting pairs (see also Chapter 6, Section 6.6).

Table 7.4 contrasts between homophonous word initial phonemes and functional prefixes

Phoneme	Noun class	Agreement class
ka-	*kalulu* (“hare”) *m(u)-a*-NC	*ka-munthu* (“small person”) *ka*-AC
	kalozera (“contour ridge”) *m(u)-a*-NC	*ka-galu* (“small dog”) *ka*-AC
	kamtema (“small hawk”) *m(u)-a*-NC	*ka-dyedwe* (“manner of eating”) *ka*-AC
ku-	*kunda* (“African giant rat”) *m(u)-a*-NC	*ku nyumba* (“at/to the house”) *ku*-AC
	kukuku (“grey lourie”) *m(u)-a*-NC	*ku phiri* (“at/to the mountain”) *ku*-AC
	kuka (“girls sleeping-quarters”) *li*-NC	*ku-dya* (“eating”) *ku*-AC
mu-	*munthu* (“person”) *m(u)-a*-NC	*mu-udzu* (“in the grass”) *mu*-AC
	mlungu (“God”) *m(u)-a*-NC	*m-dziko* (“in the country”) *mu*-AC
	mutu (“head”) *u*-NC	*m-kati* (“inside”) *mu*-AC

To make the distinctions illustrated in Table 7.4, the acquirer must learn the difference between the spurious forms on the noun and the functional nature of the phrase-initial forms that characterise the GAC. I assume this distinction is a complication that can be figured out after one has mastered the basic ANC system as outlined above. Therefore, I conclude that the GACs must emerge after the ANC system because acquiring the latter system appears to be logically less complex as learners simply need to notice the regular pairing between

Stellenbosch University https://scholar.sun.ac.za

the world of reference and the grammatical forms. The GAC, by contrast, requires the learner to know the functional status of the affixes and the grammatical relationships that hold between the affixes and the expressions they attach to. It therefore seems that the GACs involve derived notions, that is, concepts that need prior concepts to build on.

A slightly different pattern that acquirers must also notice concerns the distribution of what I have analysed as the 'expletive-like' use of some AMs. I assume that at some stage in the acquisition process, acquirers will notice that in some instances there are no specific agreement features to determine the AMs. Learners must therefore also notice that the systematic doubling pattern breaks down and placeholder AMs must be employed; hence the use of the expletive AM *zi-, ku-* and *pa-* (see again Chapter 6, Section 6.5). I argue that if children do not notice both Saussurean arbitrariness and the systematic departures from Saussurean arbitrariness, it would be difficult to account for how acquirers learn the one-to-one pairing of form and meaning and also where this pattern systematically breaks down. Following this logic, I also argue that the development of all these formal features is dependent on the semantically- and phonologically-based ANC system which I argue is the crucial developmental preliminary of the full agreement system observed in the adult grammar.

In the present section, the acquisition of the Chichewa noun and agreement class system has been characterised as a result of an acquirer noting systematic patterns and postulating rules for the noted patterns. In the next section, I appeal to Yang's (2016) Tolerance Principle to explain how these rules are generalised despite having exceptions to the relevant rules.

7.3.3 A rule-based metric for rule productivity

According to the discussion in Section 7.3.2, the acquisition of the complete Chichewa agreement system is assumed to begin with the developmental preliminaries that include the semantically- and phonologically-based ANC system. Although I have arrived at the various rules by means of corpus evidence, I assume that child acquirers also postulate the same rules from the PLD. Considering that, crosslinguistically, children are observed to be sensitive to both semantic and phonological features from very early stages of linguistic development, I hypothesise that Chichewa child acquirers begin to postulate the animate and phonological noun classification rules at that early stage. Following Yang (2016), I argue that these postulated rules are generalised, and their productivity evaluated in terms of a rule-based statistical metric, the Tolerance Principle (see again Chapter 1, Section 1.3.1).

Stellenbosch University https://scholar.sun.ac.za

According to the Tolerance Principle, a rule, *R* can be deemed productive only if the number of exceptions (*e*) to R are below a certain threshold. This threshold varies with respect to the number of cases that are subject to R. In Chapter 5, Section 5.5, I presented the various productivity calculations of the semantic and phonologically determined noun classification rules in the Chichewa nominal dataset. For reference purposes, I repeat the various rules and their productivity status in Table 7.5.

Table 7.5: Semantic and phonological noun class rule productivity

Noun class	Rule	Number of tokens	Θ_N	Attested Exceptions	Productivity status
***m(u)-a*-ANC**	animate	1222	14% (172)	12% (143)	productive
***i*-ANC**	[+cor], [+ant]	1453	13.7% (199.5)	6.7% (98)	productive
***u*-ANC**	[-cor],[+labial], [+round]	839	14.8% (124.6)	1.1% (10)	productive
***a*-ANC**	[-cor], [+labial], [+back]	357	17% (60)	0% (0)	productive
***chi*-ANC**	[+cor], [-ant]	1075	14.2% (154)	0% (0)	productive

As shown in Table 7.5, the weight of evidence supports the generalisations that I drew from the Chichewa corpus.

In more specific terms, I assume that Chichewa child acquirers notice the animate semantic basis of nouns belonging to the *m(u)-a*-ANC and generalise this pattern as a rule. According to Table 7.5, this animate rule has a 12% exception rate, which is 2% lower than the expected threshold that can offset this rule as productive. Equally, the phonologically-based rules are generalised based on their statistical weight over the observed exceptions to the rules. As shown in Table 7.5, all the phonological rules are productive. However, the exceptions and idiosyncratic forms are part of the grammar of Chichewa. According to Yang (2016:41,216), these forms are assumed to be learned by rote memorisation (cf. Chomsky & Halle 1967:172).

As argued above, the ANC rules in Table 7.5 are hypothesised to form part of the preliminary rules while the set of GACs comprising the highly predictable morphosyntactic processes are thought to form part of the later-arriving items. It is not surprising that the GAC subsystem is highly productive since by the time acquirers begin to master these syntactic rules, the

Stellenbosch University https://scholar.sun.ac.za

quantity of accessible input data - or intake - is considerable. According to the Tolerance Principle, the larger the amount of input data the higher the critical threshold, which would make the learning process harder (Yang, 2016: 109–224). It therefore makes sense that the early acquired ANC system tolerates more exceptions than it is observed in the late-arriving GAC, which are highly regular.

Having illustrated that the acquisition of the Chichewa noun and agreement class system can be accounted for by means of a rule-based metric, I further show that the various aspects of the Chichewa noun class system are hierarchically organised. In the following section, I briefly show how this aligns with general properties of linguistic organisation.

7.4 The hierarchy principle

Another key finding (which is crucially lacking in the Bleek-Meinhof system) regarding the Chichewa ANC system is that every aspect of the system is hierarchically organised. As I have shown in Chapters 4 and 5, the semantic features are manifested in the form of the animacy hierarchy, the phonological features are in the form of *feature geometry*, the entire ANC system is also organised in a hierarchical manner, as shown in the flow diagram in (8). This emerging pattern of the *noun class system* agrees with already-familiar aspects of linguistic structure that are consistently observed to follow hierarchical order, for example, syntactic structure generally, syllable structure, thematic hierarchies, the transitivity hierarchy, the prosodic hierarchy, etc. (see i.a. Keenan & Comrie, 1977; Fox, 1987; Croft, 2003; Becker, 2014: 63). In this regard, the Chichewa noun class system can then be described as a 'well-behaved' grammatical system of the kind familiar from other grammatical systems. This observed hierarchical organisation challenges the traditional views that describe the Bantu noun class system as having lost its previous order, suggesting that it is in a state of disharmony (see again Chapter 2 and the references cited there). Given the success of generation upon generation of Chichewa acquirer in acquiring the noun-class system, it seems that the hierarchical approach may be more explanatory than the traditional one. Therefore, this broader hierarchical symmetry with other grammatical systems is further support to the ANC and GAC systems proposed in the present study.

In the next section, I summarise and conclude the present chapter.

Stellenbosch University https://scholar.sun.ac.za

7.5 Conclusion

The major objective for the present chapter was to provide an analytical discussion of the underlying principles of the Chichewa noun class and agreement systems. I introduced three main principles in Section 7.2, namely the *semantic primacy principle*, the *phonological primacy principle* and the *constituent-edge principle*. In section 7.2.1, I showed that the animacy distinction that underlies the Chichewa ANC system is, cross-linguistically, at the heart of one of the most salient prototypical situations that children perceive. This also allows us to understand the relevance of the animacy distinction in the domain of grammatical encoding. In Section 7.2.2, I discussed the phonological primacy principle. I showed that the phonological features that form the phonological cue set are basic phonemes that are also observed to be primary in the phoneme inventories of many of the world's languages. These phonemes are also observed to be hierarchically organised, as predicted by the theory of feature geometry. The hierarchical order is also found to be similar to the order of phoneme inventory development in many languages. Like the semantic primacy principle, the phonological primacy principle is active at the very early stages of linguistic development. In Section 7.2.3, I discussed the constituent-edge principle, which was also found to be pivotal in grammatical structures and language acquisition crosslinguistically. The constituent-edge principle was found to be relevant in making the phonological features more salient, which aids both language acquisition and processing.

In Section 7.3, I discussed how the three principles interact in the course of language development. I argued that since the animacy features are prototypical and the agreement system is canonical in the syntax of Chichewa, infants first notice the basic word order patterns and how animacy of nouns plays a role in the distribution of NPs. Since they are also, following Biberauer (2017), assumed to be sensitive to systematic departures from the arbitrary form-meaning relations that define natural-language lexicons, they will also notice the "extra" form that is consistently present wherever a Chichewa noun appears. They therefore need to postulate rules that regulate the relationship between NPs and sets of AMs. With respect to the proposed noun class system, I have argued that these rules initially pick up on salient components of the phonological and semantic input that Chichewa-acquiring children receive. That is, Chichewa acquirers' earliest agreement rules harness salient aspects of the components that are more generally paired to give arbitrary form-meaning pairs (or 'words'): the earliest formally marked, or *grammaticalised*, agreement is keyed to relevant phonological and semantic properties. In Biberauer's terms, the earliest

Stellenbosch University https://scholar.sun.ac.za

agreement-related formal features [F]s thus rest on specific [S]- and [P]-features; further [F]-based subdivisions follow. The basic semantically- and phonologically-based ANC thus provides the foundation of the entire agreement system. The ANC system, then, forms the basis for the later-acquired morphologically predictable GACs.

Regarding how children would begin to draw generalisations from the input data to arrive at a system as proposed, I adopted Yang's (2016) rule-based metric. Here, I concluded that since all proposed noun class rules were deemed productive, it could be the same with children acquiring the Chichewa ANC system. For example, if a child acquirer notices that animate-denoting nouns trigger the *m(u)-a-* AMs, it would postulate it as a rule and generalise it beyond the available input data, whereas forms that are not productive are treated as exceptions to productive rules; such non-productive forms are assumed to be learned by rote memorisation.

In Section 7.4, I showed that the hierarchical organisation of each aspect of the *ANC system* fits into the more universal hierarchical order of linguistic structure, a property observed in many other grammatical systems. The conclusion drawn from this discussion is that the Chichewa noun class system is a well organised grammatical system, reflecting productive classes that acquisition metrics like Yang's Tolerance Principle would lead us to expect acquirers to be able to acquire; it should thus not be viewed as a system in any form of disarray, as is often suggested in the traditional literature. The fact that the noun class system is built on more general linguistic principles provides support for the proposed system.

Stellenbosch University https://scholar.sun.ac.za

CHAPTER 8
Summary and conclusion

8.1 Summary

The aims of this study were set out in Chapter 1, against the backdrop of a range of problems found in the existing accounts of the Bantu noun class system. I stated that the Bantu noun class phenomenon is typically accounted for by the traditional framework, identified as the *Bleek-Meinhof system* (Maho, 1999: 3–4; Katamba, 2006: 105). However, the Bleek-Meinhof system is based on a reconstruction of what is assumed to be the Bantu parent language (proto-Bantu), and, as such, it fails to account for several synchronic patterns related to the phenomenon. In the present study, I have argued that, based on our current knowledge of language acquisition, the diachronically based Bleek-Meinhof system does not offer a plausible account of how such a system can be acquired (cf. Lightfoot, 1979, 1991, 1999; Yang, 2016; Biberauer, 2017, 2018a). I observed that the major problem with the Bleek-Meinhof system is methodological, with evidence from Chichewa specifically showing that the system was based on a partial dataset (see also Worsley, 1954: 286 for the more general charge that discussions of Bantu noun classification are based on "casual sampling"). In this regard, my first major aim was to address this empirical gap by presenting a well sampled dataset. Informed by this type of evidence, the second aim was to propose a noun class system that is plausible when viewed from a language acquisition perspective. In line with this objective, I introduced a set of recent theoretical assumptions that would help account for the classification of Bantu nominal expressions, on the one hand, and the mode via which this is acquired, on the other. Generally, I adopted the generative approach to linguistic description and explanation. Specifically, I adopted Biberauer's (2017) Three Factors Model to account for the nature of knowledge of language and how it is acquired (Chomsky 2005). Second, I employed Yang's (2016) Tolerance Principle as a rule-based metric to account for how acquirers postulate rules and draw generalizations from sparse input data. Both Yang's (2016) and Biberauer's (2017) assumptions explain how rule postulation (from apparently sparse data) help to facilitate the acquisition of grammatical structures.

To account for the structure of nominal expressions, I adopted the Distributed Morphology approach, specifically that instantiated in Ferrari (2005). I showed that in this model, nominal expressions are assumed to minimally comprise of a noun stem, which is the result of

Stellenbosch University https://scholar.sun.ac.za

merging a functional category head *n* to a root. However, not all nouns are assumed to be derived productively; some nouns are assumed to form what Lieber (2004) identifies as the simplex lexicon, i.e. nouns that are idiosyncratically memorised.

In Chapter 2, I introduced the kind of facts that inform the previous literature on Bantu noun classification and considered how these facts have been interpreted to arrive at the accounts that are represented in the Bleek-Meinhof system. According to the Bleek-Meinhof view, nouns in Chichewa and Bantu in general are assumed to be made up of at least a prefix and a stem. It is further assumed that the prefix carries noun class and number features. Therefore, the prefix is taken as the key determinant of nominal classification, with the consequence that the Bantu noun class system was developed primarily on the basis of these so-called noun class prefixes (see i.a. Mufwene, 1980; Mugane, 1997: 33; Maho, 1999; Katamba, 2006: 105). Although not all nouns bear prefixes, linguists claim that in Proto Bantu each noun had a prefix and that the current state of affairs is a result of language change (see i.a. Richardson, 1967: 378; Maho, 1999: 63; Katamba, 2006: 116; Demuth & Weschler, 2012: 70). Each prefix was identified by means of serial Arabic numbering system, ranging from 1 to over 23. The numbering system became the standard Bantu-wide nomenclature. Strikingly, in addition to the prefix criterion, there are also claims that the Bantu noun class system is semantically motivated, specifically that the prefixes originally had consistent meanings (Burton & Kirk, 1976; Batibo, 1987; Spitulnik, 1989; Contini-Morava, 1997; Demuth, 2000; Palmer & Woodman, 2000; Selvik, 2001). As a consequence, highly speculative shared semantic properties have been postulated for each class in the literature. In this regard, the diachronic effects are assumed to have affected both the prefix system and their semantic imports. In summary, I observed that the Bantu noun class debate is characterised by overwhelming lack of consensus, especially regarding the precise principles underlying the classification system. It therefore became apparent that the problem of the Bantu noun class system has not been given a definitive answer (see i.a. Worsley, 1954; Amidu, 1997; Maho, 1999; Mchombo, 2004: 3; Katamba, 2006: 119–120). These problems provided motivation for a re-examination of the data on the basis of Chichewa in Chapter 3.

In Chapter 3, I discussed the empirical evidence and the methods followed to obtain the required evidence. To resolve the problems introduced in Chapter 1 and 2, the present study adopted a mixed-methods approach that followed quantitative and qualitative methods in data gathering, processing and analysis. The chapter introduced the computerised dataset

Stellenbosch University https://scholar.sun.ac.za

excerpted from the *Chichewa Monolingual Dictionary*. The data was then classified based on the AMs that each noun triggers, otherwise known as *noun-external distribution properties*. I identified this approach as an *agreement-based noun classification approach*. This is one major departure from the traditional approach that derived the noun classes largely based on *noun-internal distribution properties*, namely noun prefixes. The major difference emerging from these two approaches is that while the traditional approach arrived at over 23 noun classes, some of them bearing the same AMs, in the present study I only identified 12 agreement-based classes. In the first instance, I referred to these classes as *agreement classes* (ACs) rather than noun classes because in some instances the AMs were observed to be triggered by non-nominal elements such as CPs, PPs and in some instances the AMs served expletive functions. The implication of this novel observation is that we need to view the phenomenon under consideration in a broader sense as an *agreement class* problem which subsumes the noun class subsystem.

Having introduced the data and classification approach in Chapter 3, I focused in more detail in Chapter 4 on the properties of the expressions contained in each of the twelve identified ACs. Each AC was found to include expressions with distinct morpho-phonological and semantic characteristics, with some ACs, crucially, encompassing more than one traditional noun class. These ACs were further analysed to fall into two broad categories. On the one hand, the first category included typical nouns, which were in the previous literature associated with primary classification. This first group of ACs has a comparatively high representation in the corpus, with the smallest class containing 357 entries. The second category, on the other hand, is quantitatively poorly represented in the corpus, with the highest containing only 16 entries. This implies that the second group has very few lexicalised items, but consists of morphosyntactically complex and predictable expressions such as evaluative phrases, locative phrases, non-locative PPs, CPs, etc., which were traditionally associated with secondary classification. For both expository and theoretical purposes, I identified the former group as forming the typical *agreement-based noun class* (ANC) system and the latter category as forming the more *general agreement class* (GAC) system. The crucial findings from this wealth of empirical evidence was that no single ANC was made up of nouns with prefixes only; there was always more than one prefix per each class. In this regard, I showed that the prefix does not underlie the classification criterion in these classes but rather, in more general terms, the classification was based on either semantic or word-initial phonological characteristics. The so-called noun class prefixes were

Stellenbosch University https://scholar.sun.ac.za

observed to only play a part in inflectional and derivational processes. Following these observations, I set out to consider in more detail the underlying criteria of the ANC subsystem in Chapter 5 and those of the GAC in Chapter 6.

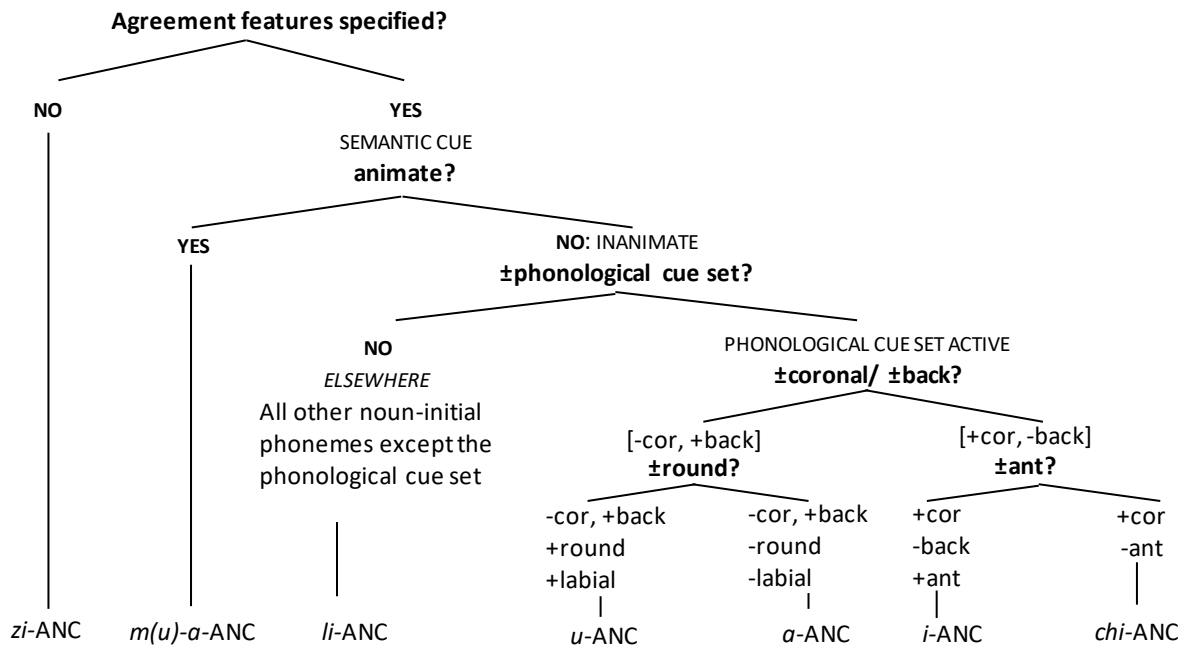
In Chapter 5, I focused on the ACs containing typical nominal expressions, which have been identified properly as ANCs. Here, I conducted a more detailed examination of the noun-internal distributional properties of nouns making up this group. As I observed in Chapter 4, the nouns primarily belong to two super-ANCs on a semantic basis, one animate/agentive and the other non- inanimate/agentive. Following cross-linguistic practice regarding treatment of animacy features, I identified these super-ANC as simply animate and inanimate. The inanimate group was then observed to involve a further subdivision based on word-initial phonological characteristics. A range of phonological features were observed to function as distinctive cues for nouns to belong to their respective phonologically determined ANCs. These were assumed to make up a *phonological cue set*, to which acquirers are sensitive; that is, they are used to establish which class nouns not belonging to the animate group should be assigned to. The key phonological features, example phonemes and proposed ANCs are summarised in (1).

(1)	**Phonological features**	**Example phonemes**	**Noun class**
a.	[+cor], [+ant], [-back]:	N-, i-, e-, mi-, me-	*i*-ANC
b.	[-cor], [+labial], [+round], [+back]:	m-, u-, o-, mu-, mo-	*u*-ANC
c.	[-cor], [+labial], [-round], [+back]:	ma-, a-	*a*-ANC
d.	[+cor], [-ant], [-back]:	ch-	*chi*-ANC

I proposed that the choice of AMs for any of the nouns in the phonologically determined classes depend on the word-initial phonemes corresponding to the groups in (1). Further to this I showed that inanimate nouns that do not begin with any of the phonemes in the phonological cue set form a different Elsewhere *li*-ANC, which can be viewed as the default within the phonologically determined super-ANC. I further observed that in constructions where the agreement trigger is not specified or has conflicting agreement features, the general default AM *zi-* is used, forming an additional more general default *zi*-ANC. In total, I suggested seven basic ANCs which take the following hierarchical structure.

