Shape morphemes in Amarakaeri: nature, origins and development

Broek, Jessie van den

Shape morphemes in Amarakaeri

Nature, origins and development

Thesis
MPhil Linguistics
Leiden University

March 2012

Jessie van den Broek

Supervisor: Prof. dr. W.F.H. Adelaar

Table of contents

Acknowledgements

This thesis is the result of a research on the Amazonian language Amarakaeri. I collected data from the language during a fieldwork trip in the department of Madre de Dios, Peru, where I resided in the indigenous community of Boca Inambari.

Several people have supported this research, in different ways, and should be mentioned here. First of all, I am grateful to my supervisor Willem Adelaar, who has raised my interest in the Amarakaeri language and who has supported me with important insights and comments during my research. I am also indebted to Rodolfo Cerrón Palomino (PUCP) and Gustavo Solís Fonseca (UNMSM), for providing much needed advice at the beginning of my field trip. I thank Klaus Quicque Bolívar from FENAMAD, for putting me in contact with the community of Boca Inambari, and Gerson Medina, for lending me the appropriate outdoor equipment.

Most of all, though, I wish to express my gratitude to all the community members of Boca Inambari, who were willing not only to accept me, but also to share their language, culture, thoughts and rituals with me. Special thanks go out to Manuel and Rosa, for hosting and teaching me; to Cornelio, for being an insightful language consultant, to Ruth, for feeding me, and to Paulita, for being my friend.

List of abbreviations

1PL	1st person plural
1SG	1st person singular
3D	three-dimensional shape
3PL	3rd person plural
3PL→1PL	subject 3rd person plural, object 1st person plural
3SG	3rd person singular
ABS	absolutive
ADJM	adjective marker
AF	affix
APPL	applicative marker
ART	article
CASE	case marker
CAUS	causative
CL	classifier
COP	copular verb
DEF	definite
DEM	demonstrative
DIM	diminutive
ER	empty root
ERG	ergative
EUPH	euphonic
F	feminine
GRN	generic relational noun
IM	imperative
INF	infinitive marker
INT	intensifier
INTERR	interrogative
N	noun
NMLZ	nominalizer
NPOS	non-possessive marker
PAST	past tense
PL	plural
POS	possessive
RP	repeater
SG→SG	singular subject, singular object
SHM	shape morpheme
V	verb
VBLZ	verbalizer

List of tables, figures and maps

Introduction

With over 300 languages belonging to some 20 linguistic families, and a large amount of apparent linguistic isolates, the Amazon basin is one of the 'linguistic black boxes' of the world (Grinevald &Seifart 2004). Research on Amazonian languages, therefore, is highly relevant: not only does it serve to describe and document these languages, many of which are endangered; it may also provide new insights into theoretical issues and change the typological debate. This is exactly what has happened in the typological debate on nominal classification. Before Amazonian languages were studied at a reasonably large scale, the theoretical framework and terminology of nominal classification was mainly based on phenomena in well-studied linguistic families, such as Indo-European gender systems and the Bantu noun classes. As more data from Amazonian languages became available, however, the existing typologies gradually lost their sustainability, since many of the Amazonian systems simply did not seem to fit into any of the established categories. New theories and frameworks had to be developed in order to accommodate these systems. Nowadays, although recent descriptions of Amazonian classifier languages have greatly enhanced our knowledge about nominal classification in the area, much remains to be discovered and understood. Some languages are known only through scarce data, and display systems that seem difficult to place into any of the available typological frameworks.

One of these languages is Amarakaeri, which is spoken in the department of Madre de Dios and belongs to the linguistic family Harakmbut. The language shows a highly interesting system of so-called 'shape morphemes' (Hart 1963), which are omnipresent in the language and occur in many different morphosyntactic contexts. However, these morphemes also display many noun properties and seem to play an important role in word formation in the language. The system has been described by Hart (1963) and Helberg (1984), but these authors do not compare it with other systems of nominal classification in the Amazon basin. Since the previous studies on the topic date from several decades ago, it is interesting to re-examine the system, now taking into account the new typological frameworks and data on other Amazonian languages that have become available in recent decades. Moreover, whereas the shape morphemes have been described before, their origins and development have never been the subject of research, although this is a highly interesting topic. My research question, therefore, is two-fold:

1. *How should the Amarakaeri shape morpheme system be categorized typologically?*
2. *How did the system originate and develop?*

These two questions are related, since the origins and development of the system may also shed light on its nature, and vice versa. In order to investigate these issues, I gathered data during a period of fieldwork in the Amarakaeri community of Boca Inambari. To learn more about the nature and development of the system, I examined the morphosyntactic contexts in which the shape morphemes are used, the degree of semantic generalization that they display, their use in neologisms and with loanwords, and the processes they undergo (such as lexicalization and merging with noun roots). I compared this data, and data from previous works on Amarakaeri, with descriptions of classifier-like phenomena in other Amazonian languages. Furthermore, I took into

account data from Katukina, which is likely to be genetically related to the Harakmbut language family (Adelaar 2000, 2007), and may thus shed light on the origins of the system.

The present research is intended to enhance knowledge about the nature of the Amarakaeri shape morpheme system and the ways in which it may have originated and developed. It will also make a small contribution to the typological discussion on the phenomenon of nominal classification in the Amazon basin. However, this research has its limitations. Since it involved only a short period of fieldwork, my data set mainly consists of word lists, short sentences and grammaticality judgements, since there was not enough time to record and analyse longer texts. Therefore, the discourse-pragmatic use of the system could not be examined properly. Moreover, my research question about the origins and development of the shape morpheme system is unlikely to be answered with certainty, since no historic data of the language is available. The answer to this question will therefore be a series of educated scenarios, rather than a solid conclusion.

The thesis is structured as follows. First of all, a short introduction to the Amarakaeri language and its speakers will be given in Chapter 1. Chapter 2 will provide an overview of the current state of affairs of the typological debate on nominal classification. This will serve as a theoretical framework for Chapter 3, in which systems of nominal classification in other Amazonian languages will be discussed. In Chapter 4, an overview of previous works on shape morphemes in Amarakaeri will be provided. Relevant features of Katukina that may shed light on the Amarakaeri shape morpheme system will be discussed in Chapter 5. In Chapter 6, my fieldwork data will be presented. Finally, all data and insights from previous chapters will be combined and assessed in Chapter 7, in order to formulate answers to the research questions.

1. Introduction to Amarakaeri

This chapter will provide basic information about the Amarakaeri language, its genetic affiliation, and its speakers. Also, the fieldwork methodology that was used in this research will be described here, and some basic characteristics and salient properties of the language will be briefly discussed.

1.1 The Harakmbut language family

Amarakaeri forms part of a relatively small language family called Harakmbut, which is spoken in the southern Peruvian lowlands in the departments of Cusco and Madre de Dios. Other varieties of Harakmbut are Wachipaeri, Arasaeri, Toyoeri, Sapiteri, Pukirieri and Kisambaeri (INEI 2008). Helberg (1984) argues that only Amarakaeri and Wachipaeri can be considered languages with certainty, while the other Harakmbut varieties may be dialects rather than languages. In a 2008 census, Kisambaeri is considered a dialect of Amarakaeri, and Amarakaeri is said to be partially mutually intelligible with Wachipaeri, Toyoeri and Sapitieri (INEI 2008). However, the language versus dialect question remains a point of discussion, since little research has been done on any Harakmbut variety other than Amarakaeri so far (Adelaar 2000: 219).

The genetic affiliation of Harakmbut has also been debated in the past decades. While some consider it an isolate (Wise 1999, Fabre 2005), others place it within the Arawakan family (Matteson 1972, Greenberg 1997). Helberg (1984) does not consider this affiliation with Arawakan languages to be proven, and argues that Harakmbut might just as well be related to the Pano-Tacanan language family, or even to Tupí-Guaraní. Similarly, Adelaar (2000: 221) argues there is very little evidence for a relation between Harakmbut and Arawakan languages, and that other possible genetic affiliations should be considered as well, such as Macro-Tucano and Macro-Puinave. Moreover, in recent years, Adelaar (2000, 2007) has proposed a genetic relation between Harakmbut and Katukina, a language which is spoken in the eastern part of the Brazilian state of Amazonas. Evidence for this relation consists mainly of similarities in basic lexicon, such as body part and kinship terms; more research would be needed in order to establish in more detail how and when the two languages separated. Since the proposed relation between these two languages is highly interesting in the light of the development of the shape morpheme system, it will be further discussed in Chapter 5.

1.2 The Harakmbut peoples

The peoples that are denominated ‘Harakmbut’ are those who speak Harakmbut languages or dialects (Gray 1996). This group is further divided into several ethnolinguistic subgroups, each of which lives in a different specific area of the region. While the names of these groups largely coincide with those of the linguistic varieties mentioned above, the group which speaks Amarakaeri

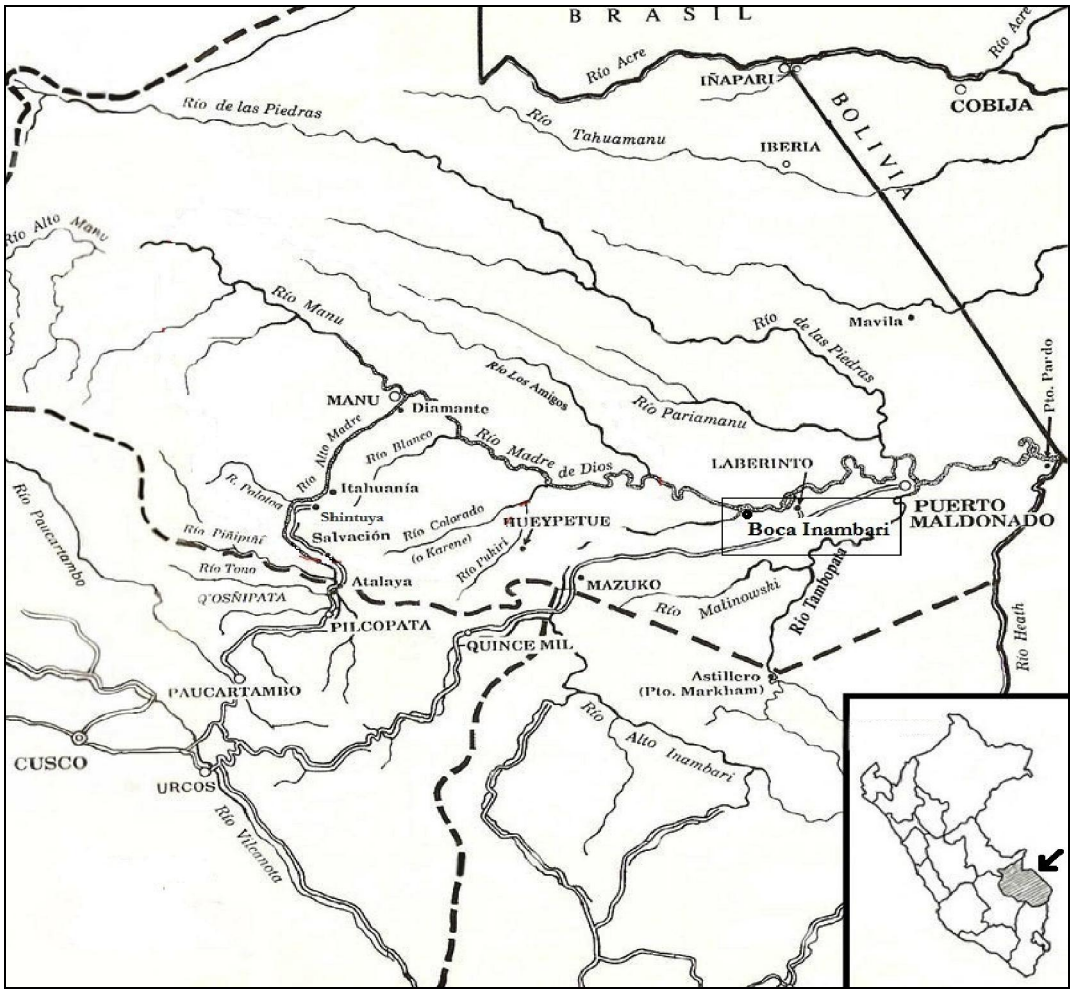
is denominated Arakmbut. Although the different Harakmbut[1] groups are usually identified as a cultural unit, Urteaga (2007 stresses that they constitute a relatively heterogeneous group, and that each ethnolinguistic subgroup distinguishes itself from other groups in terms of clothing, certain customs, and mythology.[2] However, the Harakmbut peoples do share many customs and cultural traits, and have a similar social organization. Some of the general characteristics of the Harakmbut are a traditional organization in patrilineal clans, the importance of the invisible world of spirits, and shamanism (Barriales & Torralba 1970). Traditionally, the Harakmbut were hunter-gatherers and used some form of subsistence farming based on slash and burn techniques. The Harakmbut territory used to stretch from the source of the Madre de Dios river to its confluence with the Inambari. Gray (1996: 14) estimates that by the end of the 19th century, the Harakmbut population numbered around 30.000 people. These numbers rapidly dwindled, however, due to violence and slave raids related to the rubber boom which started in the late 19th century (Gray 1996). During these years, some of the Harakmbut subgroups were almost entirely annihilated, such as the Toyoeri, of which only a handful survived the rubber boom. Contact with non-indigenous society was further intensified when Dominican missionaries entered the area in the early 20th century and moved most of the Harakmbut into newly established missions such as Shintuya, in order to convert them to Christianity. This contact caused severe smallpox and influenza epidemics in the 1940s and 1950s, which decimated the Arasaeri, Sapiteri and Kisambaeri populations (Gray 1996: 7-8).

Nowadays, the total Harakmbut population is estimated between 1000 and 2000 people, most of which belong to the Arakmbut group. The Wachipaeri population is said to count a few hundred people, while the Sapiteri, Arasaeri, Toyoeri, Kisambaeri and Pukirieri groups consist of only a few families each (Gray 1996). Most of the Arakmbut currently live in a territory between the rivers Alto Madre de Dios and Inambari, where they are spread over several riverside communities that are legally recognized by the Peruvian government (Gray 1996, INEI 2008). Some of these communities, including Boca Inambari (see Map 1), where I conducted my fieldwork, were established by Arakmbut groups that had escaped from the missions in the early 1970s. However, a medium sized Arakmbut community still resides in the mission of Shintuya, where they live next to speakers of Wachipaeri and Machiguenga. Hunting, fishing and small subsistence farming still form an important part of Amarakaeri life. Some people have their own *chacra*, on which they mainly cultivate bananas, papaya, yucca or tobacco, and communal hunting or fishing trips are common practice. In recent decades, however, the Amarakaeri have also become involved in gold extraction, which has come to play a major role in the area's economy since the second half of the 20th century (Urteaga 2007). Many Amarakaeri families now collaborate with immigrant gold miners, who live on their land and share the profit with their hosts. This has resulted in a more money-based economy within the communities. For instance, the inhabitants of Boca Inambari now frequently travel to the nearby miners' town Laberinto in order to sell their gold, and to buy 'non-traditional' foodstuffs such as soft drinks, rice, beans and biscuits, and occasionally medicines and electronic devices.

[1] In the past, and in some previous publications, such as Barriales & Torralba (1970), the Harakmbut are denominated *Mashcos*, but this is nowadays considered a derogatory term.
[2] While conducting my fieldwork, I also noticed some resentment between different Harakmbut groups, as I heard some of my Amarakaeri language consultants speaking negatively about the Toyoeris.