Stellenbosch University https://scholar.sun.ac.za

(2)



As shown in (2), the Chichewa ANC system employs two major criteria, where semantic and phonological sets of cues work together to allocate nouns to the various classes. I also showed that this ANC system has support from crosslinguistic studies where similar noun class systems have also been observed. Further support also comes from observations from previous studies of the Bantu noun class system, where both semantic and phonological criteria were noted, but a proper account of how they interrelate was lacking (see i.a. Mchombo 1978:104; Matiki 2001:76-77; Alcock & Ngorosho, 2004; van der Spuy, 2009: 201).

The major objective in Chapter 6 was to consider the morpho-semantic properties of the relatively more complex but regular expressions. In this Chapter, I focused on a range of expression-types such as evaluative phrases, locative phrases, non-locative PPs, CPs and cases where the AMs function as expletives. These were treated separately because their morpho-semantic characteristics are not typical of nouns forming the ANC system proposed in Chapter 5. Four kinds of evaluative expressions were identified, namely diminutive, augmentative, honorific and attitudinal expressions. Structure-wise, evaluatives were analysed as complex NPs that have an evaluative functional projection above the number phrase of the NP. I showed that the evaluative function in Chichewa is exponed by underspecified affixes, which play derivational and inflectional functions when lower in the nominal spine and evaluative functions when upper in the nominal spine. All these evaluative expressions were observed to trigger AMs based on their word-initial characteristics (which

Stellenbosch University https://scholar.sun.ac.za

happen to be the prefixes); hence exhibiting a phonologically determined agreement system. Evaluative expressions whose prefixes are recycled from derivational and inflectional projections associated with the relevant ANCs do not form different ACs but fall into the existing ANCs, as 'late-arrivals' in terms of acquisition order. These 'late-arrival' expressions include the AUG *chi-*, the HON *a-* and the ATT *u-* which are added to the *chi-*ANC, the *a-*ANC and the *u-*ANC, respectively. However, the DIM expressions do not coincide with any AM in the ANC agreement marker set and hence form two different GACs, namely *ka*-GAC and *ti*-GAC. The *ka*-GAC also includes *ka*-initial manner nominals due to their homophonous word-initial forms, whereas the *ti*-GAC was observed to optionally include two idiosyncratic *ti*-initial nouns, on a phonological basis.

In the second part of Chapter 6, I focused on locative phrases. Although many Bantuists have strongly argued for the nominal view of locatives, I argued for the non-nominal nature of locatives on two fronts, namely properties of the locative prefixes (LOCS) and the syntactic functions of the locative phrases (LOCP). First, following a set of diagnostic tests, LOCS were found to be morpho-semantically different from prefixes associated with nominal expressions. Second, in response to the earlier suggestion that LOCPS can serve as grammatical subjects and objects, I argued that these expressions do not necessarily fulfil these grammatical functions, but that their positions are related to information structure (including discourse functions like Topic and Focus). In addition to this, I also showed that LOCPs behave in a similar manner to non-locative PPs and CPs, which have not been treated as nominal in Bantu studies, as there is no noun class for CPs, PPs on the Bleek-Meinhof schema.

In the third part of Chapter 6, I introduced another agreement pattern not considered in the Bleek-Meinhof, namely that involving PPs, CPs and structures where AMs function as expletives. I argued that just like locative PPs, non-locative PPs and CPs trigger Topic agreement and not subject agreement. I also showed that, in addition to the default AM *zi-*, in Chichewa, there are two other AMs that also play expletive-like functions namely *ku-* and *pa-.* However, unlike *zi-* which is used for missing or conflicting nominal agreement features, the default *ku-* is used to refer to underspecified general location while the default *pa-* is used to refer to underspecified relatively specific location. The AM patterns that we find holding in relation to agreement triggers that do not fall within the Bleek-Meinhoff system and also in cases where there are no agreement triggers really do need to be considered as part of the more general agreement marking picture. Since this agreement system is not

Stellenbosch University https://scholar.sun.ac.za

restricted to nominal expressions, it would be a misnomer to refer to it as forming the noun class. However, unlike the ANC, the GAC underlying criteria were observed to primarily concern the morphological status of the functional prefixes that derive the various expressions. However, in addition to the morphological cues, there are some expressions that also trigger the relevant AMs on a phonological basis, specifically understood in terms of the *cohort effect* (Sekerina & Brooks, 2007). Therefore, the regularities that we observe in this domain are significant because they pick up in very particular ways on the patterns that are evident in the noun class domain - i.e. phonological conditioning.

In Chapter 7, I discussed the underlying principles of the Chichewa noun class system. I proposed three major principles, namely the *semantic primacy principle*, the *phonological primacy principle* and the *constituent-edge principle*. Referring to a wide range of crosslinguistic literature, I showed that all three principles are widely observed to universally underlie many aspects of grammatical structure and of the language acquisition process (Comrie 1989, Becker 2016, McCarthy 1988, Studdert-Kennedy & Goldstein 2003, McCarthy & Prince 1993). For example, I showed that the semantic primacy principle manifests through the notion of animacy, one of the universal salient prototypical situations that children perceive from a very early age. According to the consulted research (i.a. Dahl & Fraurud, 1996; Becker, 2009, 2014, 2015), the early perception of animacy is also observed to play a role in acquisition of grammatical structure such as word order, argument marking, etc. Therefore, it is not surprising to find that the Chichewa ANC system is primarily organised along the notion of animacy.

As regards the phonological primacy principle, I argued that it manifests in the form of what I refer to as a *phonological cue set*, a set of phonemes that determine classification of various nouns in the phonologically determined subsystem (see (1) above). According to the consulted literature, I showed that these phonemes are also observed to be the most salient and form the basic phoneme inventory of many of the world's languages (see i.a. Clements, 1985; McCarthy, 1988; Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003; Gierut *et al.*, 1993: 223; Chen & Kent, 2005). I also showed that the respective phonemes are crosslinguistically observed to be the first acquired and that they are acquired in a hierarchical order with the more broad and simpler coming first before the more specific and complex (Rice & Avery, 1995; Studdert-Kennedy & Goldstein, 2003; Stokes *et al.*, 2005: 825–826). These phonological forms are predicted to take the observed hierarchical order by many other theories (e.g. feature geometry (McCarthy, 1988), Place of Articulation (POA)

Stellenbosch University https://scholar.sun.ac.za

Theory (Chomsky & Halle, 1968) and Articulator Theory (Sagey, 1986; McCarthy, 1988; Mester, 1988) and also by phoneme inventory acquisition studies (Stokes *et al.*, 2005: 825–826)).

Lastly, I considered the *constituent-edge principle* where I showed that many other grammatical structures are also organised with respect to constituent edges (McCarthy & Prince, 1993; Ko, 2014; Benavides-Varela & Mehler, 2015). One important property of the constituent edges in the present context is that it enhances the salience of the phonemes that make up the phonological cue set in both language acquisition and processing. I therefore argued that the word-initial nature of phonologically determined agreement features, and the position of AMs is not a feature that is only specific to Chichewa noun and agreement class system, but is also a more universal grammatical tendency.

I then showed how the three principles work together to support the process of language acquisition. Specifically, I proposed that the animacy feature comes first in structuring the ANC system, especially following the assumptions of basic form-meaning pairing (so-called *Saussurean arbitrariness*) that characterise the basic feature of language (Biberauer, 2017). The basic one-to-one pairing of form and meaning is assumed to be further harnessed to derive the more grammaticalized patterns, so-called *systematic departure from Saussurean arbitrariness* (Biberauer, 2017). Therefore, the more basic arbitrary semantic and phonological aspects of the noun class system form the foundations on which the more regular patterns are built. With these two assumptions in place, I proposed that acquirers take notice of these patterns and postulate the observations as rules and draw generalisations following the rule-based metric proposed by Yang (2016). Having shown in Chapter 5 that all the rules drawn from my corpus data were productive, I predicted that the patterns would be mirrored in child acquirers' input. I therefore argue that the proposed Chichewa noun and agreement class system can be seen to be both learnable and amenable to explanation by Biberauer's (2017) Three Factors Model and Yang's (2016) Tolerance Principle.

In the final part of Chapter 7, I showed that all the underlying principles of the Chichewa noun and agreement class system are observed to exhibit a hierarchical structure, a property observed to hold for many grammatical aspects crosslinguistically. I therefore argue that the proposed model of the Chichewa noun and agreement class system provides a

Stellenbosch University https://scholar.sun.ac.za

plausible account of the phenomenon compared to the diachronically-based Bleek-Meinhof model.

8.2 Conclusion and topics for further research

Many works focusing on the Bantu noun class system have pointed out the various empirical and theoretical challenges with the traditional model of the Bantu noun class system (see i.a. Worsley, 1954: 286–287; Amidu, 1997; Maho, 1999), with others clearly acknowledging that a plausible account has not yet been proposed (i.a. Mchombo 2004: 3). The present study is partly motivated by these observations. I have therefore proposed a different perspective on how we ought to view the Bantu noun class system (which in essence should be viewed as the Bantu agreement system), and it is hoped that the proposals made here may meaningfully contribute to the increase in our understanding of the formal make-up of noun and agreement class systems in Bantu and other language families. By doing this, I have uncovered further problems that need to be investigated. As I have pointed out in various parts of the dissertation, the most pressing one concerns the need for a Chichewa child language corpus. Empirical evidence of this nature will help to determine to what extent the input received by Chichewa-acquiring children reflects the distributions that emerge from my investigation of the Chichewa lexicon. More specifically, what does the child-directed nominal lexicon look like? Noun acquisition study may also deepen our understanding of how phonological and semantic properties feed into the formalisation of noun and agreement class systems. In the same vein, it is also imperative that we investigate in more detail the developmental relationship between the ANC and the GAC systems.

The second important avenue of study is a comparative research within Bantu and also typologically diverse languages that exhibit related noun and agreement class systems. There are many facets that may need to be considered, for example to find out whether other Bantu languages structure their agreement class systems in the manner proposed in the present study, specifically the super-ANC distinction which is not recognised in the Bleek-Meinhof system. The other relevant aspect concerns the phonological cue set: it would be interesting to investigate whether other Bantu languages also employ a related phonological cue set. Since the Bleek-Meinhof system predicts different numbers of noun classes across the Bantu family, it would also be illuminating to find out which of the ANCs are basic or common to all Bantu languages.

Stellenbosch University https://scholar.sun.ac.za

References

Abney, S.P. 1987. The English noun phrase in its sentential aspect. Doctoral dissertation. Massachusetts Institute of Technology.

Adger, D., de Cat, C. & Tsoulas, G. Eds. 2004. *Peripheries: Syntactic edges and their effects*. New York: Kluwer Academic Publishers.

Aikhenvald, A. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, A. 2004. Gender and noun class. In *Morphology: An international handbook on inflection and word-formation*. G. Booij, C. Lehmann, J. Mugdan, S. Skopeteas, & W. Kesselheim, Eds. Berlin & New York: Walter de Gruyter. 1031–1045.

Aissen, J. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*. 21(3):435–483.

Alcock, K.J. & Ngorosho, D. 2004. Interaction between phonological and grammatical processing in single word production in Kiswahili. *Language and Speech*. 47(1):1–30.

Amidu, A.A. 1980. Locative marking and locative choice in Swahili and their semantic and grammatical implications. PhD Thesis. University of London.

Amidu, A.A. 1997. *Classes in Kiswahili*. Köln: Rüdiger Köppe Verlag.

Anderson, S. 1969. West Scandinavian vowel systems and the ordering of phonological rules. PhD thesis. MIT.

Anttila, A. & Fong, V. 2000. The partitive constraint in optimality theory. *Journal of Semantics*. 17:281–314.

Aronoff, M. 1994. Morphology by itself: Stems and inflectional classes. In *Linguistic Inquiry Monographs, 22*. Cambridge, Mass.: MIT Press.

Aronoff, M. 1998. Gender agreement as morphology. In *Proceedings of the First Mediterranean Conference of Morphology*. G. Booij, A. Ralli, & S. Scalise, Eds. Patras. 7–18.

Ashforth, A. 2014. When the vampires come for you: A true story of ordinary horror. *Social Research: An International Quarterly*. 81(4):851–882.

Ashton, E.O., Mulira, E.E.K., Ndawula, E.G.M. & Tucker, A.N. 1954. *A Luganda grammar*. Longmans, Green & Co.: London.

Baerman, M., Corbett, G.G. & Brown, D. Eds. 2010. *Defective paradigms: Missing forms and what they tell us*. Oxford: Oxford University Press.

Baker, M.C. 2001. The natures of nonconfigurationality. In *The handbook of contemporary syntactic theory*. M. Baltin & C. Collins, Eds. Malden, MA: Blackwell Publishers Ltd. 407–438.

Baker, M.C. 2008. *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Baker, C.L. & Brame, M.K. 1972. Global rules: a rejoinder. *Language*. 48:51–75.

Stellenbosch University https://scholar.sun.ac.za

Baldauf Jr, R.B. & Kaplan, R.B. 2004. Language policy & planning in Botswana, Malawi, Mozambique and South Africa: Some common issues. In *Language planning & policy in Africa: Vol. 1: Botswana, Malawi, Mozambique and South Africa*. R.B. Baldauf Jr & R.B. Kaplan, Eds. Clevedon: Multilingual Matters Ltd. 5–20.

Bateman, N. & Polinsky, M. 2010. Romanian as a two-gender language. In *Festschrift for David Perlmutter*. D. Gerdts, J. Moore, & M. Polinsky, Eds. Cambridge Mass.: MIT Press. 41–77.

Bates, G.L. 1926. *Handbook of Bulu*. Elat, Cameroun: The Halsey Memorial Press.

Batibo, H.. 1987. The Bantu ancestors' vision of the world. *Journal of Linguistics and Language in Education*. 2:1–21.

Batibo, H.. & Kgolo, N. 2016. Nasal retention of noun prefixes of classes 9/10 in Bantu languages: A case from Setswana, Kiswahili and Shisukuma. *LASU*. 4(4):21–29.

Bauer, L. 1997. Evaluative morphology: In search of universals. In *Studies in language, volume 21*. B. Comrie & M. Noonan, Eds. Amsterdam: John Benjamins Publishing Company. 533–575.

Bauer, L. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Becker, M. 2009. The role of NP animacy and expletives in verb learning. *Language Acquisition*. 16(4):283–296.

Becker, M. 2014. *The acquisition of syntactic structure: Animacy and thematic alignment*. (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press.

Becker, M. 2015. Animacy and the acquisition of tough adjectives. *Language Acquisition*. 22(1):68–103.

Bell, A. 1972. The development of syllabic nasals in the Bantu noun class prefixes mu-, mi-, and ma-. *Anthropological Linguistics*. 14(2):29–45.

Benavides-Varela, S. & Mehler, J. 2015. Verbal positional memory in 7-month-olds. *Child Development*. 86(1):209–223.

Berlin, B. 1973. Folk systematics in relation to biological classification and nomenclature. *Annual Review of Ecology and Systematics*. 4:259–271.

Bernard, R.H. 2011. *Research methods in anthropology: qualitative and quantitative approaches*. 5th ed. Lanham: Altamira Press.

Berwick, R., Pietroski, P., Yankama, B. & Chomsky, N. 2011. Poverty of the stimulus revisited. *Cognitive Science*. 35(7):1207–1242.

Biberauer, T. 2017. Factors 2 and 3: A principled approach. *Cambridge occasional papers in linguistics*. 10(3):38–65.

Biberauer, T. 2018a. Pro-drop and emergent parameter hierarchies. In *Understanding null subjects: A synchronic and diachronic perspective*. F. Cognola & J. Casalicchio, Eds. Oxford: Oxford University Press. 94–139.

Biberauer, T. 2018b. Less IS More: some thoughts on the Tolerance Principle in the context of the Maximise Minimal Means model. *Cambridge occasional papers in linguistics*. 11(6):131–145.

Stellenbosch University https://scholar.sun.ac.za

Biberauer, T. 2018c. On the significance of the periphery: a 21st century generative perspective. In *WoSSP 15*. Barcelona, 29 June.

Bird, H., Ralph, M.A.., Seidenberg, M.., McClelland, J.. & Patterson, K. 2003. Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory & Language*. 48:502–526.

Bleek, W. 1862. *A comparative grammar of South African languages*. London: Trübner & Co.

Bleek, W. 1869. *A comparative grammar of South African languages. Part 2: The concord. Section 1: The noun*. Cape Town and London: J. C. Juta and Trübner & Co.

de Blois, K. 1970. The augment in Bantu Languages. *Africana Linguistic*. 85–165.

Bloom, L. 1970. *Language development: Form and function in emerging grammars (Research Monograph No. 59)*. Cambridge MA: MIT Press.

Bokamba, E.G. 1979. Inversions as grammatical relation changing rules in Bantu languages. *Studies in the Linguistic Sciences*. 9:1–24.

Bokamba, E.G. 1993. Language variation and change in pervasively multilingual societies: Bantu languages. In *Topics in African linguistics*. S. Mufwene & L. Moshi, Eds. Amsterdam & Philadelphia: John Benjamins Publishing Co. 207–252.

Bostoen, K. & Bastin, Y. 2015. *Bantu lexical reconstruction. Oxford handbooks online*. Oxford: Oxford University Press.

Braver, A. & Bennett, W. 2015. Phonotactics of noun class disambiguation in Xhosa. In *Proceedings of AMP 2015*.

Bresnan, J. 1991. Locative case vs. locative gender. In *Proceedings of the seventh annual meeting of the Berkeley Linguistics Society: General session and parasession on the grammar of event structure*. Vol. 17. L.A. Sutton, C. Johnson, & R. Shields, Eds. 53–68.

Bresnan, J. 1994. Locative inversion and the architecture of Universal Grammar. *Language*. 70(1):72.

Bresnan, J. 1995. Category mismatches. In *Theoretical approaches to African linguistics*. Lawrenceville, NJ: Africa World Press. 19–46.

Bresnan, J. & Kanerva, J.M. 1989. Locative inversion in Chicheŵa: a case study of factorization in grammar. *Linguistic Inquiry*. 20(1):1–50.

Bresnan, J. & Kanerva, J.M. 1992. The thematic hierarchy and locative inversion in UG: A reply to Schachter's comments. In *Syntax and the lexicon*. T. Stowell & E. Wehrli, Eds. New York: Academic Press. 111–125.

Bresnan, J. & Mchombo, S.A. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language*. 63(4):741–782.

Bresnan, J. & Mchombo, S.A. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*. 13(2):181–254.

Stellenbosch University https://scholar.sun.ac.za

Bresnan, J., Cueni, A., Nikitina, T. & Baayen, H.R. 2007. Predicting the dative alternation. In *Cognitive Foundations of Interpretation*. G. Bouma, I. Krämer, & J. Zwarts, Eds. Amsterdam: KNAW. 69–94.

Broselow, E. & McCarthy, J. 1983. A theory of internal reduplication. *The Linguistic Review*. 5:25–88.

Brown, D. & Hippisley, A. 2012. *Network morphology: A defaults-based theory of word structure, volume 133*. Cambridge: Cambridge University Press.

Buell, L.C. 2007. Semantic and formal locatives: Implications for the Bantu locative inversion typology. *SOAS Working Papers in Linguistics*. 15:105–120.

Burton, M. & Kirk, L. 1976. Semantic reality of Bantu noun classes: The Kikuyu case. *Studies in African Linguistics*. 7(2):157–174.

Byamugisha, J., Keet, M. & de Renzi, B. 2018. Pluralizing nouns across agglutinating Bantu languages. In *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA, August 20-26, 2018. 2633–2643.

Caha, P. & Pantcheva, M. 2015. Locatives in Shona and Luganda. In *Workshop on building blocks, Leipzig, 21-22 November 2014*. (Ms. Masarykova univerzita, lingbuzz/002220).

Carstens, V. 1991. The morphology and syntax of determiner phrases in Kiswahili. PhD dissertation, University of California, Los Angeles.

Carstens, V. 1993. On nominal morphology and DP structure. In *Theoretical aspects of Bantu grammar*. S.A. Mchombo, Ed. Stanford, CA: CSLI Publications. 151–180.

Carstens, V. 1994. Locatives in Chichewa. In *Paper presented at the 25th Annual African Linguistics Conference*. Rutgers University.

Carstens, V. 1997. Empty nouns in Bantu locatives. *The Linguistic Review*. 14:361–410.

Carstens, V. 2005. Agree and EPP in Bantu. *Natural Language and Linguistic Theory*. 23:213–279.

Carstens, V. 2008. DP in Bantu and Romance. In *The Bantu-Romance connection: A comparative investigation of verbal agreement, DPs, and information structure*. C. De Cat & K. Demuth, Eds. Amsterdam/Philadelphia: John Benjamins. 131–165.

Cassidy, K.W. & Kelly, M.H. 2001. Children's use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review*. 8(3):519–523.

Chen, L.-M. & Kent, R.D. 2005. Consonant-vowel co-occurrence patterns in Mandarin-learning infants. *J. Child Lang*. 32:507–534.

Childers, J.B. & Echols, C.H. 2004. 2½-year-old children use animacy and syntax to learn a new noun. *Infancy*. 5(1):109–125.

Chinele, J. 2017. *Blood suckers' terror*. [Online], Available: https://www.times.mw/blood-suckers-terror/ [2019, February 01].

Cho, T. & Jun, S.-A. 2000. Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA Working Papers in Phonetics*. 99:57–79.

Stellenbosch University https://scholar.sun.ac.za

Cho, T. & Keating, P.A. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*. 29(2):155–190.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1970. Remarks on nominalization. In *Readings in English Transformational Grammar*. R. Jacobs & P. Rosenbaum, Eds. Waltham, Mass.: Ginn & Company.

Chomsky, N. 1975. *Reflections on language*. New York: Patheon.

Chomsky, N. 1980. *Knowledge of language: Its nature, origin, and use*. New York: Praeger Publishers.