MAP 1

LOCATION OF BOCA INAMBARI AND THE DEPARTMENT OF MADRE DE DIOS



(source: www.selvasamazonicas.org)

1.3 Language attitudes and degree of endangerment

With reported speaker numbers between 500 (Tripp 1995) and 1000 (Gray 1996), Amarakaeri must be considered an endangered language. The UNESCO Atlas of the Word's Languages in Danger (2010) does not list Amarakaeri as a separate language, but considers Harakmbut as a whole to be 'definitely endangered'. Although bilingual primary education exists in several of the Amarakaeri communities, and some teaching materials have been developed by the Summer Institute of Linguistics (SIL) in the 1970s and 1980s, the survival of the Amarakaeri language in the future seems uncertain. For instance, in the community of Boca Inambari, an ongoing shift towards Spanish can be observed. Apart from a handful elderly people who are monolingual in Amarakaeri, most of the adults are bilingual. At the same time, the speech of the younger adults shows frequent code switching, and these people seem to speak mainly Spanish with their children. This is quite worrisome from a conservationist point of view, since a lack of intergenerational transmission may result in abandonment of the language within one or two generations. Moreover, nearby gold

mining activities in the Madre de Dios and Inambari rivers attract a large amount of miners from coastal Peru and the Andes region, which leads to a constant presence of Spanish speaking people in the community. It must be noted, however, that because of its relative proximity to the urban centres Laberinto and Puerto Maldonado, Boca Inambari might be under heavier influence of the Spanish language and non-indigenous culture than the Amarakaeri communities that are situated further upstream along the Madre de Dios river and its tributaries.

Despite all these threats to the language, language attitudes among adults seem to be quite positive, and the Amarakaeri speakers that I worked with seemed to take pride in their native language and traditions. Moreover, the creation of neologisms for new cultural items instead of using loanwords indicates that the language is still valued as a full-fledged means of communication, and that it is kept up-to-date with new cultural and material developments. However, while according to the community members of Boca Inambari there is bilingual education in the community's elementary school, no bilingual teacher was present during the period in which I was conducting my fieldwork.

1.4 Fieldwork methodology

Data for the present research was collected during a field trip in the Peruvian department of Madre de Dios in July-September 2010. After contacts had been established with the indigenous organization FENAMAD[3] in the departmental capital Puerto Maldonado, fieldwork was conducted in the indigenous community of Boca Inambari (also known as Amarakaeri[4]). This small settlement at the confluence of the Madre de Dios and Inambari rivers is the easternmost of the Arakmbut communities, and the closest to Puerto Maldonado. The community consists of some 30 families, with a total of around 110 inhabitants. It was established in 1973, by several families that had fled from the Dominican mission of El Pilar.

I worked with three main consultants in the community:

- Manuel Kameno, 63, president and founder of Boca Inambari
- Rosa Manuaje, 45, Manuel's wife
- Cornelio Iricarine, 68, one of the few elders in the community

Since I had put my tent up on the veranda of Manuel and Rosa's house, I was in frequent contact with my consultants, and they were nice enough to teach me about their language in daily elicitation sessions. Towards the end of my stay, however, preparations for the community's anniversary party took up most of my hosts' time, and I relied more and more on the knowledge of Cornelio, who turned out to be one of the most educated people in the community. All three consultants gave me their consent to record the sessions and to use these recordings for research purposes.

[3] Federación Nativa del Rió Madre de Dios y Afluentes

[4] While both names are used among the community members themselves, 'Amarakaeri' seems to enjoy preference, for being a purely indigenous name. In this thesis, however, I will reserve the term 'Amarakaeri' for the language, and use 'Boca Inambari' to refer to the community, in order to avoid ambiguity.

I recorded all of my sessions on a Zoom H2 recorder with a built-in microphone. The files were recorded in WAV format in order to ensure a high level of sound quality. However, since a community is hardly comparable to a sterile audio studio, some of the recorded interviews are interspersed with animal sounds, children playing in the background, and sometimes the sound of a distant generator. While this does not seem to have influenced my research negatively, it does make my recordings less suitable for archiving.

My main method of data collection was elicitation, consisting mainly of translation and grammaticality judgements. I would frequently ask my consultants to translate words and short sentences from Spanish into Amarakaeri, and I also formed words myself (for instance, combining nouns with shape morphemes), asking them whether these words could be used in the language. On several occasions, I asked my consultants to analyze words or to give me the specific meaning of certain morphemes (especially prefix *wa-*), but this approach was usually met with incomprehension and impatience. While elicitation and grammaticality judgements as a main strategy seemed the best option for me, considering my specific research interest and the short period of time that I could spend in the community, this method did have its limitations. Recording more natural speech, in the form of stories or conversations, has the benefit of producing more useful data for the study of syntax and discourse pragmatics. This, however, will have to remain for future research.

1.5 Main characteristics of the Amarakaeri language

While the limited scope of this thesis does not permit me to provide an extensive grammar sketch of Amarakaeri, some basic characteristics and distinctive features of the language will be discussed in this section.

Amarakaeri has a phonemic distinction between oral and nasal vowels; it has five of each group: /a/, /e/, /i/, /o/, /u/[5] and /ã/, /ẽ/, /ĩ/, /õ/, /ũ/. This feature of Amarakaeri phonetics also influences the use of some of the consonants, i.e. those which have nasalized, prenasalized and postnasalized allophones. In a syllable with a nasal vowel, the consonant will be nasalized (for instance, **n**), while an oral vowel in the syllable triggers the use of the prenasalized allophone (e.g. **nd**) in prevocalic position, and the postnasalized allophone in postvocalic position (e.g. **dn**)[6] (Adelaar 2000: 230). This is illustrated in example (1), where phoneme /n/ appears in two different forms, according to its position with regard to the (oral) vowel:

(1) ***nd****o* *-e****dn*** *-hak*[7]
1SG -POS-house
'my house'

[5] The Amarakaeri /u/ represents the high central vowel [ʉ]; the language does not have a phoneme which represents the high back vowel [u].
[6] Different ways to deal with this phenomenon in transcription and orthography will be discussed in section 5.6.
[7] All of the examples in this section are taken from my own data.

With regard to morphology, Amarakaeri can be considered an agglutinative language. It has both prefixes and suffixes. Nominal morphology is rather simple; there is no number marking, and nominative and accusative marking are optional. Genitive marking, however, is obligatory (Tripp 1995). The language does have quite a wide array of locative and directional suffixes, which are similar to postpositions. Articles do not exist in the language.

As was mentioned before, the main topic of this thesis will be the shape morpheme system, which forms a distinctive feature of Amarakaeri morphology. The shape morphemes constitute a relatively large set of morphemes, which may refer to the shape, and in some cases to other qualities, of the word they are attached to. The shape morphemes can also be incorporated into verbs, where they can have different functions and positions. Many of these morphemes seem to originate in body part terms (Hart 1963, Helberg 1984). Some examples of the different uses of shape morphemes are given in examples (2) - (4) below:

(2) *wa* ***-po***
NMLZ **-SHM:round/box.like**
'round object, ball'

(3) *wa* ***-ku*** ***-chi*** ***-po***
NMLZ **-SHM:head** **-SHM:extension** **-SHM:round/box.like**
'thigh'

(4) *Pera* ***-po*** *o* ***-po*** *-yakay.*
rubber **-SHM:round/box.like** 3SG **-SHM:round/box.like** -kick
'He kicks the ball.'

In the remainder of this thesis, these highly interesting morphemes will be further discussed with regard to their nature, function, contexts of use, degree of semantic generalization, occurrence in lexicalizations, and productivity in word formation. Using this information, attempts will be made at finding and understanding their origins and development.

Another interesting characteristic of Amarakaeri morphology is that many nouns, including those which are formed with shape morphemes, carry prefix *wa-*. This is illustrated in example (5):

(5) ***wa*** *-mba'*
NMLZ -SHM:hand/leaf
'hand'

This prefix, which Helberg (1984: 189) defines as a 'prospective infinitive', seems to play a very important role in word formation. Interestingly, however, there are also many nouns which occur without prefix *wa-*. The function, distribution and possible origins of this prefix will also be discussed in several of the following chapters, for the prefix may be related to the system of shape morphemes.