Chomsky, N. 1995. *The Minimalist Program*. (Current studies in linguistics series). Cambridge, Mass.: MIT Press.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry*. 36(1):1–22.

Chomsky, N. 2008. On phases. In *Foundational issues in linguistic theory*. R. Freidin, C.P. Otero, & M.L. Zubizarreta, Eds. Cambridge Mass.: The MIT Press. 132–166.

Chomsky, N. 2015. *The Minimalist Program: 20th anniversary edition*. Cambridge Mass.: The MIT Press.

Chomsky, N. 2017a. The language capacity: architecture and evolution. *Psychonomic Bulletin & Review*. 24(1):200–203.

Chomsky, N. 2017b. Language architecture and its import for evolution. *Neuroscience & Biobehavioral Reviews*. 81:295–300.

Chomsky, N. & Halle, M. 1968. *The sound pattern of English*. New York: Harper & Row.

Chomsky, N. & Lasnik, H. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*. Berlin: Mouton de Gruyter. 506–569.

Choti, J. 2015. Phonological asymmetries of Bantu nasal prefixes. In *Selected proceedings of the 44th annual conference on African linguistics*. R. Kramer, E.C. Zsiga, & O.T. Boyer, Eds. Somerville, MA: Cascadilla Proceedings Project. 37–51.

Chung, S. 1998. *The design of agreement: Evidence from Chamorro*. California: University of Chicago Press.

Chung, S. 2013. The syntactic relations behind agreement. In *Diagnosing syntax*. L.L.-S. Cheng & N. Corver, Eds. Oxford: Oxford University Press. 251–270.

Cilibrasi, L. 2015. Word position effects in speech perception. PhD dissertation. University of Massachussetts, Amherst.

Cinque, G. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Cinque, G. & Rizzi, L. 2008. The cartography of syntactic structures. In *STiL (Studies in Linguistics). CISCL Working Papers on Language and Cognition*. V. Moscati, Ed. University of Siena, Siena. 43–59.

Clements, G.N. 1985. The geometry of phonological features. *Phonology Yearbook*. 2:225–252.

Stellenbosch University https://scholar.sun.ac.za

Clements, G.N. 1991. Place of articulation in consonants and vowels: a unified theory. *Working Papers of the Cornell Phonetics Laboratory*. 5:77–123.

Cole, D.T. 1955. *An introduction to Tswana grammar*. Second ed. Cape Town: Longmans Penguin Southern Africa.

Cole, D.T. 1967. The prefix of Bantu noun class 10. *African Studies*. 26(3):119–138.

Collins, C. 2004. The agreement parameter. In *Triggers*. A. Breitbarth & H. Riemsdijk, Eds. Berlin: Mouton de Gruyter. 115–136.

Comrie, B. 1989. *Language universals and linguistic typology*. 2nd ed. Chicago: The University of Chicago Press.

Comrie, B. & Thompson, S.A. 2007. Lexical nominalization. In *Language typology and syntactic description: Volume 3, grammatical categories and the lexicon*. T. Shopen, Ed. Cambridge: Cambridge University Press. 334–381.

Connelly, M. 1984. Basotho children's acquisition of noun morphology. Doctoral dissertation. University of Essex.

Contini-Morava, E. 1996. 'Things' in a noun-class language. Semantic functions of agreement in Swahili'. In *Toward a calculus of meaning. Studies in markedness, distinctive features and deixis. (Studies in functional and structural linguistics, vol. 43)*. Amsterdam/Philadelphia: John Benjamins. 251–290.

Contini-Morava, E. 1997. Noun classification in Swahili. In *African linguistics at the crossroads: Papers from Kwaluseni*. R.K. Herbert, Ed. Koln: Rudiger Koppe. 599–628.

Contini-Morava, E. 2002. (What) do noun class markers mean? In *Signal, meaning and message: Perspectives on sign-based linguistics*. Amsterdam: John Benjamins. 3–64.

Contini-Morava, E. & Kilarski, M. 2013. Functions of nominal classification. *Language sciences*. 40:263–299.

Corbett, G.G. 1991. *Gender*. Cambridge Mass.: Cambridge University Press.

Corbett, G.G. 2005. Number of genders. In *The world atlas of language structure*. M. Haspelmath, M.S. Dryer, D. Gil, & B. Comrie, Eds. Oxford: Oxford University Press. 126–129.

Corbett, G.G. 2006. *Agreement*. Cambridge, Mass.: Cambridge University Press.

Corbett, G.G. 2013. Number of genders. In *The world atlas of language structures online*. M.S. Dryer & M. Haspelmath, Eds. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Corbett, G.G. & Mtenje, A.D. 1987. Gender agreement in Chichewa. *Studies in African Linguistics*. 18(1):1–38.

Crain, S. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences*. 14(4):597–612.

Croft, W. 2003. *Typology and universals*. 2nd ed. Cambridge: Cambridge University Press.

Culbertson, J., Gagliardi, A. & Smith, K. 2017. Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language*. 92:343–358.

Stellenbosch University https://scholar.sun.ac.za

Culicover, P.W. 1999. *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford: Oxford University Press.

Dahl, Ö. 2000. Animacy and the notion of semantic gender. In *Gender in grammar and cognition*. B. Unterbeck & M. Rissanen, Eds. Berlin: Mouton de Gruyter. 99–116.

Dahl, Ö. & Fraurud, K. 1996. Animacy in grammar and discourse. In *Reference and referent accessibility*. T. Fretheim & J. Gundel, Eds. Amsterdam/Philadelphia: John Benjamins Publishing Company. 47–64.

Daltas, P. 1987. Some patterns of variability in the use of diminutive and augmentative suffixes in spoken Modern Greek Koine (MGK). *Glossologia*. 4:63–88.

Déchaine, R.-M., Girard, R., Mudzingwa, C. & Wiltschko, M. 2014. The internal syntax of Shona class prefixes. *Language Sciences*. 43:18–46.

Demuth, K. 1988. Noun class and agreement in Sesotho acquisition. In *Agreement in natural language: Approaches, theories and descriptions*. M. Barlow & C.A. Ferguson, Eds. Stanford, CA: CSLI Publications. 305–321.

Demuth, K. 2000. Bantu noun class systems: loanword and acquisition evidence of semantic productivity. In *Systems of nominal classification*. G. Senft, Ed. Cambridge: Cambridge University Press. 270–292.

Demuth, K. & Weschler, S. 2012. The acquisition of Sesotho nominal agreement. *Morphology*. 22:67–88.

Denny, J.P. & Creider, C.A. 1986. The semantics of noun classes in Proto-Bantu. In *Noun classes and categorization*. C. Craig, Ed. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Diercks, M. 2011. The morphosyntax of Lubukusu locative inversion and the parameterization of Agree. *Lingua*. 121(5):702–720.

Dimitriadis, A. 1997. Alliterative concord in phonology-free syntax. In *GLOW workshop on the morpho-syntax and phonology of African and Afro-Asiatic languages*. Rabat, Morocco.

Dingemanse, M. 2006. The semantics of Bantu noun classification: A review and comparison of three approaches. Master Thesis. Leiden University.

Dingemanse, M. 2008. *Assibi A. Amidu. 2007. Semantic assignment rules in Bantu classes. Review*. [Online], Available: http://nbn-resolving.de/urn:nbn:de:0009-10-12583 [2017, March 07].

Dixon, R.M.. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Dobrin, L.M. 1995. Theoretical consequences of literal alliterative concord. In *CLS 3 1: Papers from the 31st regional meeting of the Chicago Linguistic Society: I: The main session*. A. Dainora, R. Hemphill, B. Luka, B. Need, & S. Pargman, Eds. Chicago: Chicago Linguistic Society. 127–142.

Stellenbosch University https://scholar.sun.ac.za

Dobrin, L.M. 1998. The morphosyntactic reality of phonological form. In *Yearbook of Morphology 1 997*. G. Booij & J. van Marle, Eds. Dordrecht: Kluwer Academic Publishers. 59–81.

Doke, C.M. 1927. The significance of class 1a of Bantu nouns. In *Festschrift Meinhof*. F. Boas, Ed. Glückstadt and Hamburg: J.J. Augustin. 196–203.

Doke, C.M. 1935. Early Bantu literature: the age of Brusciotto. Johannesburg: Witswatersrand University Press. 87–112.

Doke, C.M. 1954. *The Southern Bantu Languages*. London: Oxford University Press for the International African Institute.

Doke, C.M. 1965. *Textbook of Zulu Grammar*. London: Longmans, Green & Co.

Downing, L.J. 1994. SiSwati verbal reduplication and the theory of generalized alignment. *NELS*. 24:81–95.

Downing, L.J. & Mtenje, A. 2011. Prosodic phrasing of Chichewa relative clauses. *Journal of African Languages and Linguistics*. 32(1):65–112.

Downing, L.J. & Mtenje, A. 2017. *The phonology of Chichewa*. Oxford: Oxford University Press.

Dressler, W.U. & Merlini Barbaresi, L. 1994. *Morphopragmatics. Diminutives and intensifiers in Italian, German and other languages*. Berlin & New York: Mouton de Gruyter.

Dryer, M.S. & Haspelmath, M. Eds. 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Durieux, G. & Gillis, S. 2001. Predicting grammatical classes from phonological cues: An empirical test. In *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition (Vol. 1).* J. Weissenborn & B. Höhle, Eds. Amsterdam: John Benjamins B.V. 189–229.

Elman, J.L. 1993. Learning and development in neural networks: The importance of starting small. *Cognition*. 48(1):71–99.

Embick, D. & Noyer, R. 2007. Distributed morphology and the syntax-morphology interface. In *The Oxford handbook of linguistic interfaces*. G.C. Ramchand & C. Reiss, Eds. Oxford: Oxford University Press. 298–324.

Endress, A., Nespor, M. & Mehler, J. 2009. Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences*. 13(8):348–353.

Erteschik-Shir, N. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.

Evans, N. 1997. Head classes and agreement classes in the Mayali dialect chain. In *Nominal classification in Aboriginal Australia.* M. Harvey & N. Reid, Eds. Amsterdam: John Benjamins. 105–146.

Fanselow, G. & Féry, C. 2002. Ineffability in grammar. In *Resolving conflicts in grammars: Optimality Theory in syntax, morphology, and phonology*. G. Fanselow & C. Féry, Eds. Hamburg: Helmut Buske. 265–307.

Stellenbosch University https://scholar.sun.ac.za

Ferrari-Bridgers, F. 2008. A unified syntactic analysis of Italian and Luganda nouns. In *The Bantu-Romance Connection*. C. De Cat & K. Demuth, Eds. Amsterdam: John Benjamins Publishing Co. 239–258.

Ferrari, F. 2005. A syntactic analysis of the Italian and Luganda nominal systems: How nouns can be formed in the syntax. PhD dissertation. NYU.

Flemming, E. 2003. The relationship between coronal place and vowel backness. *Phonology*. 20:335–373.

Flombaum, J.I. & Santos, L.R. 2005. Rhesus monkeys attribute perceptions to others. *Current Biology*. 15(5):447–452.

Fodor, J.D. & Sakas, W.G. 2017. Learnability. In *The Oxford handbook of universal grammar*. I. Roberts, Ed. Oxford: Oxford University Press.

Foley, W.A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Foley, W.A. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.

Fortune, G. 1955. *An analytical grammar of Shona*. London: Longmans, Green.

Fortune, G. 1970. References of primary and secondary noun prefixes in Zezulu. *African Studies*. 29(2):79–110.

Fortune, G. 1985. *Shona grammatical constructions: Parts 1 and 2*. 3rd ed. Harare: Mercury Press.

Fox, B. 1987. The noun phrase accessibility hierarchy reinterpreted: Subject primacy or the absolutive hypothesis? *Language*. 63(4):856–870.

Fromkin, V.A. 2000. *Linguistics: an introduction to linguistic theory*. Malden, MA: Blackwell Publishers Inc.

Gaeta, L. 2015. Evaluative morphology and sociolinguistic variation. In *Edinburgh handbook of evaluative morphology*. L. Körtvélyessy & N. Grandi, Eds. Edinburgh: Edinburgh University Press. 121–133.

Gagliardi, A. 2012. Input and intake in language acquisition. PhD dissertation. University of Maryland, College Park.

Gagliardi, A. & Lidz, J. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language*. 90(1):58–89.

Garbo, F. Di. 2013. Evaluative morphology and noun classification: a cross-linguistic study of Africa. *Journal of Theoretical Linguistics, 10*. 10:114–137.

Gauton, R. 1999. Locative prefix stacking as an earlier viable locativising strategy in Bantu. In *New dimensions in African linguistics and languages (Trends in African linguistics: 3)*. P.F.A. Kotey, Ed. Asmara, Eritrea: Africa World Press, Inc. 217–232.

Gelman, S. & Opfer, J. 2002. Development of the animate-inanimate distinction. In *Blackwell handbook of child cognitive development*. U. Goswami, Ed. Malden, MA: Blackwell Publishing. 151–166.

Stellenbosch University https://scholar.sun.ac.za

Gervain, J. & Werker, J. 2008. How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*. 2(6):1149–1170.

Gierut, J.A., Cho, M.-H. & Dinnsen, D.A. 1993. Geometric accounts of consonant-vowel interactions in developing systems. *Clininical Linguistics & Phonetics*. 7(3):219–236.

Gierut, J.A., Simmerman, C.L. & Neumann, H.J. 1994. Phonemic structures of delayed phonological systems. *Journal of Child Language*. 21(02):291–316.

Gil, D. 2013. Adjectives without nouns. In *The World Atlas of Language Structures Online*. M.S. Dryer & M. Haspelmath, Eds. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Givón, T. 1971a. Some historical changes in the noun-class system of Bantu, their possible causes and wider implications. In *Papers in African linguistics*. Vol. 1. C.-W. Kim & H. Stalhke, Eds. (Current inquiry into language and linguistics). Edmonton, Alberta: Linguistic Research Inc. 33–53.

Givón, T. 1971b. Historical syntax and synchronic morphology: an archaeologist's field trip. *CLS*. 7:394–415.

Givón, T. 1972. Studies in Chibemba and Bantu Grammar. *Studies in African Linguistics, Supplement #3*.

Givón, T. 1975. Serial verbs and syntactic change: Niger-Congo. In *Word order and word order change*. C.N. Li, Ed. Austin, Tex.: University of Texas Press. 47--112.

Givón, T. 1976. Topic, pronoun and grammatical agreement. In *Subject and topic*. C.N. Li, Ed. New York: Academic Press. 149–188.

Givón, T. 1979. Language typology in Africa: a critical review. *JALL*. 1:199–224.

Golston, C. 1995. Syntax outranks phonology: Evidence from ancient Greek. *Phonology*. 12(3):343–368.

Grandi, N. 2011. Evaluative affixes between inflection and derivation: a typological survey. In *Diminutives and augmentatives in the languages of the world*. P. Štekauer & L. Körtvélyessy, Eds. Lexis 6. 5–26.

Grandi, N. 2015. Evaluative morphology and number/gender. In *Edinburgh handbook of evaluative morphology*. N. Grandi & L. Körtvélyessy, Eds. Edinburgh: Edinburgh University Press. 91–107.

Greenberg, J.H. 1954. Concerning inferences from linguistic to nonlinguistic data. In *Language in Culture*. H. Hoijer, Ed. Chicago: University of Chicago Press. 3–18.

Greenberg, J.H. 1977. Studies in numerical systems I: Double numeral systems. *Studies in African Linguistics, Supplement 7*. 97–104.

Greenberg, J.H. 1978. How does a language acquire gender markers? In *Universals of Human Language vol. III: Word Structure*. J.H. Greenberg, C.A. Ferguson, & E.A. Moravcsik, Eds. Stanford, Calif.: Stanford University Press. 47–82.

Greener, I. 2012. *Designing social research: A guide for the bewildered*. London: SAGE Publications Ltd.

Stellenbosch University https://scholar.sun.ac.za

Gries, S.T. 2009. *Quantitative corpus linguistics with R: A practical introduction*. New York: Routledge Taylor & Francis Group.

Grieser, D. & Kuhl, P.K. 1989. Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*. 25(4):577–588.

Grinevald, C. 2000. A morphosyntactic typology of classifiers. In *Systems of nominal classification*. G. Senft, Ed. Cambridge: Cambridge University Press. 50–92.

Guthrie, M. 1948. Gender, number and person in Bantu languages. *Bulletin of the School of Oriental and African Studies*. 18:545–555.

Guthrie, M. 1967. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages. Vols. 1-4*. Farnborough: Gregg Press.

Guthrie, M. 1971. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages.* Farnborough: Gregg Press.

Haas, M.R. 1972. The expression of the diminutive. In *Studies in linguistics in honor of G.L. Trager*. M.E. Smith, Ed. The Hague: Mouton. 148–152.

Hale, K. 1981. On the position of Warlpiri in a typology of the base. *Indiana University Linguistics Club, Bloomington*.

Hale, K. 1982. Preliminary remarks on configurationality. In *NELS 12*. J. Pustejovsky & P. Sells, Eds. 86–96.

Hale, K. 1983. Warlpiri and the grammar of non-configurational languages. *NLLT*. 1:5–47.

Halle, M. 1997. Distributed morphology: Impoverishment and fission. *MIT Working Papers in Linguistics*. 30:425–449.

Halle, M. & Marantz, A. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20*. K. Hale & S.J. Keyser, Eds. Cambridge MA: The MIT Press. 111–176.

Harding, D.A. 1966. The phonology and morphology of Chinyanja. PhD thesis, University of California.

Hare, B., Call, J., Agnetta, B. & Tomasello, M. 2000. Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*. 59(4):771–785.

Harley, H. 2014. On the identity of roots. *Theoretical Linguistics*. 40:225–276.

Hauser, M.D., Chomsky, N. & Fitch, T.W. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science*. 298:1569–1579.

Hayes, B. 1995. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Heim, J. & Wiltschko, M. 2017. *The complexity of speech acts. Evidence from speech act modifiers*. Unpublished ms: UBC.

Heim, J., Keupdjio, H., Lam Wai-Man, Z., Osa-Gómez, A., Thoma, S. & Wiltschko, M. 2016. Intonation and particles as speech act modifiers: A syntactic analysis. *Studies in Chinese Linguistics*. 37(2):109–129.

Stellenbosch University https://scholar.sun.ac.za

Heine, B. 1976. *A typology of African languages based on the order of meaningful elements*. Berlin: Dietrich Reimer.

Heine, B. 1982. African noun class systems. In *Apprehension: Das sprachliche Erfasse von Gegenstäden*. Tübingen: Gunter Narr Verlag. 189–216.

Heine, B., Claudi, U. & Hunnemeyer, F. 1991. *Grammaticalization*. Chicago: University of Chicago Press.

Hendrikse, A.P. 2001. Semantic polysemy in the Southern Bantu noun class system. In *Polysemy in Cognitive Linguistics*. H. Cuyckens & B.E. Zawanda, Eds. Amsterdam: John Benjamins. 185–212.

Hendrikse, A.P. & Poulos, G. 1992. A Continuum interpretation of the Bantu noun class system. In *African Linguistic Contribution*. D. Gowlet, Ed. Pretoria: Via Afrika. 195–209.

Herbert, R.K. 1978. Morphological re-analysis in Bantu nasal class. *African Studies*. 37(1):125–137.

Herbert, R.K. 1985. Gender systems and semanticity: two case histories from Bantu. In *Historical semantics/historical word-formation*. J. Fisiak, Ed. Berlin, New York: Mouton Publishers. 171–197.

Herbert, R.K. 1991. Patterns in language change, acquisition and dissolution: noun prefixes and concords in Bantu. *Anthropological Linguistics*. 33(2):103–134.

Hetzron, R. 1972. Phonology in syntax. *Journal of Linguistics*. 8:251–265.

Hlungwani, M.C. 2012. Deverbal nominals in Xitsonga. PhD dissertation, Stellenbosch University, South Africa.

Hockett, C.F. 1958. *A course in modern linguistics*. New York: McMillan.

de Hoop, H. & Lamers, M. 2006. Incremental distinguishability of subject and object. In *Case, valency and transitivity*. L. Kulikov, A. Malchukov, & P. de Swart, Eds. Amsterdam/Philadelphia: John Benjamins Publishing Company. 269–287.

Hornstein, N. 2008. *A theory of syntax*. Cambridge: Cambridge University Press.

Hornstein, N. 2018. The Minimalist Program after 25 years. *Annual Review of Linguistics*. 4:49–65.

Hudson-Kam, C.L. & Newport, E.L. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language learning and development*. 1(2):151–195.

Hullquist, C.G. 1988. *Simply Chichewa: A simple yet comprehensive approach to learning and mastering the Chichewa language*. Makwasa, Malawi: Malamulo Publishing House.

Hume, E. 1992. Front vowels, coronal consonants and their interaction in non-linear phonology. PhD dissertation. Cornell.

Hume, E. 1996. Coronal consonant, front vowel parallels in Maltese. *Natural Language & Linguistic Theory*. 14(1):163–203.

Stellenbosch University https://scholar.sun.ac.za

Hyman, L.M. 1980. Reflections on the nasal classes in Bantu. In *Noun classes in the Grassfields Bantu Borderland (Southern California occasional papers in linguistics no. 8)*. 179–210.

Hyman, L.M. & Duranti, A. 1982. On the object relation in Bantu. In *Studies in transitivity (syntax and semantics 15)*. New York: Academic Press. 217–239.

Hyman, L.M. & Katamba, F. 1993. The augment in Luganda. syntax or pragmatics? In *Theoretical aspects of Bantu grammar*. S.A. Mchombo, Ed. Stanford: CLSI Publications. 209–256.

Hyman, L.M. & Mtenje, A. 1999. Prosodic morphology and tone: The case of Chichewa. In *The prosody morphology interface*. R. Kager, H. van der Hulst, & W. Zonneveld, Eds. Cambridge: Cambridge University Press. 90–133.

Idiata, J. 1998. Quelques aspects de l'aquisition de la langue isangu par les enfants. PhD Dissertation, Universite Lumiere Lyon 2.

Igartua, I. & Santazilia, E. 2018. How animacy and natural gender constrain morphological complexity: Evidence from diachrony. *Open Linguistics*. 4:438–452.

Ingram, D. & Ingram, K.D. 2001. A whole-word approach to phonological analysis and intervention. *Language Speech and Hearing Services in Schools*. 32(4):271–283.

Jackendoff, R. 2000. Fodorian modularity and representational modularity. In *Language and the brain: Representation and processing*. Y. Grodzinsky, L.P. Shapiro, & D. Swinney, Eds. San Diego: Academic Press. 3–30.

Janda, R.D. 1982. Diachronic fact vs. synchronic fiction: historical-linguistic evidence against assuming underlying grammatical uniformity for contemporary dialects of the same language. In *Papers from the 5th international conference of historical linguistics*. A. Anders, Ed. Amsterdam: John Benjamins Publishing Co. 149–162.

Jelinek, E. 1984. Empty categories, case, and configurationality. *Natural Language & Linguistic Theory*. 2:39–76.

Johnson, W. & Reimers, P. 2010. *Patterns in child phonology*. Edinburgh: Edinburgh University Press.

Jurafsky, D. 1993. Universals in the semantics of the diminutive. *Berkeley Linguistics Society*. 19:423–436.

Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language*. 3:533–578.

Jusczyk, P.W., Cutler, A. & Redanz, N.J. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development*. 64(3):675–687.

Kamwendo, G.H. 1999. Work in progress: the monolingual Chichewa dictionary project in Malawi. *Malilime: Malawian Journal of Linguistics*. 1:47–56.

Kandybowicz, J. 2009. Embracing edges: Syntactic and phono-syntactic edge sensitivity in Nupe. *Natural Language and Linguistic Theory*. 27:305–344.

Kanerva, J.M. 1990. *Focus and phrasing in Chichewa phonology*. New York: Garland.

Stellenbosch University https://scholar.sun.ac.za

Karmiloff-Smith, A. 1981. *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.

Katamba, F. 2006. Bantu nominal morphology. In *Bantu Languages*. D. Nurse & G. Philippson, Eds. London: Routledge. 103–120.

Kayambazinthu, E. 1998. The language planning situation in Malawi. *Journal of Multilingual and Multicultural Development*. 19(5&6):369–439.

Kayne, R.D. 2004. Here and there. In *Lexique, syntaxe et lexique-grammaire/syntax, lexis & lexicon-grammar: Papers in honor of Maurice Gross*. C. Leclère, E. Laporte, M. Piot, & M. Silberztein, Eds. Amsterdam: John Benjamins Publishing. 253–275.

Keenan, E.L. & Comrie, B. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*. 8(1):63–99.

Kelly, M.H. 1992. Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review*. 99(2):349–364.

Kelly, M.H. 1996. The role of phonology in grammatical category assignments. In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. J.L. Morgan & K. Demuth, Eds. New York: Psychology Press. 249–262.

Kempen, G. & Harbusch, K. 2003. A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In *Multidisciplinary Approaches to Language Production.* T. Penchmann & C. Habel, Eds. Berlin: Mouton de Gruyter. 173–181.

Kgukutli, S. 1994. The semantics of Setswana noun classes. MA thesis. Rand Afrikaans University.

Kidd, C., Piantadosi, S. & Aslin, R. 2014. The Goldilocks Effect in infant auditory attention. *Child Development*. 85(5):1795–1804.

Kilarski, M. 2013. *Nominal classification: A history of its study from the classical period to the present*. Amsterdam: John Benjamins Publishing Company.

Kinyalolo, K. 1986. On accounting for the VS word order in some Bantu languages. MA thesis. University of South Carolina, Columbia.

Kinyalolo, K. 1991. Syntactic dependencies and the Spec-head agreement hypothesis in Kilega. PhD dissertation, University of California, Los Angeles.

Kiparsky, P. 1973. 'Elsewhere' in phonology. In *A Festschrift for Morris Halle*. S. Anderson & P. Kiparsky, Eds. California: Holt, Rinehart & Winston. 93–106.

Kishindo, P. 1985. *Some word formation processes in Chichewa: A case of verb – noun derivation. A paper presented at the 3rd African Language Association of Southern African Conference held at University of Natal, Pietermaritzburg, 10 July 1985*.

Kishindo, P. 1998. Dimunition, augmentation and pejorativeness in Icindali: The semantics of class 5/6, 3/4, 7/8 and 21. *Journal of Humanities*. 12:44–55.

Kishindo, P. 2001. Authority in language: The role of the Chichewa Board (1972-1995) in prescription and standardization of Chichewa. *Journal of Asian and African Studies*. 62:261–283.

Stellenbosch University https://scholar.sun.ac.za

Kisseberth, C. 2006. Makhuwa (P30). In *The Bantu languages*. D. Nurse & G. Philippson, Eds. Nurse, Derek: Routledge. 546–565.

Ko, H. 2014. *Edges in syntax: Scrambling and cyclic linearization*. Oxford: Oxford University Press.

Körtvélyessy, L. 2015. *Evaluative morphology from a cross-linguistic perspective*. Cambridge: Cambridge Scholars Publishing.

Kramer, R. 2015. *The morphosyntax of gender*. Oxford: Oxford University Press.

Kroeger, P. 2005. *Analyzing grammar: An introduction*. Cambridge: Cambridge University Press.

Kumar, R. 2011. *Research methodology: A step-by-step guide for beginners*. 3rd ed. London: SAGE Publications Ltd.

Kunene, E. 1974. Nominalization in Zulu. UCLA papers in syntax 6. In *Approaches to the lexicon*. S.A. Thompson & C. Lord, Eds. Los Angeles: UCLA Department of Linguistics. 107–129.

Kunene, E. 1979. The acquisition of Siswati as a first language: Morphological study with special reference to noun classes and some agreement markers. PhD dissertation, University of California, Los Angeles.

Kuno, S. 1973. *Subject raising in Japanese (Current studies in linguistics series, 3)*. Cambridge MA: MIT Press.

Ladefoged, P. & Maddieson, I. 1996. *The sounds of the world's languages*. Oxford: Blackwell Publishing Ltd.

Lahiri, A. & Evers, V. 1991. Palatalization and coronality. In *Phonetics and phonology, vol. 2: The special status of coronals: Internal and external evidence*. C. Paradis & J.-F. Prunet, Eds. New York: Academic Press, Inc. 79–100.

Lambrecht, K. 1994. *Information structure and sentence form: a theory of topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lamers, M. & de Hoop, H. 2004. Animacy information in human sentence processing: an incremental optimization of interpretation approach. In *Constraint solving and language processing: First International Workshop, CSLP 2004, Roskilde, Denmark, September 1-3, 2004*. H. Christiansen, P.R. Skadhauge, & J. Villadsen, Eds. Berlin/Heidelberg: Springer-Verlag. 158–171.

Legate, J.A. & Yang, C. 2002. Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review*. 19:151–162.

Legerstee, M. 1994. Patterns of 4-month-old infant responses to hidden silent and sounding people and objects. *Early Development and Parenting*. 3:71–80.

Legerstee, M., Pomerleau, A., Malcuit, G. & Feider, H. 1987. The development of infants' responses to people and a doll: Implications for research in communication. *Infant Behavior and Development*. 10:81–95.

Stellenbosch University https://scholar.sun.ac.za

Lempert, H. 1989. Animacy constraints on preschool children's acquisition of syntax. *Child Development*. 60(1):237–245.

Levelt, W.J.. 1989. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Levy, Y. 1988. On the early learning of formal grammatical systems: evidence from studies of the acquisition of gender and countability. *Journal of Child Language*. 15(1):179–187.

Lewis, P.M., Gary, S.F. & Femig, C.D. 2015. *Ethnologue: Languages of the world*. 18th ed. Dallas,Texas. [Online], Available: https://www.ethnologue.com/ [2017, February 08].

Li, C.N. & Thompson, S.A. 1976. Subject and topic: A new typology of language. In *Subject and topic*. C.N. Li, Ed. New York: Academic Press. 457–489.

Lidz, J. & Gagliardi, A. 2015. How nature meets nurture: universal grammar and statistical learning. *Annual Review of Linguistics*. 1:333–353.

Lieber, R. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lightfoot, D.W. 1979. *Principles of diachronic syntax*. Cambridge: Cambridge University Press.

Lightfoot, D.W. 1991. *How to set parameters: arguments from language change*. Cambridge MA: MIT Press.

Lightfoot, D.W. 1999. *The development of language: acquisition, change and evolution.* Oxford: Blackwell Publishers Ltd.

Lombard, D.P. 1985. *Introduction to the grammar of Northern Sotho*. Pretoria: J.L. van Schaik.

Loogman, A. 1965. *Swahili grammar and syntax*. D.U. Press, Ed. Pittsburgh, PA: Duquesne University Press.

Lyster, R. 2006. Predictability in French gender attribution: A corpus analysis. *Journal of French Language Studies*. 16(1):69–92.

Maho, J.F. 1999. *A comparative study of Bantu noun classes*. Gothenburg: Acta Universitatis Gothoburgensis.

Mak, W.M., Vonk, W. & Schriefers, H. 2002. The influence of animacy on relative clause processing. *Journal of Memory and Language*. 47:50–60.

Mandler, J. & McDonough, L. 1993. Concept formation in infancy. *Cognitive Development*. 8:291–318.

Maratsos, M. 1983. Some current issues in the study of the acquisition of grammar. In *Handbook of child psychology, vol. 3*. P. Mussen, J.H. Glavell, & E.M. Markman, Eds. New York: Wiley. 709–777.

Maratsos, M. & Chalkley, M.A. 1980. The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In *Children's languages, vol. 2*. K. Nelson, Ed. New York: Gardner Press. 127–214.

Stellenbosch University https://scholar.sun.ac.za

Marchand, H. 1969. *The categories and types of present-day English word formation. A synchronic-diachronic approach*. Munich: Beck.

Marten, L. 2010. The great siSwati locative shift. In *Continuity and change in grammar*. A. Breitbarth, C. Lucas, S. Watts, & D. Willis, Eds. Amsterdam/Philadelphia: John Benjamins. 249–268.

Marten, L. & Kula, N.C. 2012. Object marking and morphosyntactic variation in Bantu. *Southern African Linguistics and Applied Language Studies*. 30(2):237–253.

Martin, J.B. 1986. Null elements in syntax and their ability to head govern. MA thesis, University of California, Los Angeles.

Marvin, T. 2002. Topics in the stress and syntax of words. Doctoral Dissertation. Massachusetts Institute of Technology.

Matiki, A.J. 2000. A functional category of adjectives in Chichewa and Chiyao. *Journal of Humanities*. 14:48–62.

Matiki, A.J. 2001. On type frequency, loanwords, and noun classification in Chichewa. *Malilime: Malawian Journal of Linguistics*. 2:63–87.

Matisoff, J. 1991. The mother of all morphemes: Augmentatives and diminutives in areal and universal perspective. In *Papers from the first annual meeting of the Southeast Asian Linguistics Society*. M. Ratliff & E. Schiller, Eds. Tempe: AZ. 293–349.

McCarthy, J. 1988. Feature geometry and dependency: a review. *Phonetica*. 45:84–108.

McCarthy, J. & Prince, A. 1993. Generalized alignment. In *Yearbook of morphology*. Vol. 12. G. Booij & J. van Marle, Eds. Dordrecht: Springer. 79–153.

McCarthy, J. & Prince, A. 1996. *Prosodic morphology 1986*. Vol. 13. Linguistics Department Faculty Publication Series, University of Massachusetts Amherst.

McEnery, T. & Hardie, A. 2012. *Corpus linguistics*. Cambridge: Cambridge University Press.

Mchombo, S. 1998. Chichewa (Bantu). In *The handbook of morphology*. A. Spencer & A. Zwicky, Eds. Oxford and Malden,Massachusetts: Blackwell Publishers Ltd. 500–520.

Mchombo, S.A. 1978. A critical appraisal of the place of derivational morphology within transformational grammar, considered with primary reference to Chichewa and Swahili. PhD Thesis. SOAS, University of London.

Mchombo, S.A. 1993. On the binding of the reflexive and the reciprocal in Chichewa. In *Theoretical aspects of Bantu grammar*. S.A. Mchombo, Ed. California: CLSI Publications. 181–207.

Mchombo, S.A. 1999. Argument structure and verbal morphology in Chichewa. *Malilime: Malawian Journal of Linguistics*. 1:57–75.

Mchombo, S.A. 2004. *The syntax of Chichewa*. Cambridge: Cambridge University Press.

Mchombo, S.A. 2006. Linear order constraints on split NPS in Chichewa. *ZAS Papers in Linguistics*. 36:143–160.

Stellenbosch University https://scholar.sun.ac.za

Mchombo, S.A. 2007. Argument binding and morphology in Chichewa. In *Texas linguistic society IX: The morphosyntax of underrepresented languages*. F. Hoyt, N. Seifert, A. Teodorescu, & J. White, Eds. Stanford, California: CSLI Online Publications. 203–221.

Meeussen, A.E. 1967. Bantu grammatical reconstructions. *ALS*. 3:80–122.

Meeussen, A.E. 1973. Test cases for method. *ALS*. 14:6–18.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. 1988. A precursor of language acquisition in young infants. *Cognition*. 29(2):143–178.

Meinhof, C. 1899. *Grundriss einer Lautlehre der Bantusprachen, Leipzig: F. A. Brockhaus. Revised edn. 1910*. Berlin: Dietrich Reimer.

Meinhof, C. 1906. *Grundzüge einer vergleichenden Grammatik der Bantusprachen*. Berlin: Reimer.

Meinhof, C. & Van Warmelo, N.J. 1932. *Introduction to the phonology of the Bantu languages*. Berlin: Dietrich Reimer.

Melissaropoulou, D. & Ralli, A. 2008. Headedness in diminutive formation: Evidence from modern Greek and its dialectal variation. *Acta Linguistica Hungarica*. 55(1–2):183–204.

Mester, A. 1988. *Studies in tier structure*. 1st ed. New York: Routledge Taylor & Francis Group.

Mithun, M. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.

Mkochi, W. 2009. Bimoraic word minimality conditions in Chitonga: OT analysis. *SKY Journal of Linguistics*. 22:277–285.

Mletshe, L. 2017. Deverbal nominals derived from intransitive state verbs in isiXhosa: A generative lexicon approach. *South African Journal of African Languages*. 37(1):29–39.

Mohanan, K.P. & Mohanan, T. 2003. Universal and language particular constraints in OT-LFG. In *Proceedings of the LFG03 Conference University at Albany, State University of New York*. M. Butt & T.H. King, Eds. Stanford, Calif.: CSLI Publications. 290–306.

Molina, M., van de Walle, G., Condry, K. & Spelke, E. 2004. The animate-inanimate distinction in infancy: Developing sensitivity to constraints on human actions. *Journal of Cognition and Development*. 5(4):399–426.

Morimoto, Y. 2006. Agreement properties and word order in comparative Bantu. In *Papers in Bantu grammar and description (ZAS Papers in Linguistics 43)*. L.J. Downing, S. Zerbian, & L. Marten, Eds. Berlin: ZAS. 161–187.

Morimoto, Y. 2008. From topic to subject marking: Implications for a theory of subject marking. In *Differential subject marking*. H. De Hoop & P. De Swart, Eds. Dordrecht: Springer. 199–221.

Mould, M. 1974. The syntax and semantics of the initial vowel in Luganda. In *Proceedings of the third annual conference on African linguistics*. E. Voeltz, Ed. Bloomington: Indiana University Press. 223–230.

Stellenbosch University https://scholar.sun.ac.za

Moxley, J.L. 1998. Semantic structure of Bantu noun classes. In *Language history and linguistic description in Africa*. I. Maddieson & T.J. Hinnebusch, Eds. Trenton, NJ and Asmara: Africa World Press Inc. 229–238.

Msaka, P.K. 2003. A cognitive semantic analysis of Chichewa noun classes. Undergraduate research paper. Chancellor College, University of Malawi.

Msaka, P.K. 2015. A minimalist analysis of obligatory reflexivity in Chichewa. MA thesis, Stellenbosch University, South Africa.

Mtenje, A.D. 1985. Arguments for an autosegmental analysis of Chichewa vowel harmony. *Lingua*. 66:21–52.

Mtenje, A.D. 2002. An optimality-theoretic account of Ciyao verbal reduplication. In *Trends in African Linguistics 5*. J. Mugane, Ed. Africa World Press.

Mtenje, A.D. 2007. On recent trends in phonology: vowel sequence in Bantu Languages. *SOAS Working Papers in Linguistics*. 15:33–48.

Mufwene, S. 1980. Bantu class prefixes: Inflectional or derivational? In *Papers from the 16th Regional Meeting Chicago Linguistics Society*. J. Kreiman & A.E. Ojeda, Eds. Chicago: CLS. 246–258.

Mugane, J. 1997. Bantu nominalization structures. Ph.D. dissertation. University of Arizona.

Mulford, R. 1985. Comprehension of Icelandic pronoun gender: Semantic vs. formal factors. *Journal of Child Language*. 12:443–453.

Mutaka, N. & Hyman, L.M. 1990. Syllables and morpheme integrity in Kinande reduplication. *Phonology*. 7:73–120.

Myers. 1987. Tone and the structure of words in Shona. PhD dissertation, The University of Massachusetts, Amhers.

Nankwenya, I.A. 1992. *Zofunika Mu Galamala ya Chichewa*. Revised ed. Blantyre: Dzuka Publishing Company Limited.

Ndayiragije, J., Nikiema, E. & Bhatt, P. 2012. The augment in Kirundi: When syntax meets phonology. In *Selected Proceedings of the 42nd Annual Conference on African Linguistics*. M.R. Marlo, N.B. Adams, C.R. Green, M. Morrison, & T.M. Purvis, Eds. Somerville, MA: Cascadilla Proceedings Project. 112–121.

Nespor, M., Peña, M. & Mehler, J. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*. 2:221–247.

Newport, E.L. 1990. Maturational constraints on language learning. *Cognitive Science*. 14(1):11–28.

Ngoboka, J.P. 2017. Locative markers in Kinyarwanda as determiners. *Nordic Journal of African Studies*. 26(4):292–317.

van Nice, K.Y. & Dietrich, R. 2003. Animacy effects in language production: From mental model to formulator. In *Mediating between concepts and grammar*. H. Härtl & H. Tappe, Eds. Berlin: Mouton de Gruyter. 101–117.

Nichols, J. 1986. Head-marking and dependent-marking grammar. *Language*. 62(1):56–119.

Stellenbosch University https://scholar.sun.ac.za

Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.

Nkemnji, M. 1995. On adjectives that show up as nouns. In *Trends in African linguistics 1: Theoretical approaches to African linguistics*. A. Akinlabi, Ed. Trenton, NJ: Africa World Press, Inc. 147–165.

Nyondo, E. 2013. Bitter side of kaunjika. *Malawi Nation*. 25 July.

Okhotina, I.V. 1975. Some structural changes in Bantu languages due to their specific communicative functions. In *Patterns in language, culture and society: Sub-Saharan Africa*. R.K. Herbert, Ed. Columbus: Ohio State University. 91–92.

Orr, G.J. 1987. Aspects of the second language acquisition of Chichewa noun class morphology. PhD thesis. UCLA.

Øvrelid, L. 2004. Disambiguation of grammatical functions in Norwegian: Modeling variation in word order interpretations conditioned by animacy and definiteness. In *Proceedings of the 20th Scandinavian Conference of Linguistics, Helsinki, January 7–9, 2004*. F. Karlsson, Ed. Department of General Linguistics: University of Helsinki Publications.

Paas, S. 2017. *Oxford Chichewa-English/English-Chichewa Dictionary*. Cape Town: Oxford University Press.

Pae, H.K., Schanding, B., Kwon, Y.-J. & Lee, Y.-W. 2014. Animacy effect and language specificity: Judgment of unaccusative verbs by Korean learners of English as a foreign language. *Journal of Psycholinguistic Research*. 43(2):187–207.

Palmer, G. & Woodman, C. 2000. 'Ontological classifiers as polycentric categories, as seen in Shona class 3 nouns'. In *Explorations in Linguistics Relativity*. M. Pütz & M. Verspoor, Eds. Amsterdam/Philadelphia: John Benjamins. 225–249.

Panagiotidis, P. 2015. *Categorial features: A generative theory of word class categories*. Cambridge: Cambridge University Press.

Payne, T.E. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.

Perfors, A., Tenenbaum, J.B. & Regier, T. 2006. Poverty of the stimulus? A rational approach. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Picardi, C.A. & Masick, K.D. 2014. *Research methods: Designing and conducting research with a real-world focus*. London: SAGE Publications Ltd.

Pinker, S. 1984. *Language learnability and language development*. Cambridge MA: Harvard University Press.

Pinker, S. 1996. *Language learnability and language development*. Cambridge MA: Harvard University Press.

Du Plessis, J.A. 1982a. Sentential infinitives and nominal infinitives. *South African Journal of African Languages*. 2(1):1–18.

Du Plessis, J.A. 1982b. The analysis of the infinitive. *South African Journal of African Languages*. 2(2):18–47.

Stellenbosch University https://scholar.sun.ac.za

Du Plessis, J. & Visser, M.W. 1992. *Xhosa syntax*. Bellville: Via Afrika.

Plowright, D. 2012. *Using mixed methods: frameworks for an integrated methodology*. Los Angeles, London: SAGE Publications Ltd.

Pollard, C. & Sag, I.A. 1994. *Phrase structure grammar*. Chicago: The University of Chicago Press.

Polomé, E. 1967. *Swahili language handbook*. Washington DC: Centre for Applied Linguistics.

Ponsonnet, M. 2018. A preliminary typology of emotional connotations in morphological diminutives and augmentatives. In *Morphology and emotions across the world's languages*. M. Ponsonnet & M. Vuillermet, Eds. Amsterdam: John Benjamins Publishing Company. 17–50.

Postal, P. 1969. Anaphoric islands. In *CLS 5*. R.I. Binnick, A. Davison, G.M. Green, & J.L. Morgan, Eds. Chicago: CLS.

Poulos, G. & Bosch, S. 1997. *Zulu*. Munchen: Lincom Europa.

Poulos, G. & Msimang, C.T. 1998. *A linguistic study of Zulu*. Cape Town: Via Afrika Limited.

Preminger, O. 2014. *Agreement and its failures*. S.J. Keyser, Ed. Cambridge, Mass.: The MIT Press.

Price, T. 1958. *The elements of Nyanja for English-speaking students*. 2nd ed. Blantyre: Church of Scotland Mission.