In contrast with the nominal morphology of the language, the verbal morphology is relatively complex. Amarakaeri has an elaborate verbal cross-referencing system in which both subject and object are marked on the verb by prefixes and suffixes which indicate person and number. This type of system resembles verbal cross-referencing systems of Andean languages such as Quechua (Adelaar, p.c.). There are different sets of person marking affixes for declarative, dubitative and imperative mood (Tripp 1976a: 15). Furthermore, Amarakaeri verbs are marked for aspect. Helberg (1984: 276) distinguishes 'primary' aspect, which is integrated with the person/mood system, and 'secondary' aspect, which is marked by separate aspect suffixes. The three main tenses are present, past and future; past tense is further divided into recent and remote past, and is also marked for evidentiality (own experience / hearsay) (Helberg 1984). The basic word order of Amarakaeri is SOV.

1.6 Notes on orthography

As was mentioned in the previous section, Amarakaeri phonology has a system in which consonants can be pre- and postnasalized, in correspondence with the position and nature of the vowel in the syllable. Such a system poses a challenge for linguistic transcription and orthography development. This is illustrated below, where different representations of the prenasalized vowel in the Amarakaeri word for 'hand' by different authors are shown:

Hart (1963)	Tripp (1995) / SIL	Helberg (1984)	FENAMAD (2006)
*wa**m**a'*	*hua**b**a'*	*wa^{m}**b**a'*	*wa**mb**a*

One the one hand, the nasal, prenasalized and postnasalized consonants are allophones of the same phoneme, and could therefore be considered unworthy of distinct representation in the transcription. Hart (1963) and Tripp (1995), as well as the SIL bilingual teaching materials (e.g. Moqui 1974) seem to have followed this rationale, representing the different allophones of these phoneme by one and the same letter. Hart (1963) has chosen to use only the nasal consonant (m, n, ŋ) while Tripp (1995) and the SIL represent the same set of sounds by the plosive variant (b, d, g). Although theoretically well-grounded, it can be argued that this method of transcription does little justice to the actual sounds of the language; when one hears Amarakaeri speech, the pre- and postnasalized consonants are among the most striking features of the language's overall 'sound'. Helberg (1984) has opted for a more representative transcription, while still conserving the theoretical side of the matter, by adding the pre- and postnasalization of consonants in superscript. However, using superscript in orthography seems less ideal for practical reasons. In 2006, a standardized orthography of Harakmbut[8] which had been proposed by the indigenous organization FENAMAD was approved by the Dirección Nacional de Educación Intercultural Bilingue y Rural, which is part of the Peruvian Ministry of Education (DINEIBIR-DEIB 2006). In this orthography, pre- and postnasalized allophones are presented as independent sounds, and are represented by letter combinations such as **dn** and **nd**. This solution seems to be the most practical one, since it is easy to work with on computer keyboards. In the same fashion, palatalized allophones of /t/ and /s/ are

[8] It must be noted that this is a standardized orthography which is intended to be used for all Harakmbut varieties, not just Amarakaeri.

written with letter combinations (**ch** and **sh**). Also, nasal vowels are represented by adding dieresis, which is practical from a computer user's point of view.

Throughout this thesis, I will transcribe Amarakaeri sounds according to the 2006 FENAMAD orthography. While not ideal from a theoretical point of view, it seems to me the most representative and practical of the options stated above. More importantly even, it is the orthography which was chosen by the speakers themselves as the best way to represent their language. In Table 1 below, this orthography is presented.

TABLE 1
TRANSCRIPTION OF AMARAKAERI PHONEMES
(after DINEIBIR-DEIB 2006)

VOWELS

Phoneme	Transcription
/a/	*a*
/ã/	*ä*
/e/	*e*
/ẽ/	*ë*
/i/	*i, y*
/ĩ/	*ï*
/o/	*o, w*
/õ/	*ö*
/u/	*u*
/ũ/	*ü*

CONSONANTS

Phoneme	Transcription
/'/	'
/h/	*h*
/k/	*k*
/m/	*m, mb*
/n/	*n, nd, dn*
/ɳ/	*ñ*
/ŋ/	*ng, gn*
/p/	*p*
/r/	*r*
/s/	*s, sh*
/t/	*t, ch*

2. Nominal classification

In this chapter, the phenomenon of nominal classification will be introduced. Section 2.1 will provide a brief introduction to nominal classification and a description of the main concepts and terminology related to this topic. The typological challenges that have marked the debate in the past decades will be touched on in 2.2. Finally, in 2.3, theories about the origins and development of nominal classification systems will be presented.

2.1 Nominal classification: an introduction

As is noted by Senft (2000: 11), the human mind needs classification devices in order to function properly; either consciously or unconsciously, we are constantly categorizing the objects, beings, concepts and phenomena that we perceive in the world around us. It is therefore not surprising that many of the languages in the world possess some means of noun categorization. These systems of nominal classification organize and group together the referents of nouns on the basis of their inherent properties. A well-known example is the gender system of Spanish, which marks the gender of the head noun on its determiners and modifiers, as is shown in example (6):

(6) *la* *mujer* *vieja*
ART.DEF(F) woman(F) old(F)
'the old woman'

While Indo-European nominal classification systems usually categorize the noun on the basis of gender (Masculine, Feminine, and in some cases Neuter), many other languages have noun categorization devices which function on the basis of other semantic categories. For instance, the Eastern Tucanoan language Tuyuca has a classification system which categorizes nouns according to their shape, dimensionality, arrangement, and several other categories (Barnes 1990). This is illustrated in example (7) below:

(7) *atí* *-hoó* ***-poro***
DEM -banana **-CL:cylindrical**
'this banana'

(Barnes 1990: 281)

Apart from being a linguistic reflection of the classification devices that exist in the human mind, systems of nominal classification often play an important role in discourse, as referent tracking devices. In many languages, class markers are used as anaphora, which make the overt expression of the noun unnecessary and express reference without obstructing the information flow (Derbyshire & Payne 1990, Grinevald 2000). Furthermore, class markers may function as nominalizers, relativizers, derivational affixes, and markers of definiteness or individuality (Senft 2000: 26).

The overarching term that will be used for noun categorization devices in this thesis is 'systems of nominal classification', following Senft (2000) and Grinevald (2002). The most widely accepted definition of the individual classifying elements in these systems is given by Allan (1977: 285), stating that 'classifiers[9] occur as morphemes in surface structure under specifiable conditions', and that they denote 'some salient perceived or imputed characteristic of the entity to which an associated noun refers'. It must be noted, however, that while all systems of nominal classification share these properties, they differ from one another in many other respects. Some of the parameters that are used to distinguish different types of nominal classification systems are the morphosyntactic locus that the class markers occur in, the degree of grammaticalization of the system, its semantic organization, and whether the system shows agreement or not (Aikhenvald 2000a). The different types of systems that occur cross-linguistically, and the challenges these systems pose to the establishment of a typology, will be discussed with more detail in Section 2.2.

The phenomenon of nominal classification is found in a wide range of languages from all over the world. However, some types of systems are more frequent in certain areas or language families than in others, such as numeral classifiers in East Asia, or gender systems in Indo-European languages. This predominance of certain types of systems in certain areas is often due to areal diffusion rather than common origin (Dixon 1986, Aikhenvald 2000a). Nominal classification is often highly culture-specific, and may provide interesting insights into how speakers of different languages categorize the world around them along different semantic lines (Aikhenvald 2000a: 5). However, this is not necessarily the case. An example that is often cited in this context is the noun class system of the Australian language Dyirbal, in which feminine human beings are grouped in the same noun class as fire and fighting, and other things that are associated with danger (Lakoff 1987); this is said to tell us something about Dyirbal speakers' world view and culture. However, Plaster & Polinsky (2007) argue that noun class assignment in Dyirbal is in fact more formally than semantically based, and that the categorization of females, fire and fighting in the same class has very little to do with cultural concepts. This example is instructive, in that it shows us to be careful when relating categories of nominal classification to speakers' conceptual categorization of the actual world, and to refrain from exoticizing languages and their speakers by overemphasizing typologically rare linguistic features.

2.2 Classifying nominal classification: typological challenges

Ever since systems of nominal classification started to be studied typologically, about three decades ago, their typological categorization has been the subject of debate. This lack of consensus is reflected by a terminological confusion in descriptive and typological works[10], which complicates cross-linguistic comparison (Aikhenvald 2000a: 1). One of the first attempts at establishing

[9] Allan (1977) and Aikhenvald (2000a) use 'the term 'classifiers' as an umbrella term to designate the individual elements of noun categorization devices. However, considering the fact that the word 'classifiers' is also used as a denomination of a subclass of this group (cf. 2.2), I will be using the term 'class markers', following Seifart (2005).

[10] See Aikhenvald (2000a: 8-10) for a discussion on the use of different terminology within different linguistic traditions.

typological categories for nominal classification is presented by Allan (1977). Based on a sample of 50 languages that possess nominal classification, Allan distinguishes four main types of 'classifier languages':

- *numeral classifier languages*, in which class markers occur in expressions of quantity;
- *concordial classifier languages*, in which the class markers appear on the noun and its modifiers;
- *predicate classifier languages*, in which the class markers occur on the verb;
- *intra-locative classifier languages*, in which class markers are embedded in locative expressions.

As we see, these Allan's typology is mainly based on the morphosyntactic environment that the class markers occur in. However, as is pointed out by Dixon (1986), morphosyntactic locus should not be the only criterion to distinguish different types of nominal classification, for the noun phrase may host two very different types of class markers: noun class systems[11] and noun classifier systems. The criteria which Dixon uses to distinguish the two types are presented in Table 1 below:

TABLE 2

CHARACTERISTICS OF NOUN CLASS AND NOUN CLASSIFIER SYSTEMS

(after Dixon 1986)

Noun class	**Noun classifier**
relatively small number of classes (2-20)	relatively large number of classes (100 is common)
all nouns are classified	not all nouns are classified
nouns belong to only one class	nouns may occur with several classifiers
closed grammatical system	open system
can form a morphological unit with the noun	never forms a morphological unit with the noun
often forms a portmanteau morpheme with case and number	is not fused with case and number

[11] In the literature, gender systems are usually considered a subtype of noun class systems (Dixon 1986, Grinevald 2000).

agreement	no agreement
obligatory system → little variation between speakers	variation between speakers is possible

Dixon considers numeral classifiers a subtype of noun classifiers, but signals that verbal classifiers do not fall clearly into one of the categories that he has established (Dixon 1986: 107). This illustrates the main shortcoming of Dixon's typology: although it applies to 'prototypical' systems, such as the noun class systems of Bantu languages and the noun classifiers which are typical of South East Asian languages, it has proven to be problematic for many other languages. In particular, new data on nominal classification in Amazonian languages, which has become available in recent decades, has shown that many of the Amazonian systems do not fit into either of Dixon's categories (Aikhenvald 2000a, Grinevald & Seifart 2004). In an overview of nominal classification in the Amazon region, Payne (1987) and Derbyshire & Payne (1990) argue that many Amazonian languages combine characteristics of several of the system types that have been proposed by Allan (1977) and Dixon (1986), and that most of these systems cannot be labelled as one discrete type. While this deviation from the pre-established categories might feed the temptation to define Amazonian nominal classification systems as 'aberrant' or 'exotic', Grinevald & Seifart (2004: 244) rightly argue that these systems are no more exotic than others; they are simply the ones which were most recently encountered. In recent years, therefore, authors such as Aikhenvald (2000a) and Grinevald (2002) have proposed a revision of the typological parameters which are used in the description of nominal classification systems, in order to grant the Amazonian systems a more legitimate place within the typology. Departing from the idea that systems of nominal classification are constantly evolving, and that one type of system may develop into another, the aforementioned authors use a 'prototype-continuum approach' (Aikhenvald 2000a) or 'grammaticalization perspective' (Grinevald 2002), which places different systems along a continuum of grammaticalization, instead of trying to fit them into already established categories. Within this framework, many of the Amazonian noun categorization devices can be accommodated as emergent noun class systems; although they show a much lesser degree of grammaticalization than the fully grammaticalized gender or noun class systems such as exist in Bantu languages, the differences between these systems can be considered as a matter of degree rather than essence (Grinevald & Seifart 2004: 282). While the prototypical noun class systems appear at the most grammaticalized end of the continuum, the lexical end is represented by class terms (e.g. the morpheme *-man* in words such as *mailman* and *policeman*), and measure terms (e.g. 'a *slice* of bread', 'a *cup* of coffee'), which exist in most languages of the world (Grinevald 2002: 260). This model is illustrated in Figure 1 below:

FIGURE 1

SYSTEMS OF NOMINAL CLASSIFICATION ALONG THE GRAMMATICALIZATION CONTINUUM

after Grinevald (2002: 260)

< Lexical----------Grammatical >

class terms
measure terms

classifiers

noun classes

As is shown in Figure 1, halfway along the spectrum lies the category which is commonly denominated as 'classifiers' (Aikhenvald 2000a, Grinevald 2002). A more detailed typology of this fairly broad category is provided by Aikhenvald (2000a) in her book *Classifiers: a typology of noun categorization devices*, in which the different types of classifiers are categorized on the basis of the morphosyntactic locus which they occur in. Aikhenvald (2000a) distinguishes five main classifier types[12], which will be briefly described below:

- *Noun classifiers* are free morphemes which occur within the scope of the noun phrase; they are not involved in agreement. The choice of a noun classifier is semantically based, and in some cases different classifiers can be used with the same noun, resulting in a change in meaning. The properties of this type of classifier largely coincide with those of the category 'noun classifier' that was established by Dixon (1986), although Aikhenvald (2000a: 81) argues that noun classifiers are not necessarily part of a large set, and that they can be grammaticalized to varying extents. Noun classifiers are widespread among Australian and Western Austronesian languages. In the Americas, noun classifiers occur in Mesoamerica and in a few Makú and Jê languages.

- *Numeral classifiers* appear in numeral noun phrases and in expressions of quantity. Just like noun classifiers, they are not involved in agreement. Moreover, numeral classifiers are usually chosen on a semantic basis. Numeral classifiers are, after noun classes, the most frequent type of nominal classification system, and occur in many languages of Southeast Asia and Oceania, North and Central America, and in several Amazonian language families.

- *Classifiers in possessive constructions* (denominated *genitive classifiers* by Grinevald (2002)) have three different subtypes. They may categorize either the possessed noun (*possessed classifiers*), the possessor (*possessor classifiers*), or categorize the semantic nature of the relation between the two (*relational classifiers*). These classifier types are relatively uncommon cross-linguistically; they mainly occur in Oceanic languages and in some languages of North and South America.

- *Verbal classifiers* appear on the verb. They serve to categorize the referent of one of its arguments, which may or may not be overtly expressed by a noun, in terms of its inherent properties such as shape, size and consistency. Their choice is thus mainly semantic. Aikhenvald (2000a: 149) distinguishes three ways in which verbal classifiers may be

[12] The same classifier types, except for the locative and deictive classifiers, are mentioned in Grinevald (2000, 2002).

realized: by means of *classificatory noun incorporation*, as *verbal affixes*, or as suppletive verbs which in themselves categorize the argument in terms of its inherent properties (*classificatory verbs*). Verbal classifiers occur mainly in the Americas, including in some Amazonian languages, and in a few languages of northern Australia.

– *Locative and deictic classifiers* appear in locative noun phrases and on deictic modifiers, respectively. These types of classifiers are quite rare in the languages of the world. Locative classifiers have been reported to occur in some Arawak and Carib languages; deictic classifiers occur in a few North American languages, such as Eskimo.