Prieto, V.M. 2005. Spanish evaluative morphology: Pragmatic, sociolinguistic, and semantic issues. PhD Dissertation. University of Florida.

Prieto, V.M. 2015. The semantics of evaluative morphology. In *Edinburgh handbook of evaluative morphology*. N. Grandi & L. Körtvélyessy, Eds. Edinburgh: Edinburgh University Press. 21–31.

Pullum, G.K. & Scholz, B. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review*. 19:9–50.

Pullum, G.K. & Zwicky, A.M. 1988. The syntax-phonology interface. In *Linguistics: The Cambridge survey: Volume 1, linguistic theory: Foundations*. Cambridge, Mass.: Cambridge University Press. 255–280.

Rakison, D. & Poulin-Dubois, D. 2001. Developmental origin of the animate–inanimate distinction. *Psychological Bulletin*. 127:209–228.

Rauh, G. 2010. *Syntactic categories: Their identification and description in linguistic theories*. Oxford: Oxford University Press.

Renfrew, C.E. 1966. Persistence of the open syllable in defective articulation. *Journal of Speech and Hearing Disorders*. 31:370–373.

Rice, K. & Avery, P. 1995. Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In *Phonological acquisition and phonological theory*. J. Archibald, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 23–42.

Richards, N. 2016. *Contiguity theory*. Cambridge, MA: MIT Press.

Stellenbosch University https://scholar.sun.ac.za

Richardson, I. 1957. *Linguistic survey of the northern Bantu borderland, volume 2*. London: Oxford University Press for the International African Institute.

Richardson, I. 1967. Linguistic evolution and Bantu noun class systems. In *La Classification Nominale dans les Langues Négro-Africaines*. Aix-en-Provence: Éditions du Centre National de la Recherche Scientifique. 373–390.

Ritter, E. & Rosen, S.T. 2010. Animacy in Blackfoot: Implications for event structure and clause structure. In *Syntax, lexical semantics and event structure*. M.R. Hovav, E. Doron, & I. Sichel, Eds. New York: Oxford University Press. 124–152.

Ritter, E. & Wiltschko, M. 2009. Varieties of infl: tense, location, and person. In *Alternatives to cartography*. J. van Craenenbroeck, Ed. Berlin & New York: Mouton de Gruyter. 153–202.

Roca, I. & Johnshon, W. 1999. *A course in phonology*. Malden, MA: Blackwell Publishing.

Sagey, E.C. 1986. The representation of features and relations in nonlinear phonology. Thesis. Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy.

Salzmann, M.D. 2004. Theoretical approaches to locative inversion. MA thesis, University of Zurich, Switzerland.

Salzmann, M.D. 2011. Towards a typology of locative inversion - Bantu, perhaps Chinese and English—but beyond? *Language and Linguistics Compass*. 5/4:169–189.

Sande, H. 2015. An interface model of phonologically determined agreement. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*. K. Kim, P. Umbal, T. Block, Q. Chan, T. Cheng, K. Finney, M. Katz, S. Nickel-Thompson, & L. Shorten, Eds. Somerville, MA: Cascadilla Proceedings Project. 339–350.

Sande, H. 2016. An interface model of phonologically determined agreement. In *Proceedings of the 33rd West Coast Conference on Formal Linguistics*. K. Kim, Ed. Somerville, MA: Cascadilla Proceedings Project.

Sande, H. 2017. Distributing morphologically conditioned phonology: Three case studies from Guébie. PhD dissertation. University of California, Berkeley.

Saussure, F. de. 1966. *Course in general linguistics*. C. Bally, A. Sechehaye, & A. Riedlinger, Eds.Translated by Wade Baskin. New York: McGraw-Hill Book Company.

Scalise, S. 1986. *Generative morphology*. Dordrecht – Holland: Foris Publications.

Schachter, P. 1992. Comments on Bresnan and Kanerva's "Locative inversion in Chichewa: A case study of factorization in grammar. In *Syntax and semantics: Syntax and the lexicon 26*. T. Stowell & W. Wehrli, Eds. San Diego: Academic Press, Inc. 103–110.

Schadeberg, T.C. 1990. *A sketch of Umbundu*. W.J.G. Möhlig & B. Heine, Eds. Köln: Rüdiger Köppe Verlag.

Schadeberg, T.C. 2001. Number in Swahili grammar. *Swahili Forum VIII. Afrikanistische Arbeitspapiere*. 68:7–16.

Scheer, T. 2011. *A guide to morphosyntax-phonology interface theories: How extra-phonological information is treated in phonology since Trubetzkoy's Grenzsignale*. Berlin, New York: De Gruyter Mouton.

Stellenbosch University https://scholar.sun.ac.za

Schenker, A.M. 1955. Gender categories in Polish. *Language*. 31:402–408.

Sekerina, I.A. & Brooks, P.J. 2007. Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*. 98(1):20–45.

Selvik, K. 1996. Setswana noun classes: Conceptual categories marked by grammar? Cand. philol. Thesis. University of Oslo, Oslo.

Selvik, K. 2001. When a dance resembles a tree: a polysemy analysis of three Setswana noun classes. In *Polysemy in cognitive linguistics: Selected papers from the Fifth International Cognitive Linguistics Conference*. H. Cuykens & B. Zawanda, Eds. Amsterdam/Philadelphia: John Benjamins. 161–184.

Senft, G. 2007. Nominal classification. In *The Oxford handbook of cognitive linguistics*. D. Geeraerts & H. Cuyckens, Eds. Oxford. 676–996.

Simango, S.R. 2006. Verb agreement and the syntax of ciNsenga relative clauses. *Southern African Linguistics and Applied Language Studies*. 24(3):277–290.

Singleton, J.L. & Newport, E.L. 2004. When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology*. 49(4):370–407.

Slobin, D. 1981. The origins of grammatical encoding of events. In *The child's construction o.f language*. W. Deutsch, Ed. New York: Academic Press. 185–199.

Smith, J.L. 2005. *Phonological augmentation in prominent positions*. New York: Routledge.

Smith, N. & Tsimpli, I.-M. 1995. *The mind of a savant: Language learning and modularity*. Oxford: Blackwell Publishing.

Smoczyńska, M. 1985. The acquisition of Polish. In *Crosslinguistic study of language acquisition, Vol. 1*. D.I. Slobin, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 595–686.

Spitulnik, D.A. 1987. *Semantic superstructuring and infrastructuring: Nominal class struggle in ChiBemba*. Bloomington, Indiana: Indiana University Linguistics Club.

Spitulnik, D.A. 1989. Levels of semantic structuring in Bantu noun classification. In *Current Approaches to African Linguistics (Vol. 5)*. Newman & Botre, Eds. Dordrecht – Holland: Foris Publications.

Sproat, R. 1985. On deriving the lexicon. PhD. dissertation. MIT.

van der Spuy, A. 2006. Wordhood in Zulu. *Southern African Linguistics and Applied Language Studies*. 24(3):311–329.

van der Spuy, A. 2009. Zulu noun affixes: a generative account. *South African Journal of African Languages*. 29(2):195–215.

van der Spuy, A. 2014. The morphology of the Zulu locative. *Transactions of the Philological Society*. 112(1):61–79.

Steele, S. 1978. Word order variation: a typology study. In *Universals of Human Language, IV: Syntax*. J.H. Greenberg, C.A. Ferguson, & E.A. Moravcsik, Eds. Stanford: Stanford University Press. 585–623.

Stellenbosch University https://scholar.sun.ac.za

Štekauer, P. 2015. Word-Formation processes in evaluative morphology. In *Edinburgh Handbook of Evaluative Morphology*. L. Körtvélyessy & N. Grandi, Eds. Edinburgh: Edinburgh University Press. 43–60.

Steriopolo, O. 2008. Form and function of expressive morphology: A case study of Russian. PhD Thesis. The University of British Columbia.

Steriopolo, O. 2009. Form and function of expressive morphology: A case study of Russian. *Russian Language Journal / Русский язык*. 59:149–194.

Steriopolo, O. 2013. Diminutive affixes in the number domain: A syntactic variation. *Questions and Answers in Linguistics*. 1(2):33–56.

Steriopolo, O. 2014. Parameters of variation in the syntax of homophones. *Poljarnyj vestnik: Norwegian Journal of Slavic Studies*. 17:46–73.

Stoel-Gammon, C. 1985. Phonetic inventories, 15-24 months: A longitudinal study. *Journal of Speech and Hearing Research*. 28:505–512.

Stoel-Gammon, C. & Herrington, P.B. 1990. Vowel systems of normally developing and phonologically disordered children. *Clinical Linguistics & Phonetics*. 4(2):145–160.

Stoel-Gammon, C. & Pollock, K. 2008. Vowel development and disorders. In *The handbook of cinical linguistics*. M.J. Ball, M.R. Perkins, N. Müller, & S. Howard, Eds. Malden, MA: Blackwell Publishing. 525–548.

Stokes, S., Klee, T., Carson, C. & Carson, D. 2005. A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of speech, language, and hearing research : JSLHR*. 48:817–833.

Stowe, L. 1989. Thematic structures and sentence comprehension. In *Linguistic structure in language processing*. G. Carlson & M.K. Tanenhaus, Eds. Dordrecht: Kluwer Academic Publishers. 319–357.

Studdert-Kennedy, M. & Goldstein, L. 2003. Launching language: The gestural origin of discrete infinity. In *Language evolution*. M.H. Christiansen & S. Kirby, Eds. New York: Oxford University Press. 235–254.

Stump, G. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.

Suzman, S. 1980. Acquisition of the noun class system in Zulu. *Papers and Reports on Child Language Development.* 19:45–52.

Svenonius, P. 2006. The emergence of axial parts. In *Nordlyd, special issue on adpositions: Tromsø working papers in linguistics no. 33*. P. Svenonius & M. Pantcheva, Eds. Tromsø: CASTL, University of Tromsø. 1–22.

de Swart, P., Lamers, M. & Lestrade, S. 2008. Animacy, argument structure, and argument encoding. *Lingua*. 118(2):131–140.

Swingley, D., Pinto, J.P. & Fernald, A. 1999. Continuous processing in word recognition at 24 months. *Cognition*. 71(2):73–108.

Taraldsen, K.T. 2010. The nanosyntax of Nguni noun class prefixes and concords. *Lingua*. 120(6):1522–1548.

Stellenbosch University https://scholar.sun.ac.za

Taylor, J. 2007. Things and places: The case of the Zulu locatives. *Language Matters*. 38(1):105–131.

Tegey, H. 1975. The interaction of syntactic and phonological processes: examples from Pashto. In *PCLS II*. P.R. Clyne, F. Hanks, William, & C.L. Hofbauen, Eds. 571–582.

Tettamanti, M., Alkadhi, H., Moro, A., Perani, D., Kollias, S. & Weniger, D. 2002. Neural correlates for the acquisition of natural language syntax. *NeuroImage*. 17(2):700–709.

Thwala, N. 2006. Parameters of variation & complement licensing in Bantu. *ZAS Papers in Linguistics*. 43:209–232.

Trithart, M.L. 1977. *Relational grammar and Chichewa subjectivization rules*. Bloomington: Indiana University Linguistics Club.

Tseng, Y.-C. 2008. Branching consistency as a syntactic OCP constraint to Hakka relative construction. In *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*. M. Chan & H. Kang, Eds. Columbus, Ohio: The Ohio State University. 861–873.

Tsimpli, I.-M. 2014. Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*. 4(3):283–313.

Tsonope, J. 1987. The acquisition of Setswana noun class and agreement morphology, with reference to demonstratives and possessives. PhD dissertation, State University of New York at Buffalo,.

Vail, L.H. 1971. The noun class of Tumbuka. *African Studies*. 30:35–59.

Visser, M. 1989. The syntax of the infinitive in Xhosa. *South African Journal of African Languages*. 9(4):154–185.

Voorhoeve, J. 1980. Noun classes in Adere. In *Noun classes in the Grassfields Bantu Borderland*. L.M. Hyman, Ed. Los Angeles: University of Southern California. SCOPIL 8. 57–72.

van der Wal, J. 2015. Bantu syntax. In *Oxford Handbooks Online*. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-50 [2018, April 18].

Ward, G., Sproat, R. & McKoon, G. 1991. A pragmatic analysis of so-called anaphoric islands. *Language*. 67:439–474.

Watkins, M.H. 1937. A grammar of Chichewa: A Bantu Language of British Central Africa. *Language*. 13(2):5–158.

Weckerly, J. & Kutas, M. 1999. An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*. 36:559–570.

Welmers, W.E. 1971. The typology of the proto-Niger-Kordofanian noun class system. In *Papers in African linguistics*. K. Chin-Wu & S.F. Herbert, Eds. Champaign and Edmonton: Linguistics Research, Inc. 1–16.

Welmers, W.E. 1973. *African Language Structures*. Berkeley: UCP.

Stellenbosch University https://scholar.sun.ac.za

Wiltschko, M. 2017. Ergative constellations in the structure of speech acts. In *The Oxford Handbook of Ergativity*. J. Coon, D. Massam, & L.D. Travis, Eds. Online Publication.

Wiltschko, M. & Steriopolo, O. 2007. Parameters of variation in the syntax of diminutives. In *Actes du congrès annuel de l'Association canadienne de linguistique 2007. Proceedings of the 2007 annual conference of the Canadian Linguistic Association*.

Worsley, P. 1954. Noun classification in Australian and Bantu: Formal or semantic? *Oceania*. 24:275–288.

Yamamoto, M. 1999. *Animacy and reference: A cognitive approach to corpus linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Yamamura, S. 2010. Development of adjectives used as in the history of English. *English Linguistics*. 27(2):344–363.

Yang, C. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. Cambridge, Mass.: The MIT Press.

Yang, C., Crain, S., Berwick, R.C., Chomsky, N. & Bolhuis, J.J. 2017. The growth of language: Universal Grammar, experience, and principles of computation. *Neuroscience and Biobehavioral Reviews*. 81:103–119.

Zaliznjak, A.A. 1964. K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom russkom jazyke. *Voprosy Jazykoznanija*. 4:25–40.

Zawanda, B. & Ngcobo, M. 2008. A cognitive and corpus-linguistic re-analysis of the acquisition of the Zulu noun class system. *Language Matters: Studies in the Languages of Africa*. 39(2):316–331.

Zeller, J. 2017. Locatives in Bantu. Unpublished manuscript, University of KwaZulu-Natal, Durban.

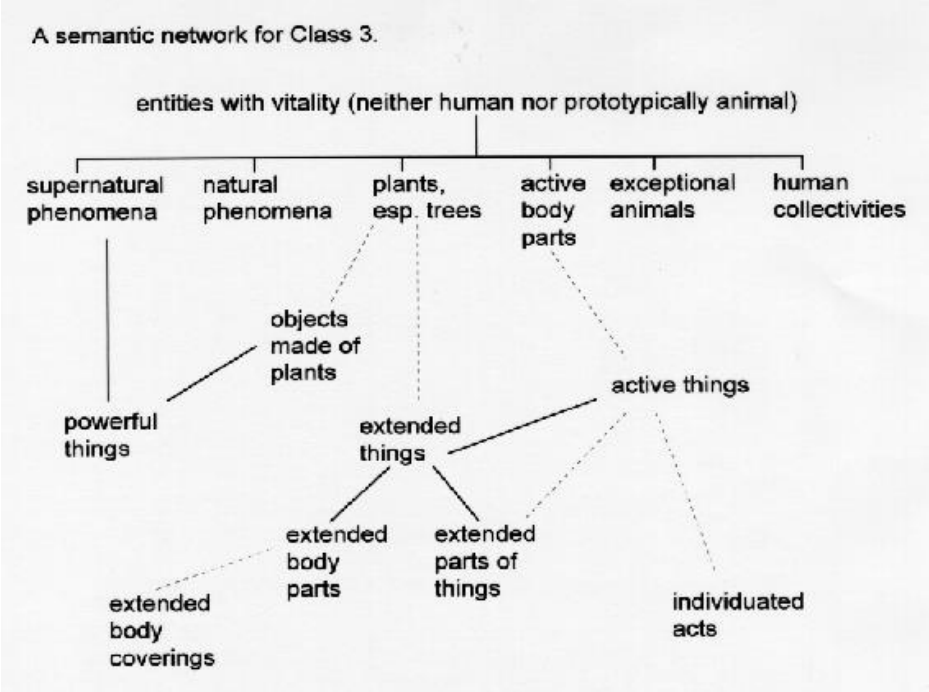
Ziervogel, D. & Mabuza, E.J. 1976. *A grammar of the Swati language*. Pretoria: Van Schaik.

Stellenbosch University https://scholar.sun.ac.za

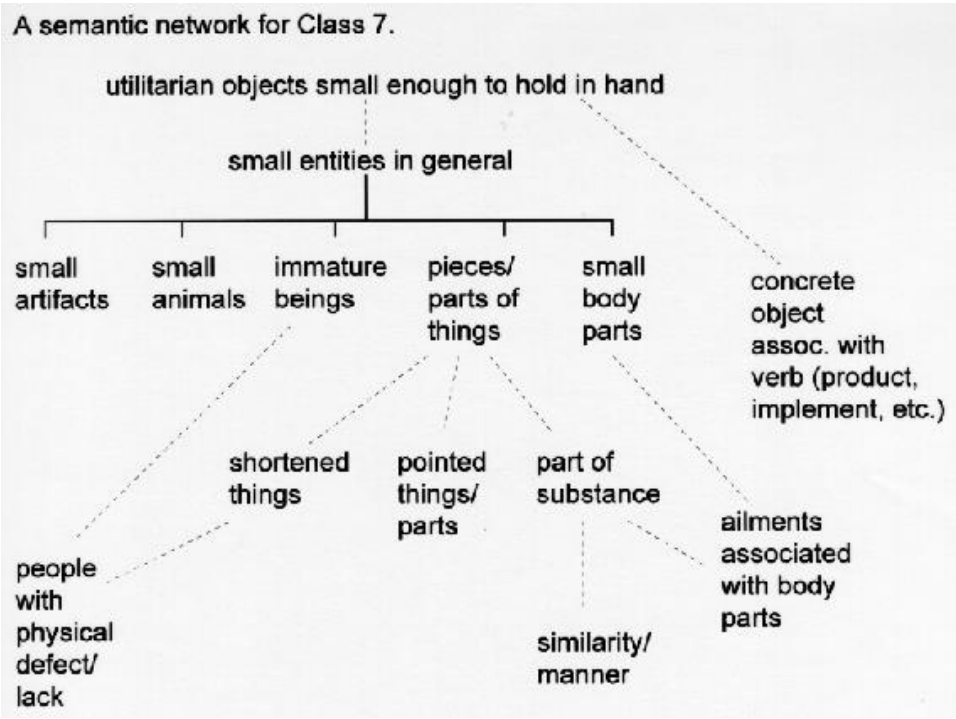
Appendices

Appendix 1

Appendix 1A: Contini-Morava's semantic network analysis for NC3 in Kiswahili



Appendix 1B: Contini-Morava's semantic network analysis for NC7 in Kiswahili



Stellenbosch University https://scholar.sun.ac.za

Appendix 2: Type 1 nouns in the m(u)-a-AC (≈NC1)

1	bwanthulira	mouse/bird trap
2	bwaponde	peanut butter
3	finye	land frog
4	fukizi	beggar
5	fuse	corn weevil
6	galauza	new ridges for planting
7	gogoda	high-heeled shoe/ stiletto
8	gogodera	type of old rifle
9	gogomole	a kind of woodpecker
10	goli	goalkeeper
11	gong'oli	unintelligent person, stubborn person
12	gong'ontha	woodpecker
13	gonthi	deaf person
14	gowe	kind of parrot
15	gowo	kind of hoe
16	gulule	a childish person
17	gunda	Nyau dance's sixth drum
18	gundani	house rat
19	gwirize	successor/representative
20	konya	string made from the bark of a tree
21	konye	cramp/ spasm, convulsion
22	kuwe	grey lourie
23	kwakwase	a character in Nyau dance
24	kwinimbira	a type of mask dancer
25	lende	swing/dangling
26	nyenga	liar/deceiver/mongoose
27	pemphe	kind of small bird which looks like a francolin
28	pote	putty
29	sosoke	hen that has no feathers on its neck
30	tambalale	kind of children's game
31	tambe	witch/sorcerer
32	tsipe	type of grass
33	tsokwe	dry maize seeds stripped from the cob ready for pounding
34	tsope	small kind of wild foxglove
35	zonde	frog with long legs

Appendix 3: Type 2 nouns derived via the prefix *ka-* in the m(u)-a-AC (≈NC1)

1	kababa	tick
2	kabali	hasp
3	kabanga	type of liquor, courageous person
4	kabawi	Wahlberg's eagle
5	kabawu	a type of insect
6	kabementha	type of Nyau cult mask character
7	kabengere	unripe pumpkin
8	kabichi	cabbage
9	kabili	language

Stellenbosch University https://scholar.sun.ac.za

10	kabo	strong man
11	kabondololo	ankle
12	kaboni	carbon
13	kabwanda	fat mouse
14	kachasu	type of locally distilled liquor
15	kachawa	earlobe
16	kachenzi	lesser cane rat
17	kacherenthundu	type of kachere tree
18	kachikhutu	type of mouse
19	kachikhweru	type of sweet potato
20	kachimbo	pipe for smoking, cigar
21	kachipapa	type of Nyau cult mask character, type of winnower
22	kachisi	temple, church, shrine
23	kachiti	type of three-legged stool
24	kachoso	sunbird
25	kachukulu	decayed malted grain, perfumed powder
26	kadainswa	type of dimple
27	kadale	Combretum zeyheri
28	kadaulo	expert
29	kaderuderu	type of beetle
30	kadingoli	kind of gourd
31	kadiri	spy, detective, gossip, albino
32	kadukuta	type of antelope
33	kadyaakolo	type of fish
34	kadyali	a type of plant with edible sweet roots
35	kadyamkwichi	dormouse
36	kadyamlamu	dormouse
37	kadyankhadze	type of Nyau cult mask character
38	kadyole	bootlicker, sharp-witted person
39	kadyolidyoli	spy, detective, gossip, albino
40	kadzanjatema	failing to cut something repeatedly
41	kadzaza	funnel
42	kadzidzi	owl
43	kadzotche	a big bee that does not bite, a worm with irritant hairs
44	kafadongo	snouted beetle
45	kafaluvi	snouted beetle
46	kafaluvu	snouted beetle
47	kafambira	chimney, ship funnel
48	kafucheche	cow's second stomach, appendix
49	kafuchefuche	borer-beetle, weevil
50	kafulansana	fat mouse
51	kafumbwefumbwe	corn weevil
52	kafwenge	maize rotten due to overstay
53	kafwitsa	stomach, kind of butterfly
54	kagolo	kind of tree, kind of maize
55	kagonda	loincloth

Stellenbosch University https://scholar.sun.ac.za

56	kagongono	elbow
57	kagwinthi	small unripe pumpkin
58	kajintchi	pneumonia
59	kakazi	eagle, hen with feathers like an eagle
60	kakhadwe	birth scar
61	kakhutagona	easy chair
62	kakhwerekhwere	hen with stunted growth
63	kakolo	ankle
64	kakolwa	ankle
65	kakombo	Nile cabbage
66	kakozi	kite
67	kakukuku	type of rat
68	kakwiyo	cheetah
69	kalakalaka	tissue of flesh at the lower part of the tongue
70	kalakaswe	type of worm
71	kalaliki	clerk
72	kalambulabwalo	harbinger, team that plays before the main team
73	kalanje	float
74	kalasawene	small thorny plant
75	kalenjeka	collection, flock
76	kalicheche	chicken with stunted growth
77	kaligande	an arrow's trajectory
78	kaligo	type of musical instrument
79	kaligondo	slender mongoose
80	kalijosolo	cricket
81	kalikongwe	slender mongoose
82	kalikwikwiti	type of snake
83	kalimandolo	thick black cloth
84	kalimba	type of musical instrument
85	kalimbwi	boil, abscess, pimple
86	kalim'khundu	chief's representative
87	kalim'tsinde	chief's representative
88	kalindelinde	hesitation
89	kalingumbwa	type of snake
90	kalirombe	chameleon
91	kalisewe	type of mango
92	kalitushu	cartridge case
93	kaliwo	clay smoking pipe
94	kaliwondewonde	AIDS
95	kalizelize	centipede
96	kalonga	head of the Phiri clan, deity name, banana variety
97	kalulu	hare
98	kalulumkuluwabwalo	a kind of Nyau cult mask character
99	kalumbalumba	tick
100	kalumbelumbe	foot sore
101	kalungusese	creeping plant

Stellenbosch University https://scholar.sun.ac.za

102	kalungwe	type of plant
103	kama	bed, nursery
104	kamanya	type of insect, pumpkin variety
105	kamasanza	mushroom variety
106	kamba	tortoise
107	kamba	food
108	kambaapudza	type of tortoise
109	kambakamba	type of gourd
110	kambalame	small bird
111	kam'bale	birthmark, spot or deformity which a person is born with
112	kam'bambe	headband, headcloth
113	kamberembere	cunning person, crook
114	kambinimbini	worker bee
115	kambirinji	dwarf, sufferer from agenesis
116	kambowa	armed robber, gangster
117	kambuku	leopard
118	kambulanje	dwarf
119	kambulumbulu	dangerous person, fierce animal
120	kambulumbunde	cheetah
121	kambuni	type of chicken
122	kambuzi	type of fish
123	kambwi	water cabbage, water lettuce
124	kambwinda	type of mouse
125	kambwiri	old hoe
126	kambwita	hen with short legs
127	kambzidyo	canary
128	kamchacha	type of Nyau cult mask character
129	kamchikope	type of sweet potato
130	kamchiputu	small hawk, Eulophia cucullata
131	kamdawe	beer tasted to check its quality
132	kamfalansa	type of insect
133	kamgodya	packed meal while travelling
134	kamgwamba	eagle
135	kamkukute	type of owl
136	kamnjerezi	dragonfly
137	kamnyata	cunning person, crook
138	kampango	type of fish, snake bean
139	kampapalala	touchy person, slender, tall & thin person
140	kamphambe	a hen with feathers like an eagle
141	kamphimbi	dwarf
142	kamphondo	a small thorny plant (Amaranthus spinosus)
143	kamtang'a	shorts
144	kamtayitsemphanga	role model
145	kamtazi	type of bug
146	kamtinga	bottle/container with a sling
147	kamting'indi	thin &short person

Stellenbosch University https://scholar.sun.ac.za

148	kamtolagwidi	witch who dies after taking witch-hunting concoction
149	kamtsenenge	type of die, the biggest skittle in nsikwa game
150	kamundi	bush baby
151	kamunthu	small person
152	kamuonga	token of appreciation to a bride's parents
153	kamvamgugu	species of mouse
154	kamvulumvulu	whirlwind, gale
155	kamwachale	type of pot/pitcher
156	kamwana	small child, young of an animal, foetus
157	kamwanakhule	small heap of maize left in the granary
158	kamwendom'njira	mobile person
159	kanamkalimbwe	boil, abscess
160	kanamng'wang'wa	type of ant
161	kanamtsonje	canine tooth
162	kanamwali thengo	beautiful fresh girl
163	kanamwalicheche	type of shrub
164	kanamzerenga	potsherd, broken pot fragment
165	kanankhuwe	type of fish, swallow, one who always strolls
166	kandeya	a bee's nest full of honey
167	kandunde	a type of gourd for storing/drinking beer
168	kangaude	spider, cobweb
169	kangombwa	leopard
170	kangulengule	hornbill
171	kangwambe	type of bird
172	kang'wing'wi	type of nyau
173	kanise	little finger
174	kanjama	clever animal that warns others in danger
175	kanjanda	fishnet
176	kanjedza	Phoenix reclinata
177	kanjerenjere	quick ripening maize
178	kanjinji	boil, abscess, pimple
179	kanjipiti	dwarf, sufferer from agenesis
180	kanjunga	dining chair
181	kankhande	buffalo thorn
182	kankhoma	klipspringer (kind of antelope)
183	kansang'azi	blackjack seedlings
184	kansanthi	type of snake
185	kansatsi	Eumetia cervina
186	kanserekete	type of mango
187	kansilanga	type of chicken
188	kantchindwi	wild date palm
189	kanthing'anthi	firefly
190	kanthona	tangerine/mandarin orange
191	kanthulula	type of mango
192	kanthusha	type of termites
193	kanungu	hedgehog

Stellenbosch University https://scholar.sun.ac.za

194	kanyambwebwe	type of grasshopper
195	kanyamenyame	type of plant
196	kanyamphande	maize flour
197	kanyendera	various kinds of mushroom
198	kanyeragudu	scorpion
199	kanyimbi	zorille
200	kapachi	a variety of small & long fish
201	kapado	domestic pig, swine
202	kapalala	cucumber-like plant
203	kapalaliza	one who takes things without the owner's permission
204	kapalepale	time/process of weeding plants
205	kapamba	spleen
206	kapambiro	spleen
207	kapandamneni	phrase
208	kapandaphata	type of plant
209	kape	strainer, sieve, fool
210	kapechepeche	corn weevil
211	kapeni	blade of sprouting grass, maize cob for a budding cob
212	kapeta	fat mouse
213	kapetera	roof of round house
214	kaphadambo	mushroom variety
215	kapherephere	type of creeping plant
216	kaphikaphula	hectic activity, industrious wife
217	kaphirikwete	mound of soil to be used with planting crop
218	kapichi	door fastener
219	kapirikoni	deceitful person, traitor
220	kapirimwana	clitoris
221	kapitagwirire	thorn of a thorn tree that works like a hook
222	kapitawo	guard, captain, supervisor
223	kapolo	slave, servant
224	kapombodza	aged person
225	kapondabwino	a dangerous spot (e.g. because of murderers and ferocious animals)
226	kapondo	armed robber, gangster
227	kapsala	deceiver, crook, robber
228	kapuchi	crop pest, stalk borer
229	kapuku	type of mouse
230	kapunthabuye	poor/miserable/wretched person
231	kapwayi	gadfly, biting fly of cattle
232	karuweruwe	type of fish, swallow, one who always strolls
233	kasambala	calf, young cow
234	kasambwe	viper, adder
235	kasekese	stunted domestic animal, mischievous person
236	kasemasema	a hoe-shaped axe, adze
237	kasenye	grysbok
238	kashole	one who sets fire to dead wood to clear the bush
239	kashukushuku	elbow

Stellenbosch University https://scholar.sun.ac.za

240	kashusha	bat
241	kasokolowe	Uapaca nitida
242	kasolowondo	ankle
243	kasongo	cucumber-like vegetable
244	kasozo	downpour, goods carried by a balancing on a pole
245	kasumbu	leg calf
246	kasusha	type of bat
247	katakwe	clever/wise person, champion
248	katala	pitcher for bathing
249	katanda	ladder inside a granary
250	katangale	corruption
251	katawa	hamerkop/hammerhead stork
252	katayamadzi	last pole of the roof of veranda
253	katchenjeneko	clever/wise/intelligent person
254	katcheretchere	chicken with stunted growth
255	katenthem'tsidya	waterbuck
256	kathusi	type of termites
257	kathuve	an arrow's trajectory
258	kathuvu	moving very fast
259	katikwi	type of gourd
260	katolotolo	pole on top of round roof
261	katswiri	expert, champion, hero
262	katulutulu	fingernail infection, whitlow
263	katumba	leg calf muscle
264	katumbu	otter
265	katundu	goods, luggage
266	katundulu	turtle dove
267	katungwa	kind of pigeon
268	katungwe	swing
269	kaulakhokho	pumpkin variety
270	kaulatsitsi	earthen jar, beer brewing place
271	kaundula	census
272	kaungo	last pole of the roof of veranda
273	kautsa	first witness in a court case
274	kauwuzi	type of baboon
275	kavalatsetse	burglar
276	kavalo	horse
277	kavumba	whirlwind
278	kavumbu	one who reveals
279	kavuwa	prayer hut
280	kavuwevuwe	one who causes hatred, gossip
281	kawangwi	a hen with feathers like an eagle
282	kaweruweru	kind of fish/bird
283	kawinini	Mondia whitei
284	kawolowolo	type of shrub
285	kayera	haricot bean

Stellenbosch University https://scholar.sun.ac.za

286	kayimbi	type of cucumber
287	kazimoto	miniskirt

Appendix 4: Type 2 nouns derived via the prefix *na-* in the *m(u)-a-*AC (≈NC1)

1	nabiri	sore
2	nabuta	kind of chicken with short legs
3	nachigonkhiro	Type of lizard
4	nachinundu	paradise whydah bird
5	nachiula	old toothless elephant
6	nachumwe	tall poles used as building material
7	nadzambwe	large bee that does not sting
8	nadzuche	large bee that does not sting
9	nafadala	snouted beetle
10	nafungwe	Acacia macrothyrsa
11	nafwaka	type of hardwood tree
12	nagoya	type of large skittle
13	nakalongo	main root of a maize plant
14	nakambe	rattle/fruit of msekese (camel's foot) tree
15	nakanyimbwi	dribble of saliva/ mucus, mouth uncleanness after meal
16	nakapeneka	fragment of a broken pot
17	nakapumbo	type of frog
18	nakhodzwe	waterbuck
19	nakhonyere	loincloth
20	nakonako	spy/detective
21	nakwiaka	skilled catcher of objects thrown towards them
22	nalifuku	period when the rains have stopped
23	nalikoma	hornbill
24	nalikukuti	type of small snake
25	nalinsi	pin
26	naliti	needle
27	naliwole	maize that has begun to rot/ arrow poison
28	namadzi	noise made by swimming children
29	namakhoswe	chief representative
30	namalindi	type of bird with long beak and legs
31	namalowe	vibration/echo
32	namasamba	western or eastern green snake
33	nambera	small drum
34	nambumbira	wasp -like insect
35	namdembe	large skittle in a sikwa game
36	namdumu	large navel
37	namdzadzalangwe	red bishop bird
38	namdzambwe	large bee that does not sting
39	namdzikambe	chameleon
40	namdzikwi	a type of bird (pin tailed whydah bird)
41	namdzumbulumbu	large bee that does not sting
42	namgodo	porridge prepared from beans

Stellenbosch University https://scholar.sun.ac.za

43	namgogoda	a type of old rifle
44	namgoma	roasted maize
45	namgomba	woven basket made from reeds
46	namgumula	incision of two or three lines made above the nose
47	namkakala	type of fish that swims sideways
48	namkalimbwi	boil
49	namkandiwu	scorpion
50	namkaputu	flappet lark
51	namkawisi	cob of green maize
52	namkungwi	counsellor
53	namkwantha	type of beer brewed and consumed the same day
54	namkwichi	pin-tailed whydah bird
55	namndaya	maize flour
56	namng'omba	ground hornbill
57	namntchichi	large owl
58	namondwe	storm
59	namowe	kind of large mushroom
60	namphangala	kind of tree
61	namphwaphwa	octopus cabbage tree
62	nampsipsi	kind of small bat
63	namtadza	root
64	namtibule	pounded green maize
65	namtindi	multitude, crowd, scores
66	namtsadzande	sunbird
67	namwali	virgin/initiate
68	namwalinyala	plant whose leaves shrink when touched
69	namweza	crow
70	namwiengeza	type of creeping plant
71	namwimbi	big housefly
72	namwino	nurse
73	namwiyo	green spotted dove
74	namyengeza	sponge
75	namyimbwi	dried dribble of saliva or mucus of the face
76	namzambala	kind of tree
77	namzambwe	large bee that does not sting
78	namzengwe	widow bird
79	namzeze	swallow
80	nanchidwe	type frog
81	nanchinundu	paradise whydah bird
82	nandaya	pounded green maize
83	nandere	a spinning stick used in playing the game
84	nandiula	spring of water in a well
85	nandolo	pigeon peas, Cajanus cajan
86	nandombe	kind of fish
87	nanduwinduwi	flufftail
88	nangalire	beetle

Stellenbosch University https://scholar.sun.ac.za

89	nangananga	mousebird
90	nango	wild duck
91	nangwape	small axe
92	nanjerezi	a kind of fly
93	nantchengwa	hammerhead stork
94	nanyati	spotted beans
95	nanyembwenyembwe	a dribble of dried saliva or mucus on the face
96	nanyongo	clitoris
97	napantha	influenza, flu
98	naphini	kind of tree/ kind of maize with white and yellow seeds
99	naphulundwe	sorghum that has dropped its seeds
100	napweri	albino
101	nathando	okra/ ladies' finger
102	natserenga	potsherd
103	nazale	nursery
104	nazimbuzimbu	a sore

Appendix 5: Nouns derived via the prefix sa- found in the m(u)-a-AC (≈NC1)

1	sabwira	person who places a corpse in a grave
2	sachoka	a type of incision
3	sadyang'ona	kind of fish (Synodontis nyassae)
4	saguga	various cichlid fish (Otophrynx, etc.)
5	sakhuta	glutton
6	sakwata	old man who is not married
7	salumphanjira	a mouse with long lips and bad smell
8	samvamkunkhu	brave and resilient person
9	samwamowa	bottle-nosed fish
10	saonandege	pig
11	sapangwa	someone who does not take advice
12	satha	endless
13	sathamagazi	spleen
14	savala	snail without shell, slug
15	salowam'balani	marijuana

Appendix 6: *ma-* derived nouns in the *m(u)-a-*AC (≈NC1)

1	madomola	wasp
2	maiye	kind of banana
3	makanja	Nyau dancer who performs on stilts
4	make	mother of
5	makechulu	mother/ queen ant
6	makekhutu	eardrum, kind of big termite
7	malasankhuli	kind of worm
8	malemu	late, deceased, dead person
9	malindi	gossip
10	maluku	shy person

Stellenbosch University https://scholar.sun.ac.za

11	malume	maternal uncle
12	mambala	untrustworthy person
13	manakule	A kind of bird
14	mang'ani	buttonquail
15	manthu	mother ant, woman who has given birth to many
16	mapondera	a type of witchdoctor
17	masalamusi	magic tricks
18	masana	noon, afternoon, daylight
19	masikini	poor person, beggar, street child
20	mataya	affluent person
21	matsakamula	rainmaker

Appendix 7: Inanimate nouns in the *m(u)-a-*AC (≈NC1)

Appendix 7A: The plants set

1	bale	burley
2	bandawe	a kind of maize
3	bononGwe	wild spinach
4	bulugamu	bluegum, eucalyptus
5	bwembanyani	a tree with fruits like beans
6	dzanje	mushroom
7	fipe	grass (Beckeropsis uniseta)
8	gombamluzu	kind of large mushroom
9	gondolosi	plant for sexual desire arousal (Mondia whitei)
10	gonkho	sorghum/millet with small white grains
11	gumu	maize cob before developing grains
12	gwafa	guava
13	khadigo	pumpkin flowers & unripe pumpkin relish
14	khobo	kind of tree (Commiphora africana)
15	khobwe	cow pea
16	khocho	Commiphora africana
17	kholobowa	variety of large bananas
18	kholowa	sweet potato leaf
19	khome	grass for making fish traps
20	khonje	sisal, Agave sisalana
21	khuwo	first rain for planting crops, kind of tree
22	khwanya	bean leaves
23	khwisa	nettle
24	koko	coconut fruit
25	longolwe	variety of red mushroom
26	lungululu	fresh beautiful vegetation that appear just before rainy season
27	lunguzi	mysore thorn
28	ombwe	hairy beans with poisonous leaves that some use for killing fish
29	phingo	African ebony
30	phwetekere	small species of tomato

Stellenbosch University https://scholar.sun.ac.za

31	pichesi	peach
32	pupa	maize
33	puwa	maize
34	pwidzi	disease of swollen cheeks/ mumps
35	tanje	pumpkin
36	thikinya	yellow maize
37	thingo	hairy shrub
38	thombozi	kind of tree (Diplorhyncus condylocarpon)
39	tirigu	wheat, corn
40	tiyi	tea
41	tsipe	type of grass
42	tsonyo	bushy shrubs
43	tsoyo	bitter leaf, Vernonia spp
44	tsumbwi	bitter tasting sorghum/ millet
45	wanje	grass, Setaria palustris
46	zumba	herb (Crotalaria brevidens)

Appendix 7B: The instruments set

1	bangwe	a guitar-like musical instrument, harp
2	banya	a kind of stick or club
3	dambalitsiro	discarded material
4	deka	trap
5	gubu	one-stringed bow shaped musical instrument
6	gumbagumba	gramophone, record player, juke box
7	gumbu	musical instrument, animal/ soldiers' track in bush
8	gwang'wa	mattock, pickaxe
9	gwichi	bar for closing a grass door
10	Juba	chisel
11	khunguniza	smooth a pot, smooth the ground by raking up leaves
12	kombe	fishing net
13	komichi	cup
14	konya	string made from the bark of a tree
15	kulimbo	a pole that holds a grass door from falling
16	kundwe	red ochre used for colouring pots
17	kwenga	a kind of drum beaten to warn, signify and inform people of something
18	kwenje	a kind of drum used to alert people in an emergency
19	lamba	belt, lace
20	ogala	auger
21	pango	a guitar-like musical instrument with seven strings, harp
22	phwitika	panga-knife, machete
23	poto	pot
24	psetete	rafter
25	talasada	panga knife
26	tasho	earthen jar
27	thibulo	men's traditional medicine for semen production
28	tikwi	kind of gourd with a narrow middle

Stellenbosch University https://scholar.sun.ac.za

29	zenga	curved knife/axe, sickle
30	gututu	chamber pot
31	kokoko	a small pole for constructing the platform of a granary
32	phini	paper-clip, safety pin
33	phinifolo	safety pin

Appendix 7C: Miscellaneous inanimate nouns

1	budabuda	duodenum
2	buledi	bread
3	firizesi	frozen lollipops popular with children
4	gumbasi	a small worshipping hut
5	gumbwa	thick gun's end/ butt, alga
6	khanga	guinea fowl like dotted cloth, spinning stick in kind of game
7	khumbako	a small worshipping hut
8	zambwe	west
9	zipi	zip
10	gonamkumange	mat
11	pidigoliro	big toe/ forefinger
12	bandira	earth up, bank up, enlarge ridges
13	banyira	prize, makeweight, addendum
14	basera	prize, makeweight
15	betchi	competition, contest, rivalry
16	fumba	wrist jewellery, herb for crop production increment
17	galauza	new ridges for planting
18	gonamkumange	mat
19	gowo	kind of hoe
20	konye	cramp/ spasm, convulsion
21	pamkomera	kind of rain
22	pamkonda	showers which come after harvest, rain that wets crops in the field after ripening
23	phathiphathi	blade of grass, when just sprouting, a just budding maize cob bud
24	pidigoliro	big toe/ forefinger
25	pinyolo	pawned goods, mortgage, lien
26	pitipiti	visit, act of running across one another
27	pokopoko	cacophony, pandemonium, mayhem
28	pondereza	domination, discrimination, oppression
29	tambalale	kind of children's game

Appendix 7D: The a- initial nouns

1	abwito	eye discharge, rheum
2	achale	achar/ pickle
3	adakulasa	the Pleiades
4	adyo	garlic
5	akamwiniatsatana	the Pleiades
6	aleka	kind of liquor

Stellenbosch University https://scholar.sun.ac.za

7	alumali	rack
8	anyezi	onions/ onion
9	asipirini	aspirin
10	atulo	dreams, ignorant people
11	ayisikirimu	ice cream

Appedix 8: Type 1 nouns in the *li*-AC (≈NC5)

1	biyo	fish enclosure (trap), breakwater
2	chenjerero	act of outdoing somebody
3	chenjezo	warning, alarm, advice
4	dalitso	blessing
5	dandaulo	complaint, concern
6	dangaliro	space, gap
7	dipo	fine, fee, award, ransom, pledge
8	doko	port, harbour
9	dombolo	ransom, liberator
10	dongosolo	system, order, plan, structure
11	dontho	drop
12	dziko	country, land, earth, planet
13	fanizo	image, parable, allegory
14	fano	idol, effigy
15	fule	heap, pile, mound
16	fulufute	edible black and red flying ant
17	funso	question
18	ganizo	thought, idea, concept, opinion
19	gogodo	a big bone
20	golo	goal, goalpost
21	gomo	hill with trees, forest
22	gonero	quiver
23	gongo	strand, single piece of thread
24	gowero	boy's house, dormitory
25	gwero	origin, source, basis, cause
26	khomo	entrance, doorway
27	khumbuko	small house for consulting spirits
28	khwerero	ladder, steps, stairs
29	lamulo	law, rule
30	lemba	script, word, vowel
31	lembetso	dictation
32	lingaliro	intention, decision
33	lingo	hole in the mound from which termites come out
34	londo	line on a cloth made as a result of ironing
35	lonjezo	promise, vow
36	loto	dream
37	lulute	ululation
38	lumo	razor blade
39	lupsa	ashes of burnt grass or burnt bush