Although this typological framework, which combines the grammaticalization perspective with a morphosyntactically driven subcategorization of the group of classifier systems, manages to accommodate most of the nominal classification systems in the languages of the world, it does not account for some of the more complex systems, which are quite common in Lowland Amazonia. In many languages of that region, different classifier types not only coexist, but in fact share a single set of classifier morphemes which appears in many different morphosyntactic environments (Aikhenvald 2000a). For instance, the Tupí language Mundurukú has a large set of classifiers which are used as verbal classifiers, numeral classifiers, deictic classifiers, and noun class markers, and function both as agreement markers and derivational affixes. Aikhenvald (2000: 204) denominates the languages which show this type of system 'multiple classifier languages' and distinguishes these from the languages in which different classifier types coexist but each have their own set of classifiers. Languages that show only a certain amount of overlap in the morpheme inventories of their different classifier sets are considered 'fuzzy types', or systems in transition (Aikhenvald 2000a: 230). In contrast, Seifart (2009: 351) criticizes the concept of 'multiple classifier language' for failing to identify these complex systems as being coherent. Indeed, the term may be a bit problematic, since it seems to imply that systems of this type consist of multiple classifier types, while it might be more accurate to say that these languages challenge the entire notion of different classifier types, by using a single set of classifiers in multiple environments. Grinevald & Seifart (2004: 260) take this discussion one step further, stating that some of the 'multiple classifier' systems might have to be analyzed as noun class systems with agreement markers on multiple targets, rather than as a combination of several classifier types. In the light of this ongoing debate about use of terminology and the establishment of typological categories, Seifart (2009) proposes the use of a 'multidimensional typology', which consists of more detailed and more varied parameters and lacks any pre-established types, in order to obtain a more fine-grained image of every individual language and nominal classification system. Considering the fact that all languages are different and no two systems of nominal classification share exactly the same properties, this approach seems hardly superfluous.

In this thesis, as was mentioned before, I will also focus on the way in which the Amarakaeri nominal classification system might have developed. In this light, the grammaticalization perspective which is used by Aikhenvald (2000a) and Grinevald (2002) will prove useful. Especially in the context of Amazonian languages, which tend to undergo relatively rapid language change due to frequent migration and extensive language contact (Adelaar, p.c.), a framework in which a language is primarily perceived as an evolving entity seems to be the most accurate one. However, in my research on nominal classification in Amarakaeri I will depart from the idea that

not all systems of nominal classification are necessarily undergoing a unidirectional process of grammaticalization, and I will not exclude the possibility that the Amarakaeri system might have moved in either direction - or in both directions - along the grammaticalization continuum. With regard to terminology, I will adopt most of the terms that are used in Aikhenvald (2000a), since this is so far the most comprehensive and detailed work on nominal classification. I will thereby keep in mind that the typological categories which Aikhenvald uses represent focal points along a continuum, rather than completely separate categories (Aikhenvald 2000a: 434). Therefore, upon encountering phenomena that do not seem to fit into the existing typology, I will focus on a detailed description and cross-linguistic comparison of these phenomena, rather than trying to assign them to one of the already established categories.

2.3 The emergence and evolution of nominal classification systems

In the light of the typological debate, and especially within the framework of the grammaticalization continuum, it is highly interesting to investigate how systems of nominal classification originate and develop. According to Mithun (1986: 388), 'all classificatory stems begin life as nouns'. Although this has proven to be an oversimplification, since other word classes such as verbs and determiners may also give rise to these systems (Aikhenvald 2000a: 353), nouns indeed seem to be the most common source for the development of nominal classification. The process in which lexical items develop into class markers can be considered an instance of grammaticalization, since it involves a transformation of the lexical item whereby it acquires a more abstract and grammatical meaning. It is widely assumed that grammaticalization processes are marked by unidirectionality: once a linguistic entity has grammaticalized, the process is irreversible (Heine & Kuteva 2009, Aikhenvald 2000a). However, although most of the emerging nominal classification systems share the same point of departure (lexemes) and direction of development (towards grammaticalization), they differ cross-linguistically in a wide range of respects. First of all, different types of lexical sources seem to give rise to distinct types of systems, involving different kinds of semantic changes. This will be discussed in 2.3.1. Secondly, as will be shown in 2.3.2, the development from lexeme into class marker may follow several different paths, which pass through distinct intermediate stages along the grammaticalization continuum. Thirdly, systems of nominal classification in different languages show different levels of dynamism and development; this will be discussed in 2.3.3. Finally, the evolution - and decay - of nominal classification systems is often influenced by language contact, which may bring about different types of changes in these systems. These will be touched on in section 2.3.4.

2.3.1 Development of lexical items into classifiers

As was mentioned before, class markers are most commonly derived from nouns. Which semantic subgroup of nouns becomes grammaticalized as class markers is often specific to language families and linguistic areas (Aikhenvald 2000a: 353). However, some sets of nouns can be distinguished as frequent and typical sources of class markers. Aikhenvald mentions five main groups:

- *body part nouns*
- *kinship terms and nouns referring to humans and higher animates*
- *generic nouns*
- *unit counters*
- *culturally important items*

Each of these subgroups tends to give rise to a different type of system (Aikhenvald 2000a: 353). Body part nouns frequently grammaticalize into verbal classifiers, through classificatory noun incorporation (Mithun 1986: 391), but may also become numeral or locative classifiers. In contrast, nouns referring to kinship, humans and higher animates usually give rise to noun classifiers, which may further grammaticalize into noun class systems. Generic nouns (such as 'person', 'bird' or 'tree') are reported to develop into noun classifiers, verbal, numeral and possessed classifiers (Aikhenvald 2000a: 360), while unit counters (e.g. 'heap', 'handful') are only found to become numeral classifiers. Finally, culturally important items may give rise to any type of nominal classification system (Aikhenvald 2000a: 361).

Other word classes may also be the source of class markers, although this is relatively rare cross-linguistically. In some languages, posture and motion verbs develop into classificatory verbs, while they become deictic classifiers in others. Verbs of handling may develop into classifiers that appear in possessive constructions, or they may be the source of numeral classifiers (Aikhenvald 2000a: 362). Closed classes, such as demonstratives or locative adpositions, may also develop into systems of nominal classification. Items from these classes tend to grammaticalize into closed noun class systems, rather than into larger, open systems of classifiers. Furthermore, Aikhenvald (2000a: 366) signals that some languages have a nominal classification system of mixed origins; a single set of classifiers may have been derived from different lexical classes (e.g. nouns and verbs).

The process of grammaticalization in which lexical items become class markers tends to go hand in hand with semantic change. The most commonly attested semantic change associated with this process is that of specific > generic, or concrete > abstract (Aikhenvald 2000a: 401). That is, nouns with a specific reference usually develop into class markers with a broader or more abstract meaning. For instance, in the Iroquoian language Cayuga, the noun stem for 'car' has grammaticalized into a classifier that is used with all types of vehicles (Aikhenvald 2000a: 403). According to Heine & Kuteva (2009: 166), this kind of change is an instance of desemanticization (also known as semantic bleaching): during its transformation into a classifier, the lexical source loses some of its semantic content, thereby becoming applicable to a wider range of referents. However, grammaticalization of lexemes into class markers does not always involve desemanticization. Nouns which already have a generic reference, such as generic nouns, tend to develop into generic noun classifiers or noun classes, keeping more or less the same range of referents. Moreover, Aikhenvald (2000a: 403) mentions cases in which the classifier has acquired a more specific meaning than its source noun, such as the classifier for 'old man, respectfully' in the Mayan language Mam, which is derived from the generic noun for 'man'.

While some of the semantic content is lost in the process of desemanticization, the semantic value of a lexeme that is being grammaticalized is often extended into a specific context, serving as the basis for semantic generalization. Different types of such semantic extensions can be discerned.

Aikhenvald (2000a: 404) distinguishes three main types: extensions by material makeup (e.g. water > all drinkable liquids), extensions by function (e.g. car > all vehicles), and extensions by shape (e.g. tree > vertical things). Extensions which involve a combination of these types are also attested. According to Grinevald (2000: 71), there often seems to be a correlation between the type of class marker and the semantic domain into which it is extended: while numeral classifiers often express physical categories such as shape, genitive constructions tend to classify nouns according to function, and noun classifiers are usually extended into the domain of material makeup and essence (e.g. animacy).

2.3.2 Grammaticalization paths

Although the development of lexemes into classifiers and noun classes always seems to involve grammaticalization, there are different paths through which this grammaticalization process may take place. These paths can be distinguished mainly on the basis of the intermediate stages that they entail, the main types being *compounding*, *repeaters*, and *noun incorporation*. These three grammaticalization paths will be discussed below.

Compounding is a productive process in most languages of the world. The defining property of this process is the combination of lexemes into larger words. In its most straightforward form, compounding involves the combination of two words, one of which semantically modifies the other (Booij 2005: 75). An example of this is the compound *apple tree*, where *apple* modifies *tree*, and the compound as a whole denotes a certain type of tree. As is noted by Heine & Kuteva (2009: 163), this type of compounds - coined 'modifying compounds' - often develops into noun-classifier (N + CL) combinations. A possible explanation for this is that in many noun-noun combinations one of the nouns assumes a more general semantic value, which, as we saw before, is one of the semantic changes that may be involved in the development of nouns into classifiers. This type of grammaticalization is common in Australian languages, many of which have compound-like structures of generic and specific nouns as a stylistic or discourse device (Aikhenvald 2000a: 372). This is illustrated in example (8) from the Pama-Nyungan language Yidiny:

(8)

bamaal	*yaburuNgu*	***minya***	***gangu:l***	*wawaal*
Person-ERG	girl-ERG	**animal-ABS**	**wallaby-ABS**	see-PAST

Lit.: 'the person girl saw the animal wallaby'

(Dixon 1982: 186, cited in Grinevald 2002: 268)

As is argued by Grinevald (2002: 269), this type of generic-specific pairings can be considered a form of emergent noun classifier systems, and they should be placed at the most lexical end of the grammaticalization continuum. In the course of time, these systems may further develop into more grammaticalized systems of nominal classification, as the use of the generic noun becomes obligatory rather than being a stylistic device. Subsequently, the generic noun may undergo phonological reduction and further grammaticalize into a noun class affix. This has happened in languages such as Dyirbal, which now has a closed grammatical noun class system that originated as a system of generic-specific compounding (Aikhenvald 2000a: 372). While the generic-specific pairings seem to be an areal feature in Australia, many languages in other parts of the world show similar patterns. As is pointed out by Aikhenvald, it is often difficult to determine whether these

'generic nouns' should be considered classifiers or nouns; the obligatoriness and predictability of use of the generics is usually the decisive factor (Aikhenvald 2000a: 87). The idea that all these systems constitute points on the grammaticalization continuum rather than forming part of any pre-established category, should be kept in mind here.

Repeaters constitute another intermediate stage between nouns and classifiers. They are best defined as nouns which may appear in slots that classifiers usually occur in. Repeaters may either classify 'themselves', or semantically closely related nouns. They serve as 'ad hoc classifiers' in a number of languages. Some repeaters occur as a truncated form of the noun; this type is coined 'partial repeaters' (Aikhenvald 2000a: 361). Repeaters form a clear example of an intermediate stage between noun and classifier, particularly since they exist alongside full nouns. That is, one and the same item may be used both as an independent noun and as a repeater synchronically, even within a single sentence. An instance of this is presented in example (9) from the Bora-Witotoan language Miraña. In this language, only classifiers tend to occur suffixed to the copula verb and to numerals. The 'nouns' in these slots can therefore be considered repeaters. The independent noun, which appears at the end of the sentence, is optional in this construction.

(9) *ó -ʔdi ihka* ***-báhu*** *tsá* ***-báhu*** *(báhu)*
1SG-POS COP-**RP:forest** one **-RP:forest** (forest)
'I have one (stretch of) forest.'

(Seifart 2005: 94)

What distinguishes repeaters from classifiers in Miraña is that the former have undergone no or very little semantic generalization, and are thus almost identical to nouns in terms of their semantic value (Seifart 2005: 95). In this sense, they seem to occur at the more lexical end of the grammaticalization continuum - Seifart (2005: 95) considers them as a phenomenon 'at the margin of nominal classification'. However, repeaters may grammaticalize into classifiers, and even into noun classes. In Miraña, for instance, which has an extensive system of nominal classification, most classifiers seem to have entered the system as repeaters, before undergoing semantic generalization and phonological reduction. In this process, repeaters lose the ability to occur as independent nouns (Grinevald & Seifart 2004: 279), a development which marks their complete transition into grammaticalized classifiers.

Noun incorporation is another path that nouns may take in the course of grammaticalization. It can be defined as a process in which a noun and a verb are combined into a verbal compound (V + N); usually, this produces verbs that express an institutionalized activity[13]. Furthermore, incorporated nouns are non-referential, in that they do not refer to any object in particular (Booij 2005: 92). How the process of noun incorporation may develop into a system of nominal classification is extensively described by Mithun (1986). First, noun incorporation may come to play an important role as a referent tracking device in certain languages. After a referent has been introduced in a noun phrase, noun incorporation provides a means of keeping the reference clear without interrupting the information flow, by simply narrowing the scope of the verb (Mithun 1986: 381). However, as noun incorporation comes to play an increasingly important role in the language, some

[13] A similar process occurs in English, resulting in institutionalized verbs such as *rock climbing* and *tooth brushing*.

of the V + N compounds may be used so frequently, and may become so well established, that they cease to fulfil their task as referent tracking devices. It is at this point that the argument of the verb, which is implied by the incorporated noun, begins to be expressed overtly. This type of system, in which an incorporated noun somehow characterizes an argument of the verb, which is also referred to by a noun phrase in the same sentence, is defined as 'classificatory noun incorporation' (cf. 2.2). There is often a generic-specific relationship between the incorporated noun and the overtly expressed noun. An example from the Northern Iroquoian language Cayuga is given in (10) below:

(10) ***So:wá:s*** *akh**náhsk**waę'*.
dog 1SG-**CL:domestic.animal**-have
'I have a (pet) dog.'

(Mithun 1986: 387)

The final stage of this development, as is described by Mithun (1986: 390) occurs when verb-incorporated classifiers shift from indicating kinds of entities, as we saw in (10), to indicating qualities, such as shape, size or consistency (Aikhenvald 2000a: 150). This can be illustrated with an example from the North American language Caddo (11):

(11) ***Kapí:*** ***kan**čâ:ni'ah*
coffee **CL:liquid**-buy-PAST
'He bought (liquid) coffee.'

(Mithun 1986: 386)

As we see in (11), the object of the verb ('coffee') is classified by the incorporated verb on the basis of its consistency ('liquid'). Interestingly, as is noted by Mithun (1986: 390), the verb-incorporated classifier which serves to qualify a certain semantic group is usually not chosen for being its most prototypical member; rather, it is selected simply for being the most frequently incorporated noun of the group. For instance, lexical items that refer to ubiquitous objects or phenomena, such as 'water', will be frequently used in the language, and will therefore appear in a large number of frequently occurring V + N compounds; this is the main cause for the grammaticalization of the noun for 'water' into the classifier that qualifies all liquids. For the same reason, verb-incorporated classifiers often find their origins in body part nouns, since these are among the most frequently incorporated nouns[14] (Mithun 1986: 393). As will be shown in Chapter 3, these body part based classifiers often qualify their referents in terms of shape.

When the process of noun incorporation ceases to be productive in a language, this may cause the classifiers to lose their semantic transparency. The set of verb-incorporated classifiers then becomes closed, and often semantically opaque, but continues to be used (Mithun 1986: 392).

[14] Aikhenvald (2000a) uses the term 'body part incorporation' to refer to this type of classificatory noun incorporation.

2.3.3 Stage of development of nominal classification systems

Regardless of which grammaticalization path they follow, systems of nominal classification find themselves in different stages of development cross-linguistically. While some systems are young and emerging, others are relatively old (Grinevald 2000: 84). Furthermore, some systems are in decay; this is often related to language obsolescence and areal diffusion (cf. 2.3.4). Although it is difficult to determine the (relative) age of nominal classification systems, some indicators can be distinguished.

First of all, systems which are semantically transparent, and whose class markers seem to have a clear lexical origin, are assumed to be relatively new (Aikhenvald 2000a, Grinevald 2000). In contrast, semantic opacity and an advanced degree of grammaticalization are indicative of old systems. Phonological reduction of class markers also indicates a relatively high age. Systems in decay tend to be marked by simplification and the loss of some of the class markers, rather than the loss of the classifier construction itself (Grinevald 2000: 85). As is argued by Aikhenvald (2000a), within systems of nominal classification that employ different sets of class markers in several morphosyntactic loci, the different sets often show different stages of development and grammaticalization.