Stellenbosch University https://scholar.sun.ac.za

40	pangano	contract, agreement, promise
41	pempho	request
42	phikiso	oiling beer porridge for the second time
43	phiko	wing
44	phinde	kind of small mat
45	phitso	large pitcher used for brewing beer
46	phokoso	noise
47	phote	eddy
48	phukuto	the season in which some trees shed their leaves
49	phukwa	homesickness
50	phuma	breathlessness, acting too quickly
51	phumulo	a resting place for people during a funeral on their way to the grave
52	phunziro	lesson, topic, subject
53	phunzo	disrespect, audacity, impudence
54	sewero	play
55	somo	banana stalk, green palm leaf
56	sosa	act of clearing the garden
57	tamo	praise, dewlap
58	tanthauzo	meaning, sense, interpretation
59	tchimo	sin
60	tchinga	protection, fence, impediment
61	tchingo	trap for catching fish
62	tembedzo	promise
63	temberero	curse, swearword
64	thambo	sky
65	thamo	arrogance, boastfulness, pride
66	thangata	hard labour, work without pay
67	thangato	help, assistance
68	thawo	internode
69	thema	thick forest, wilderness
70	thendero	groundnut flour
71	thereko	big pot for brewing beer
72	thungo	end
73	tsankho	segregation, apartheid, racism
74	tsogolo	future
75	tsokomole	kind of large mushroom
76	tsonga	height of a man or animal
77	vunga	part in the mice's hole where mice sleep
78	vuto	problem
79	vuwo	pot (used by traditional healers), haul of fish
80	yankho	answer, response, solution
81	yesero	temptation
82	zunzo	crisis, persecution

Stellenbosch University https://scholar.sun.ac.za

Appendix 9: Animate nouns in the *li*-AC(≈NC5)

1	balang'ombe	big wasp
2	bavu	wasp, hornet
3	bavunyondo	kind of big wasp
4	bunda	young dove
5	doda	adult man
6	dole	type of fish
7	dolo	hero, expert
8	dondwe	gerbil
9	doole	gerbil
10	dzinzi	infertile domesticated female animal
11	dzombe	locust(s)
12	godi	ant-eater
13	hule	whore
14	Jaha	strong adult man
15	jeremusi	germ, micro organism
16	kadini	veterinarian
17	khanda	baby
18	kholo	parent
19	khunguzi	carpenter bee, large wasp
20	khuwi	vulture
21	ligondo	slender mongoose
22	likonyani	calf
23	likumba	kind of fish
24	likungwi	mamba
25	limeni	large male cane rat
26	lipedwa	bagworm moth
27	lipumi	type of flying termite
28	liwenya	kind of pest
29	lizinyani	calf
30	phandauzi	parent goat
31	phanya	gerbil/ mouse like animal
32	thadzi	female animal
33	thamba	fish of Ciprinidae group
34	thebwerebwe	very fat person
35	thekenya	jigger flea
36	themba	barbel, small cyprinid fish
37	thodwe	gerbil
38	thole	calf, rail trolley
39	thomo	children's character
40	thong'ondo	mouse like animal
41	tsotsi	thief, crook, villain
42	tsungwa	kind of fish from Cichlidae group

Stellenbosch University https://scholar.sun.ac.za

Appendix 10: *ma*-initial mass and uncount nouns

	Noun	gloss
1	maale	sign for dwindled friendship & enmity
2	mabadwidwe	person's build, birth-rate, citizenship
3	mabenga	spaces in a grass door or roof
4	mabi	faeces
5	mabimbi	wave, billow
6	mabindi	kind of girl's game
7	maboti	marabou stork
8	mabubusa	girl's breast, poor person
9	mabuka	blister or boil with unknown causes
10	mabutsutsu	girl's breast, poor person
11	machenga	kind of black cricket
12	machenya	foot sores
13	macherefutsa	filtrate
14	macherekwete	traditional harp
15	machesi	match/ matches
16	machikumba	large lone baboon
17	machira	hammock/ stretcher, ransom
18	machitachita	doings, liveliness
19	macholicholi	slanting place/ steep slope
20	made	woven cloth fabric
21	madede	sleeping habit, pleasant voice in singing
22	madekufana	scarecrow
23	madibi	spotted animal
24	madomba	kind of livestock's disease
25	madotiro	water spilt from full jar
26	madukaduka	bamboo/ timber waste from chiselling, planing/ carving
27	madumbo	kind of Ngoni traditional dancewear
28	madyaidya	eating in different places, bad behaviour
29	madyeramphoto	bribery, corruption, greed
30	madyodyo	medicine/ love token, food, pleasant voice in singing
31	madzedze	newly bought hoe blade
32	madzeradzera	strange thing
33	madzeram'phuno	nasal sound
34	madzi	water, liquid
35	madziachitengwa	large amount of run off water from hills
36	madziamphumphuli	large amount of run off water from hills
37	madzimwira	hot spring
38	madzudzu	red maize variety
39	madzulo	evening
40	madzuwa	breastmilk
41	mafululu	spotted animal
42	mafundi	sunheat, sunstroke
43	mafuta	fat, oil, fuel

Stellenbosch University https://scholar.sun.ac.za

44	mafutambi	yellow maize
45	mafutu	shrubs/ small forest
46	magadi	strong salt made from ashes
47	magamba	black soot, cooked fresh peas eaten without main meal
48	magampha	planting method, meal of peas eaten without nsima
49	magawagawa	disease involving leg ulceration
50	magazi	blood
51	magetsi	electricity, electric lamps/ lights
52	mago	kind of women's game
53	magodagoda	respect that a wife gives her husband for more love, backing sound
54	magodi	food prepared & kept for husband by wife
55	magogobi	very steep & stony place
56	magololo	kind of musical instrument
57	magwankhwa	slim body, very thin person's bone appearance
58	maidyaidya	eating in different places, bad behaviour
59	maiko	place for keeping things
60	maimba	unlawful wedlock
61	majiga	crossroad
62	majodi	watermelon's sweetless variety
63	majolijoli	king's bodyguards
64	makajo	women's cheek decorations
65	makaka	dry peeled cassava, decoration, pride
66	makako	lifting machine, gallows, small poles for drying tobacco in shed
67	makale	finger rings
68	makambo	piece of cloth
69	makandatchembere	varicose veins
70	makande	black clay soil, split bamboo for making granary
71	makani	stubbornness, mischief, rudeness
72	makata	cooked unpeeled peas eaten without main meal, high place
73	makate	groundnuts' disease
74	makavu	slag, maize sown very late
75	makeadafa	kind of plant
76	makemadzi	black water insect
77	makemfika	stick into which another stick is twirled to make fire
78	makha	fish gill
79	makhala	big roof pole
80	makhalang'oma	diaphragm
81	makhavi	slag, maize sown very late
82	makhololo	phlegm, sputum
83	makhoti	judiciary
84	makidi	spotted animal
85	makina	machine, engine, device
86	makokoto	leftovers of the pot after food has been shared
87	makoloanyamu	black male ants
88	makope	ripe red cucumber
89	makoti	kind of skin disease

Stellenbosch University https://scholar.sun.ac.za

90	makukutsi	things deposited on riverbank by water
91	makulikuli	water waves
92	makumba	kind of tilapia fish
93	makumbuka	large variety of banana
94	makuna	labia
95	makunkhu	soaked rice eaten raw, relish remnant
96	makupala	mat, body hair
97	makutuangoma	kind of plant
98	makwa	scales
99	makwale	grass blade, kind of fish, Astatotilapia calliptera
100	makwatakwata	consecutive marrying of several wives
101	makwekwetembe	cooked fresh little pumpkin piece
102	makwenje	teeth gap
103	makwetekwete	gait of an old person
104	makwinja	wrinkles on an old person's skin
105	mala	things that mouse uses for its nest
106	malagada	rocky soil
107	malaina	gmelina tree
108	malakwi	fish gill
109	malala	garden natural germinated maize/ sorghum
110	malambadothi	soil binding substance
111	malandi	flames
112	malasha	coal
113	malaya	shirt, clothes
114	malechereche	end, conclusion
115	malekwa	grass blade, kind of fish
116	malemya	big alligator, cuckoo hawk
117	malendebza	swarm of bees
118	malenga	space between the wall & roof
119	malesa	molasses, bees' nest full of honey
120	malimbanjira	beer leftovers
121	malimidwe	agriculture, cultivation, crop/animal husbandry
122	malinde	ambush, plot, attack
123	malipenga	kind of dance
124	malipopola	roof strengthen material
125	malipsata	nudity, nakedness
126	maliragudu	red ants
127	maliralira	frequent cries
128	malire	boundary, limit, end/ lifespan
129	maliroliro	sticky substance that oozes from a tree when cut
130	maliseche	genitals, nakedness/ nudity
131	maliyano	garden boundary
132	malo	place, plot, land
133	malobzadzuwa	past time
134	malombo	invocational dance
135	malonda	business, commerce, goods to be sold/ bought

Stellenbosch University https://scholar.sun.ac.za

136	malopa	meal prepared at a funeral
137	maloto	dream, instinct, guess
138	malovu	continuous speaking
139	malubwerubwe	act of talking while in deep sleep, hallucination
140	malula	dry unmature groundnut or rice grain
141	maluli	shoot without miss, coincidental meeting
142	malulu	marksmanship, offal
143	malumbambewa	fish found at upper river source after flood
144	malunga	abstinence, sensitivity, having self control
145	malunje	thick forest
146	mamba	fish/ snake scales
147	mambedza	sap, snail's slime, albumen
148	mambulu	boiled maize
149	mame	dew, temporal thing
150	mamgwedza	very big mud fish
151	mana	manna
152	manandi	adhesiveness
153	manda	grave
154	mandalena	mandarin oranges
155	mande	tracks trodden in the grass, pleasant voice in singing
156	Mande	Monday
157	mandionekedwe	ostentation, showy display
158	mandongo	bulrush millet's disease
159	manenanena	failure to keep secret
160	manera	ladder, steps, staircase
161	mang'a	cracks
162	mangadzi	maize crop planted after its planting season
163	mangalangala	clever & impudent act for showing off to others
164	mang'anzo	unwell ground flour
165	mangawa	grudge, debt, suspicion
166	mang'ombe	response, comment/ remark
167	mangukuludwa	porridge from nsima
168	mangungu	boiled dehusked maize
169	mangwedza	very big mud fish
170	mangwere	type of musical instrument
171	mangwesa	very big mud fish
172	mangwindo	very old mouse
173	manjenje	palsy
174	mankhanana	sputum
175	mankhananda	dysentery, kind of sexually transmitted disease
176	mankhasha	exaggeration
177	mankhokwe	shoot
178	mankhomwa	unripe sugarcane
179	mankhukwa	maize grain husks
180	mankhupete	maize grain husks
181	mankhusu	chaff, holes

Stellenbosch University https://scholar.sun.ac.za

182	mankhwala	medicine, drug
183	manoamcheni	sharpened teeth
184	manoansadzu	large gaped front teeth
185	manong'onong'o	hearsay/ rumour, gossip
186	manoni	fatty things
187	mantchedza	bulrushes, sedge
188	mantha	fear, anxiety
189	manthongo	eye rheum
190	manthova	sap, snail's slime, albumen
191	manya	old person's body wrinkle, crow's feet
192	manyama	okra, ladies' fingers
193	manyame	kind of mushroom
194	manyazi	shame, shyness, embarrassment
195	manyenje	body hair, lion's mane, torn clothes
196	manyi	dung, faeces/ stool, excrement
197	manyonyo	sitting with legs crossed
198	manyowa	manure, compost, muck
199	manyuchi	honey, sweet, coitus
200	manzanda	maize grain husks
201	manzenene	hoarse, raucous
202	manzere	left
203	maoza	act of composting unripe grass with soil
204	mapazianjiwa	kind of tattoos
205	mapembe	edges of a hoe
206	mapenda	kind of ordeal, shooting skill of hitting target
207	mapeto	end, tip, conclusion
208	maphephe	hearsay/ rumour
209	mapinga	kind of disease
210	mapira	sorghum
211	mapolipoli	slanting place/ steep slope
212	mapumba	lumpy bit of uncooked flour in nsima
213	mapundi	clever doing of things
214	maputa	young mice that are unable to open eyes
215	mapwenga	fat meat
216	mapwitimanoni	fat meat
217	masa	mouse nest, beehive
218	masabani	baldness through shaving
219	masagwidi	disease of swollen cheeks/ mumps
220	masala	uncultivated area
221	masalamusi	bad luck, uncanny/ supernatural happenings
222	masalanga	pounded maize
223	masangwesangwe	cloud indicating rain fall, mice trap
224	masanje	kind of children's game, childishness
225	masano	graveyard, sepulchre
226	masanzi	vomit
227	masenga	growing of crops after its season

Stellenbosch University https://scholar.sun.ac.za

228	masese	kind of beer, dregs/ lees of beer, traditional dance
229	masewe	fish eggs
230	mashini	machine
231	mashosha	head's bald part
232	masika	harvesting time
233	masika	dregs/ lees of beer
234	masoabambo	kind of tattoos
235	masokanthanga	coveting of several things at once
236	masoma	small poles placed vertically when building granary
237	masomphenya	vision/ foresight, dream, phantasm
238	masukupiza	salt production place
239	maswaswani	refuse/ scattered leaves
240	masweswe	baldness about the temples
241	matambi	untargeted arrow shoot, promiscuous behaviour
242	matambo	grave/ graveyard
243	matanga	watermelon's sweetless variety
244	matangadza	tying of person's legs, hindrance to free movement
245	matanthi	malted maize in beer making process
246	matatalazi	struggle, dispute, disagreement
247	mate	saliva, spittle
248	matebani	play
249	mateka	beer drunk before field preparation, beer tasted for its quality
250	matenda	disease/ illness, patient
251	matete	bee, reeds
252	matha	untargeted arrow shoot, promiscuous behaviour
253	mathambo	grave/ graveyard
254	mathanyula	homosexuality
255	mathithi	cataract, waterfalls
256	mathotho	drop, blemish, spot/ dot
257	mathulinga	tonsillitis
258	mathuthu	stammer/ stammering, stumbling
259	mathutsi	sap, snail's slime, albumen
260	mathwe	kind of mushroom
261	matipa	women's incisions on their backs & thighs
262	matiti	malaria, person's body joints
263	matolatola	petty theft
264	matombozi	small edible worms
265	matondo	kind of caterpillar, mtondo tree fruit
266	matsatule	hairy worms
267	matseka	chest cavity, diaphragm
268	matsire	hangover
269	matsiriro	confluence, convergence, end
270	matsite	a kind of mushroom
271	matso	edges of something
272	matsolo	part of a stick for making bow
273	matsonono	kind of beans

Stellenbosch University https://scholar.sun.ac.za

274	matsundi	thing struggles against water/ wind flow
275	matukutuku	defiance, rebelliousness, unruliness
276	matundu	vitiligo/ leukoderma
277	matundulu	pod
278	matutcha	wave, billow
279	maufutu	fresh beautiful flowers
280	mauta	the Pleiades
281	mauwa	bamboo's/ timber's waste from chiselling, planing/ carving
282	mavalanguwo	meat fixed to hide after animal skinning
283	maviko	leaves put in a water container for preventing spilling
284	mavokoko	new leaves on a branch
285	mavule	horse's/ lion's mane, animal's bristle
286	mavulevule	meat fixed to hide after animal skinning
287	mavume	chorus
288	mavumukira	quick job
289	mavundu	water with residue in a container
290	mavuvu	riot/ violence, controversy, argument
291	mawala	kind of fish
292	mawangamawanga	multicolour
293	mawende	gap between teeth
294	mawere	breasts, breastmilk
295	mawinu	body pains after hard work
296	mawiri	civet cat
297	mawonde	unmature reeds not fit for making a mat
298	mawoza	act of covering weeds with soil
299	mawu	words, sound
300	mayawa	dried cassava
301	mayemba	kind of white beans
302	mayepezi	rainwater gushing into the house by wind
303	mayere	method for finding lucky/wrong person, lot, tricks
304	mayimwamna	maternal uncle
305	mazame	ruin
306	mazangazime	hallucination, big problem, dream
307	mazeze	kind of small white fish
308	mazinga	screw
309	maziru	foot sore
310	maziwa	breastmilk
311	mazizi	coldness/ cold, coolness, slowness
312	mazo	feathers fixed to arrow's tail
313	mazoma	boy's initiation ceremony
314	mazombwe	kind of grasshopper
315	mazowe	kind of tobacco/ local fruits/ plant, eggs
316	mazwa	feathers fixed to arrow's tail
317	macheka	sawing industry, saw
318	macheza	conversation, chat, discussion, hobby
319	machokero	reasons for leaving, finality

Stellenbosch University https://scholar.sun.ac.za

320	madyerero	feast, celebration
321	madyo	delicious food, appetite
322	maferano	finality, determination to die together with the opponent
323	mafinya	pus
324	magwera	leftover (especially of the maize in the field)
325	makombo	leftover food, oddments
326	makupe	very big mat
327	makupo	leaves used to fan or extinguish fire
328	makwinya	wrinkles, frown-marks
329	malaulo	bad omen
330	malekano	road junction, crossroads
331	malekezero	end, boundary
332	maliro	funeral
333	malodza	bewitchment, evil tidings
334	maluma	joint, place where two things join
335	malungo	malaria person's body joints
336	mambala	place with unburnt grass
337	mamina	nasal discharge
338	manyenya	edible black and red flying ant
339	mapserera	burnt food substance in the pot
340	masambiso	place where the pall bearers stop for a rest before they reach the graveyard with a corpse
341	masendera	shelled maize that is boiled
342	masiye	an opharn, house where parents have both died
343	matama	pride, arrogance
344	matanda	scaffold
345	matchera	fishing net
346	matewe	rickets
347	matimati	tomato
348	matokoso	mixture of water, sugar & maize husks in the process of distilling liquor
349	matola	unlicensed transport of people
350	matsitso	slope, gradient
351	matsukuluzwa	dregs, filtrate
352	matsukwa	grey water left in the pot after the soaked pounded maize has been taken out
353	matupsa	festering of a wound
354	mauka	genital warts
355	maziko	foundation, origin, basis
356	malawi	flames

Appendix 11: *u*-AC animate nouns

Appendix 11A: Animate *m*-initial nouns in the *u*-AC (≈NC3)

1	m'bampha	a kind of termite with a big red head
2	mcheni	kind of fish of the Cichlidae group with sharp teeth

Stellenbosch University https://scholar.sun.ac.za

3	mdulachala	scorpion
4	mdulamphuno	praying mantis
5	mdyakuwe	a kind of small bird with a crest
6	mdzukulu	grandchild, great-nephew/ great-niece, a person who assist the bereaved family
7	metete	a bare-necked hen
8	mfule	castrated bull
9	mkango	lion
10	mkoka	an animal that has died on its own
11	mkokowogona	a clan's oldest member who has authority over the junior members
12	mkombozi	person who takes care of a disabled person
13	mkota	a parent/mother animal
14	mkuku	bush-baby
15	mkunga	eel, lazy person, dull person, poor person, fool
16	mkung'ung'u	dwarf farm animal
17	mlamba	kind of catfish/ mudfish
18	mleme	a bat
19	mlotamwera	kind of dark chambo fish
20	m'mbulu	wild dog
21	mnjiri	boar, warthog
22	mpasa	kind of fish of the Ciprinidae
23	mpatsachokolo	monkey pod/ eared cassia
24	mpekesa	a cichlid fish
25	mpepedwa	a bagworm moth
26	mpolo	stillborn child/ child born without strength
27	mpoto	an old male animal with curly horns
28	mpsoli	a kind of mouse
29	msambalitsiro	long black mudfish
30	msambululu	lizard
31	msulu	banded mongoose
32	msululu	an impotent goat, a young female mouse that has reached the stage for mating
33	msundu	leech
34	mswachulu	animal which breaks anthills and feeds on the ants
35	mswankhono	stork
36	mtantha	unfaithful person
37	mtanthamphuno	praying mantis
38	mtcheni	kind of fish
39	mtheno	castrated bull
40	muswe	soldier termite (without wings)
41	muuka	ghost, person with strange appearance
42	muwale	brother/sister, relative/ family member, kin, friend, neighbour
43	mwakule	a kind of weaver bird
44	mzenga	civet cat
45	mzime	last born child
46	mzwanya	skilled person, specialist, knowledgeable person

Stellenbosch University https://scholar.sun.ac.za

Appendix 11B: u-initial animate nouns

	Noun	gloss
1	udzudzu	mosquito
2	udzungulu	a kind of bird
3	ukonokono	sea snails
4	ukuku	fleas
5	ulumira	mosquito like insects that bite
6	ulungu	type of pain e.g. snake bite
7	usimwi	a bare-necked hen
8	usipa	type of small fish (Ciprinidae group)
9	utaka	type of fish (Cichlidae group)
10	utakafumbi	mouse
11	utitiri	fleas

Appendix 12: *U*-initial concrete nouns that control AM *u- (≈NC14)*

1	ubongo	brain
2	ubweya	fur
3	ubwinga	herb given to girls who have reached puberty
4	uchema	palmwine
5	uchi	honey
6	ude	cobweb
7	udzu	grass
8	udzudzu	mosquito
9	udzungulu	a kind of bird
10	ufa	flour
11	ukala	sperm/semen
12	ukalisitiya	eucharist
13	ukanga	long hairs of an animal/bracelet
14	ukonde	net
15	ukonokono	sea snails
16	ukopo	brains
17	ukuku	fleas
18	ulalo	bridge
19	ulembe	poison applied to a spear/arrow
20	ulimbo	glue trap
21	ulumira	mosquito like insects that bite
22	ulusi	thread
23	uluwa	husks of sorghum seeds
24	umbwinya	curd
25	umuna	semen, manhood
26	una	burrow, mouse hole
27	undi	flour from fried grains e.g. pumpkin seeds
28	unguwo	fleas

Stellenbosch University https://scholar.sun.ac.za

29	unongo	brain
30	unotsi	eye discharge
31	ununda	type of beetle
32	unyang'anya	type of very tiny flea on chickens
33	unyewa	butt-end of an arrow
34	unyinji	crowd of people
35	unyolo	shackle, fetter
36	unyoma	glue trap
37	upotche	rheum, foam out of fermenting beer
38	usanjo	bridge
39	usimwi	a bare-necked hen
40	usinini	gums
41	usipa	type of small fish (Ciprinidae group)
42	usufu	kapok
43	uta	bow
44	utaka	type of fish (Cichlidae group)
45	utakafumbi	mouse
46	utata	fleas
47	utatavu	cobweb
48	utawaleza	rainbow
49	utawo	place for preservation of meat or fish
50	utitiri	fleas
51	utomoni	sticky sap/glue
52	utonde	semen, fertility, fecundity
53	utongole	dung
54	utoto	dye, colour, paint
55	utsi	smoke
56	utuchi	sawdust
57	utululu	kind of tall grass
58	uvuni	oven
59	uwende	clotted animal blood as food
60	uyalo	platform made from grass/reeds for drying plates, crops
61	wonga	gunpowder
62	wulu	wool

Appendix 13: Nouns with other word-initial characteristics in the *u*-AC (≈NC3&14)

1	bawe	waterside's grass
2	binda	black & red grained maize cob
3	bombwe	oil residues
4	bongo	brain
5	bovu	cheek, Brachystegia longifolia
6	bowa	mushroom
7	bowe	a cooking pot with a small opening
8	bundu	things found in bees' nests

Stellenbosch University https://scholar.sun.ac.za

9	buwa	hunting animals
10	buye	wild potato, Plectranthus esculentus
11	bwemba	fruit of the tamarind tree
12	bweya	fur
13	bwinga	herb given to girls who have reached puberty
14	chono	kind of trap for small animals
15	dabwi	poisonous substance
16	dembera	sorghum/ millet with large white grains
17	derema	kind of mushroom
18	dowadowa	cichlid fish
19	dulu	after first rain sprouting vegetation
20	fuluwenza	influenza/ flu
21	fumbefumbe	shavings, sawdust
22	fwifwi	plant (Coccinia adoensis)
23	gamba	kind of grasshopper
24	gebede	bead put on string as jewellery
25	gompho	very short trousers
26	gongonya	a line on a cloth made as a result of ironing
27	wonga	gunpowder
28	wulu	wool

Appendix 14: Nouns controlling AMs *m(u)-a-* that were incorrectly allocated to NC14

1	gombamluzu	kind of large mushroom
2	gulamamine	kind of grasshopper
3	gulewamkulu	masked dancer, character in gule wamkulu dance
4	gumbu	musical instrument, animal/ soldiers' track in bush
5	gumbwa	thick gun's end/ butt, alga
6	gwafa	guava
7	gwamula	force entry, molestation after forcibly entering house
8	hereche	incomplete haircut, brown mouse
9	Januwale	January
10	Jejula	a kind of sour beer
11	Julayi	July
12	Juni	June
13	khanga	guinea fowl like dotted cloth, spinning stick in kind of game
14	khanyamba	scorpion
15	khobo	kind of tree (Commiphora africana)
16	khocho	Commiphora africana
17	khokolole	things taken away by run off water
18	kholobowa	variety of large bananas
19	kholowa	sweet potato leaf
20	khoswe	house rat
21	khovani	spiderwort
22	khwakhwakhwa	emptiness, bankruptcy

Stellenbosch University https://scholar.sun.ac.za

23	khwawe	crippleness
24	khwiya	kind of grasshopper
25	komichi	cup
26	kondaine	love potion
27	konya	string made from the bark of a tree
28	kupe	kind of insect
29	lendebza	bees' gathering
30	limba	musical instrument
31	lomwa	type of termite
32	longolwe	variety of red mushroom
33	Novembala	November
34	O, o	O, o
35	Ogasiti	August
36	Okutobala	October
37	onga	gunpowder, semen (figuratively)
38	P, p	P, p
39	pasinjala	passenger
40	phingo	African ebony
41	phinifolo	safety pin
42	phwetekere	small species of tomato
43	pichesi	peach
44	pidigoliro	big toe/ forefinger
45	pilo	pillow, hassock
46	S, s	S, s
47	singano	needle, injection/ syringe
48	T, t	T, t
49	talasada	panga knife
50	thombozi	kind of tree (Diplorhyncus condylocarpon)
51	tsonyo	bushy shrubs
52	tsumbwi	bitter tasting sorghum/ millet
53	V, v	V, v
54	W, w	W, w
55	waya	wire
56	Y, y	Y, y
57	Z, z	Z, z
58	zumba	herb (Crotalaria brevidens)
59	zumu	poisoned concoction, kind of sauce, wild rodent
60	bwaponde	peanut butter
61	gang'a	a big termite
62	achale	achar/ pickle
63	adyo	garlic
64	aleka	kind of liquor
65	alumali	rack
66	alungu	decedent's spirit, spirit visitation
67	amanda	the dead's spirits
68	anyezi	onions/ onion

Stellenbosch University https://scholar.sun.ac.za

69	ayisikirimu	ice cream
70	bale	burley
71	batala	fish, butter
72	biriwita	thick black cloth
73	bononGwe	wild spinach
74	bulugamu	bluegum, eucalyptus
75	buthu	young/ pretty girl
76	bwalero	fast acting poison
77	bwana	boss, master, rich person
78	chale	porridge for beer brewing
79	Disembala	December
80	dududu	kind of plant, yellow mall agama ground lizard
81	Epulo	April
82	Febuluwale	February
83	firizesi	frozen lollipops popular with children
84	fundabwi	gule wamkulu initiates' medicine for promoting memory
85	gondolosi	plant for sexual desire arousal (Mondia whitei)
86	N, n	N, n
87	nakanyimbwi	dribble of saliva/ mucus, mouth uncleanness after meal
88	namndaya	maize flour
89	namtibule	pounded green maize
90	namtindi	multitude, crowd, scores
91	nandaya	pounded green maize
92	nandolo	pigeon peas, Cajanus cajan
93	nanyati	spotted beans
94	napantha	influenza, flu
95	nathando	okra/ ladies' finger

Appendix 15: Non-plural marker noun initial *mi-*

1	micheta	immature maize
2	michewu	a small hat made of ropes, palm leaves (sometimes worn by women to show their grief)
3	mihawu	doing things without patience
4	mikhwithi	a large dark cloud which covers the sky, being dismal
5	mileniyamu	millennium
6	miligalamu	milligram
7	milikisikono	milk scone
8	mililita	millimetre
9	milimo	building material (obtained from the bush)
10	militare	military
11	miliyoni	million
12	mimba	pregnancy, stomach
13	mimbu	wild potato
14	mina	nit
15	mindwi	roots of a crop
16	minganzolo	paper-back acacia

Stellenbosch University https://scholar.sun.ac.za

17	mingazona	a kind of acacia tree
18	minibasi	minibus
19	minisiketi	miniskirt
20	minisitile	ministry
21	miniti	minute
22	minju	roots of a crop
23	mirakuli	miracle
24	misa	mass
25	misailo	missile
26	mishoni	mission
27	misiteki	mistake
28	mitala	meter
29	mitala	polygamy
30	miting'i	meeting

Appendix 16: nouns with a variety of word initial characteristics in the *i*-AC (≈NC9)

A-initial

1	adiresi	address
2	aironi	iron
3	ambulasi	ambulance
4	ambulera	umbrella
5	andiroko	skirt
6	apilo	appeal
7	asikolo	frame of a window which opens on hinges
8	ayata	incisions on the skin

B-initial

1	bafuta	smooth white linen cloth (for covering a dead body/worn by widow)
2	baji	budge
3	bampala	bumper
4	banda	a type of big ant
5	bandi	band
6	basi	bus
7	basiketi	basket
8	benesi	lancet
9	binzo	a ceremonial meal cooked to formalise an engagement between a boy and a girl
10	bisiketi	biscuit
11	bizinesi	business, company/ industry
12	butchala	butcher, butcher's shop
13	bwalidzo	small wooden poles used on a granary rack
14	bwimvi	water with residue in a container
15	bwivu	gums

Stellenbosch University https://scholar.sun.ac.za

16	bwizi	kind of caterpillar eating fresh leaves

D-initial

1	dobwe	boiled and dried bean leaves
2	dobza	kind of mudfish
3	dombe	drum with holes
4	dzinthu	agricultural produce

E-initial

1	edzi	AIDS
2	emvulopu	envelope
3	epuloni	apron

F-initial

1	faindi	fine
2	fakitale	factory
3	famu	farm
4	fendeule	bamboo pith
5	feremu	frame
6	firiji	fridge
7	firimu	film
8	foloko	fork
9	fomu	form
10	foni	phone
11	fosholo	shovel
12	fulasiki	flask
13	fulumba	caltrop
14	fumakazi	queen
15	futali	boiled potatoes in groundnut soup
16	fwataki	type of rifle

G-initial

1	galamafoni	gramophone
2	galandifolo	gramophone
3	galimoto	car
4	gemeswa	very big mat
5	gindu	remains (of maize in a mortar
6	golokolo	small of the back, lower vertebrae
7	golosale	glossary
8	gomwa	bean porridge, potage, infertile person
9	gunthe	a type of cultivation
10	gwanda	short-sleeved shirt

Stellenbosch University https://scholar.sun.ac.za

H-initial

1	habu	hub
2	hakisowo	hacksaw
3	hamala	hammer
4	handulo	handle
5	hatchi	horse
6	hayala	hire
7	hedi	headmaster, heading a soccer ball
8	holo	hall
9	hotera	hotel
10	hutala	hooter

J-initial

1	jando	circumcision initiation ceremony
2	Jekeseni	injection
3	jele	jail
4	Jenjere	fish poison
5	jerekani	jerry can
6	Jinje	small drum
7	Jombo	boot
8	Jontcho	mixture of tomato and powdered groundnuts used as relish
9	Juga	toolbox

K-initial

1	kakavu	ravenous appetite for meat
2	kalata	letter
3	kampani	company
4	kango	kind of fly
5	kankhumu	sudden death
6	kantini	canteen
7	kapitolosi	part of a traditional hut which supports the roof
8	kapu	cup
9	kasenjere	small variety of sugarcane
10	keyala	address
11	khendulo	candle
12	khukhuto	salty soil
13	khukuwi	chest cavity, diaphragm
14	khuli	stem, bulb
15	khulo	red object at the very tip of a banana bunch
16	khundu	red object at the very tip of a banana bunch
17	khunguluzi	space
18	khuvya	part of a place or thing
19	kompyuta	computer
20	kwacha	Malawian currency

Stellenbosch University https://scholar.sun.ac.za

L-initial

1	lamya	telecommunication, telephone/ phone, call
2	lenikoti	rain coat
3	lole	lorry
4	lumwe	type of termite which swarms when the rain begins
5	lunzule	small kind of owl
6	milimita	any measurement's smallest part

O-initial

1	ofesi	office
2	ogala	auger
3	oloto	alto
4	opareshoni	operation/ surgery

P-initial

1	pasi	type of fly
2	patisi	leg-cloth formerly worn by soldiers
3	phekete	scrap, piece, the sound of breaking
4	phembereza	flour falling when pounding/milling
5	phemphe	Prince-of-Wales feathers
6	phukira	type of vegetable relish that sprouts during the dry season
7	phundudwa	sloughed off snake's skin
8	piki	pickaxe
9	psipe	type of grass

S- initial

1	sabata	sabbath
2	sadaka	gift, memorial ceremony
3	sadulo	saddle
4	safuliya	sauce pan without handle
5	saha	saw, tool with sharp tooth edge
6	saini	signature
7	saiti	a hole in a grave to place a corpse
8	saizi	size
9	sakachulu	type of mouse
10	sakambwa	kind of shrub
11	sakasa	cold season
12	sakasi	suckers, new shoot
13	sale	a kick
14	saluti	salute
15	samani	summons
16	sambale	nsima made from maize flour mixed with ground bran
17	sanda	shroud
18	sandasi	sandal

Stellenbosch University https://scholar.sun.ac.za

19	sandipepa	sandpaper
20	sando	hammer
21	sanga	olive grass snake
22	sangala	cheerfulness
23	sangale	type of sweet potato
24	sangazinji	kind of climbing plant
25	sangulutso	entertainment
26	sangwe	small basket used as a rattle
27	sangweji	sandwich
28	sanjika	kind of fish
29	sansi	type of musical instrument
30	sanza	worn or torn piece of cloth
31	sanzamvula	skin rash from too much sun
32	sapoti	support
33	sapulano	soprano
34	satifiketi	certificate
35	sawa	groundnut, peanut
36	sawatcha	salute
37	sayansi	science
38	sebwa	dehusked maize
39	sefa	sieve, safe
40	semberere	water lily, types of tubers, onion-like plant
41	sensasi	census
42	senti	scent
43	sepusepu	timber chippings
44	serula	cell phone
45	sesenya	a kind of beetle
46	shaka	hole, gap
47	shanzi	second wife to a man as appreciation
48	shasha	crook, expert
49	shati	long-sleeved shirt
50	shirini	shilling
51	shomo	a cry by a hunter who finally kills an animal
52	shonga	pain in the stomach
53	shopu	shop
54	shuga	sugar
55	sika	sickle
56	sikelo	scales
57	siketi	skirt
58	silamba	a kind of bell used in nyau performance
59	simba	initiation camp
60	simbi	iron
61	sipanala	spanner
62	sirika	silk
63	siro	clay, dirt
64	sitampa	stamp

Stellenbosch University https://scholar.sun.ac.za

65	sitediyamu	stadium
66	siteji	stage
67	sitima	train, ship
68	sitolo	store/shop
69	siwiti	sweet
70	sokosi	socks
71	sompho	axe
72	songeya	type of STI
73	songo	black mamba
74	songwe	sunbird
75	sonso	sweetless variety of watermelons
76	sonsokera	stye (a swelling on the edge of the eyelid)
e77	sopo	soap
78	sowo	saw
79	soya	soybean
80	sukulu	school
81	sukundu	hip joint, head of femur
82	sulo	anvil, forge
83	sumbulere	umbrella
84	sumpha	type of swelling, lump
85	sumu	poisoned concoction, type of sauce
86	sundwe	maize left over from previous year, surplus
87	sungo	land protruding into the sea
88	sunsi	large &dangerous weapon, missile
89	sunzo	basket for straining beer
90	sunzu	type of swelling, lump
91	susuli	type of pot, type of mouse
92	suti	suit
93	sutikesi	suitcase
94	swalulu	type of lizard
95	swanda	type of bag
96	swiri	part of a stomach of an animal

T- initial

1	takisii	taxi
2	tambula	tumbler
3	tangani	kidney
4	tayipi	typewriter
5	tchitchi	smell of burnt food
6	tchuthi	holiday
7	telefoni	telephone
8	tepi	tape
9	thale	kind of mushroom
10	thanda	ladies' fingers (okra)
11	thando	communal food eating
12	thenesi	tennis, white canvas shoes

Stellenbosch University https://scholar.sun.ac.za

13	thibo	immature pumpkin cut and mixed with relish
14	thindi	thick bush along riverbanks, crowd
15	thireyi	tray
16	thokoli	round shape
17	thozi	peak of a tree
18	thuluthulu	green/unripe maize
19	thulutso	childbirth ceremony
20	thumbzi	mother goat
21	thunduluzi	thickly entwined trees, stunting crop/tree
22	thunga	type of snake, type of gule wankulu
23	thuphwe	fatigue
24	tiki	tickey, obsolete South African coin
25	tikiti	ticket
26	tipoti	teapot
27	tsamacheya	type of die, type of skittle
28	tsampingu	type of drum
29	tsekaumbereko	incision, tattoo
30	tsiga	faggot, bundle of long things
31	tsonye	disease affecting breastfeeding babies' eyes
32	tsungulo	fool
33	tsuwi	sloughed off snake's skin
34	tupa	sharpening metal, sharpener, file

V- initial

1	vero	veil
2	vesiti	vest
3	visende	animal's testicles
4	voko	the act of being vocal

W- initial

1	wailesi	radio
2	wigi	wig
3	wiribala	wheelbarrow
4	wisikoti	waistcoat
5	wotchi	watch

Y- initial

1	yadi	yard
2	yembe	mango
3	yembedodo	variety of mango
4	yunifomu	uniform
5	yunivesite	university

Stellenbosch University https://scholar.sun.ac.za

Appendix 17: Animate denoting nouns in the *chi*-agreement class (≈ NC7)

1	chagaga	type of insect
2	chambo	type of fish
3	chamnjiri	Name of God
4	chenkha	a kind of fish (Otopharynx spp.)
5	chezi	male dove
6	chibenthe	kind of fish (Copadichromis)
7	chibungu	kind of caterpillar
8	chidakwa	drunkard, alchoholic
9	chidangwaleza	albino, ghost
10	chidangwe	bushbuck
11	chidazi	type of wasp
12	chidengwa	lame person
13	chidindire	old person believed to have wisdom
14	chidombe	kind of small fish
15	chidude	spirit of dead person, fool, person with strange appearance
16	chidyamakanda	sugar daddy, sugar mummy, paederast
17	chidyaonga	white stork
18	chidyupaya	a kind of mouse
19	chidzete	stupid/foolish person
20	chifukufuku	a moth
21	chifukulamchenga	kind of fish
22	chifukumbu	maggot
23	chifumbu	caterpillar
24	chigaga	a type of centipede, a type of herb
25	chigawenga	bandit, gangster, rebel
26	chigwekwe	big old mouse
27	chigweto	lover
28	chigwiri	experienced functionary, professional, veteran
29	chikala	wild animal, dangerous person
30	chikhambwa	scorpion
31	chikhwaya	rich person
32	chikhwiya	kind of grasshopper
33	chikudzu	kind of red mouse
34	chikufenji	kind of fish
35	chikumbe	male cane rat
36	chikumbu	kind of flying insect
37	chikwangwala	crow
38	chikwatu	hawk
39	chilalamadzi	kind of snake
40	chilandamphuno	kind of grasshopper
41	chimphona	type of baboon
42	chipangwi	hornbill
43	chiphadzuwa	beautiful girl, beauty queen
44	chiuli	ratel

Stellenbosch University https://scholar.sun.ac.za

45	chalo	chief, traditional authority
46	chikwangula	woman's last-born child, last item in a basket
47	chilekwa	foolish person
48	chilembwe	roan antelope, character of gule wamkulu dance
49	chilimamine	kind of grasshopper
50	chilimphundi	hornless cow
51	chilinundu	kind of bird
52	chilombo	wild beast/ animal, insect, partaker in Chinyau masquerade
53	chilumphirabere	type of caterpillar
54	chilungulungu	type of caterpillar
55	chimbalangondo	fierce animal, dangerous person
56	chimbanga	eagle
57	chimbano	character in gule wamkulu dance
58	chimbenje	kind of fish
59	chimbu	cleg, hippopotamus fly, gadfly
60	chimbumbuzi	solitary bee
61	chimbuwira	infertile man, impotent man
62	chimbwi	kind of fish, hyena, second hole in a bawo board's inner row
63	chim'dakulasa	scarecrow
64	chimphamba	eagle
65	chimphanga	eagle, gadfly
66	chimphangwi	eagle
67	chimpse	kind of parrot
68	chimunthu	big crowd, derogatory term for a person
69	chimwamowa	drunkard
70	chimwendom'mphako	snake eagle
71	chinengo	kind of flying termite
72	chingolopiyo	a kind of bird
73	chingoti	kind of fish
74	chingowe	fat mouse, boiled maize cob
75	chinkhambo	greedy person
76	chinkhanga	type of cloth, scorpion
77	chinkhanira	scorpion
78	chinkhulo	tilapia
79	chinkhupiti	type of bat
80	chinsanga	poisonous type of snake
81	chinsansa	wild duck, comb duck
82	chinthumbwi	kind of small fish of the Ciprinidae group
83	chinyakwalala	chicken with stuck-out feathers
84	chinyau	masquerade
85	chinziri	common quail
86	chiombamadzi	aquatic python
87	chiombankhanga	martial eagle
88	chipembere	rhinoceros
89	chiphaniphani	firefly
90	chiphona	giant, strong person

Stellenbosch University https://scholar.sun.ac.za

91	chipsolopsolo	young rooster
92	chipungu	vulture
93	chipwale	otopharynx fish
94	chisamba	a family's first born
95	chisidze	elegant' grasshopper
96	chisoni	hedgehog
97	chiswambiya	ichneumon wasp
98	chiswankhono	stork
99	chiswe	termite
100	chitemera	kind of small white fish
101	chitemetsakhasu	chicken used at a funeral
102	chitsekulamuvi	chicken given to estranged woman to reunite her with her husband
103	chitsenda	very short person, foolishness, caterpillar
104	chitserekwete	fool
105	chitsiru	fool
106	chitsitsansepe	family's second born
107	chiumbambiya	ichneumon wasp
108	chiunda	small dove, young pigeon
109	chivumbi	kind of flying termite
110	chivuvu	big black wasp
111	chiwangwi	eagle
112	chiwaula	kind of caterpillar
113	chiwonga	type of small fish
114	chiwunda	chick of a bird
115	chiyabwe	itchy caterpillar
116	chizangala	type of gule wamkulu dance
117	chizimamawere	last born
118	chizingiri	ichneumon wasp
119	chizuma	type of flying ants
120	chizunyani	young dove/pigeon
121	chombwe	civet
122	choso	sunbird

Appendix 18: lexicalised *ku*-initial locative nouns (≈NC17)

1	kumudzi	home village, rural area
2	kumsitu	Nyau camp site
3	kumwezi	menstruation period
4	kunkhadze	Nyau camp site
5	kumadzi	Nyau camp site
6	kumaso	private parts

Appendix 19: lexicalised *pa*-initial locative nouns (≈NC16)

1	pathupi	pregnancy
2	panyo	anus

Stellenbosch University https://scholar.sun.ac.za

3	pansi	under
4	pamtanda	converging of people at a funeral house
5	pakati	pregnancy
6	pachimake	highest point, climax
7	padera	special place
8	pafupi	near
9	paindeinde	climax
10	pakamwa	brim, talking a lot
11	paliponse	everywhere
12	panja	outside
13	patali	far