Furthermore, Grinevald (2000) argues that systems may show different levels of dynamism; while some systems still create new class markers, others seem to be 'frozen', with a closed set of class markers. According to Grinevald (2000: 84), the level of dynamism is independent of the relative age of the system. As will be discussed in the next section, the dynamism, direction and stage of development of nominal classifier systems not only depends on language-internal factors; it may also be influenced by language contact and areal diffusion.

2.3.4 The influence of language contact on nominal classification

Most languages of the world are not isolated entities; they are under constant influence of other, often neighbouring, languages, which tends to result in lexical, grammatical and phonological changes. It is therefore not surprising that language contact may also leave its mark on systems of nominal classification. Usually, it is the dominant language - either in terms of prestige, or with regard to speaker numbers - which exercises influence on the other language (Aikhenvald 2000a: 382). This influence may take two different forms: direct diffusion (i.e. borrowing of lexemes or forms) or indirect diffusion (i.e. borrowing of certain patterns or structures). In the context of nominal classification systems, indirect diffusion is much more common than direct diffusion (Aikhenvald 2000a: 383, Seifart 2007: 440).

Language contact may produce different types of changes in systems of nominal classification, depending on the structural properties of the languages that influence each other. According to Aikhenvald (2000a), systems of nominal classification may be either created, restructured, or reduced under influence of language contact. *Creation* of nominal classification tends to occur when the dominant language has a system that is closer to the lexical end of the grammaticalization continuum; diffusion of closed grammatical systems is quite rare. As Aikhenvald argues, however, it is often difficult to establish the direction in which diffusion took place, since systems of nominal classification tend to be areal features (Aikhenvald 2000a: 384). *Restructuring* of nominal

classification systems may take different forms. In some cases, influence from the dominant language results in the use of already existing classifiers in new morphosyntactic loci, or changes in the composition of noun classes. In other cases, a system of classifiers may undergo rapid grammaticalization under areal pressure, and develop into a noun class system (Aikhenvald 2000a: 385). Both creation and restructuring of nominal classification may be accompanied by the borrowing of actual classifier morphemes or noun class affixes, which is an instance of direct diffusion. A situation in which the dominant language has no nominal classification, or only a simple system, may cause the *reduction* of nominal classification in the language under influence. In some cases, reduction is caused by extensive lexical borrowing. A large influx of lexemes that are not classified or marked for noun class may result in the overall loss of these functions in the language (Aikhenvald 2000a: 388). This, once more, shows the strong influence of language contact and bilingualism on nominal classification.

In this chapter, it was shown that nominal classification systems come in many shapes and sizes. They may differ with regard to their function, morphosyntactic locus, semantic characteristics, origins, type of development, dynamism, and stage of grammaticalization. The typological discussion that was presented here will serve as the theoretical framework for Chapter 3, in which the characteristics of nominal classification systems in Amazonian languages will be discussed.

3. Systems of nominal classification in Amazonian languages

The Amazon basin, stretched over nine countries, comprises a vast area in South America. The Amazon river, which originates in the Andes, is fed by tributaries that come from Colombia, Venezuela and the Guyanas in the north, Ecuador and Peru in the west, and Bolivia and Brazil in the south. This entire area is known as the Amazon basin. Apart from its staggering size, the region is known for being the most linguistically diverse area in the world (Grinevald & Seifart 2004). It is home to over 300 languages which belong to 20 different families[15], plus a high number of genetic isolates (Dixon & Aikhenvald 1999). Moreover, many of the linguistic families are relatively small, consisting of only a few members. Some efforts have been made at establishing genetic relationships between several of the Amazonian language families through linguistic comparison, such as Matteson (1972) and Greenberg (1987), but these seem rather speculative, for they are based on little data. In fact, the Amazon basin is said to be one of the least known linguistic regions in the world. Many of the languages that are spoken in the area remain to date undescribed, and many of the grammars that have been published so far are incomplete. At the same time, the vast majority of Amazonian languages display a certain degree of endangerment. This increases the importance and urgency of descriptive research on the languages of the region.

Dixon & Aikhenvald (1999) mention several factors that may account for the high degree of linguistic diversity in the Amazon basin. First of all, some of the tribes in the area have a system of exogamous intermarriage, in which marrying someone from the same language group is considered incestuous. This has caused extensive language contact between different ethno-linguistic groups over a long period of time. Furthermore, shortly after European colonizers entered the area, epidemic diseases quickly spread, to which the indigenous population had no immunity. As a consequence, many tribes were decimated, while others were completely wiped out. In many cases, the survivors of several tribes joined forces and started living together, which brought their languages in close contact with one another. Another cause of the linguistic diversity in the area was the frequent displacement of tribes, which occurred as an effect of the rubber boom around 1900 and during missionary activities in the 20th century. All these factors have created a 'linguistic area': a region in which genetically different language groups have come to share linguistic features through areal diffusion (Dixon & Aikhenvald 1999: 8). Some of the common traits of the Amazonian linguistic are contrastive vowel nasalization, an agglutinating structure, noun incorporation of only obligatorily possessed nouns, and split-ergative alignment systems. Another important areal feature is the presence of extensive systems of nominal classification, often with characteristics that deviate from previously established typological 'universals' (Dixon & Aikhenvald 1999). All this makes nominal classification in Amazonian languages a highly interesting field of studies, in which much remains to be discovered.

This chapter will set the stage for my research on nominal classification in Amarakaeri. Although the Harakmbut language family is considered to be either an isolate (Wise 1999) or to be genetically

[15] In their overview of Amazonian languages, Dixon & Aikhenvald (1999) distinguish the following majour linguistic families: Carib, Arawak, Tupí, Macro-Jê, Tucano, Pano, Makú, Nambiquara and Arawá. Harakmbut is listed as a one of the smaller language families in Peru, and as being a genetic isolate.

related only to Katukina[16] (Adelaar 2000, 2007), the fact that complex systems of nominal classification are considered an areal feature of Amazonia makes a discussion of nominal classification in other languages of this region highly relevant to the present research. In section 3.1, some of the comparative studies that have been undertaken so far will be discussed. The next section, 3.2, will provide a closer look at a few individual languages whose systems of nominal classification show interesting similarities with that of Amarakaeri. Finally, the possible origins and development of nominal classification systems in Amazonian languages, and the role that language contact and areal diffusion may have played in these developments, will be discussed in 3.3.

3.1 Comparative studies

As was mentioned in the previous chapter, recently described systems of nominal classification in Amazonian languages display many 'uncommon' features, which makes them very relevant to the typological debate. However, while detailed descriptions of individual languages have been published, few attempts have been made at mapping, comparing and categorizing the systems of nominal classification in the entire area, and at finding correlations between typological subtypes of systems and the geographic location of the languages they occur in. The two main studies that have been dedicated to this so far are Payne (1987) and Derbyshire and Payne (1990). These will be discussed in this section.

Payne (1987) focuses on the Western Amazon, an area which stretches from southern Colombia to southern Peru, and from the Ecuadorian lowlands to western Brazil. In the article, the many languages that are spoken in this area are divided into four groups, according to the type of nominal classification systems that they do or do not possess. Payne distinguishes two main types of nominal classification: 'noun classification' (a broad group, which includes noun classifiers, numeral classifiers and noun classes), and 'verbal incorporation' (which seems to be the same as classificatory noun incorporation such as is described by Mithun (1986)), and illustrates the typological features of each of these systems with examples from individual languages. Furthermore, attention is paid to the geographic location of each of these subgroups.

1. *Languages with noun classification*

   Group 1 consists of languages that have noun classification but have no verbal incorporation. The classifiers in these systems tend to show both inflectional and derivational properties, which makes them hover somewhere in between the categories of 'noun classifiers' and 'noun classes' that were established by Dixon (1986). Moreover, many of these languages display a combination of a numeral or noun classifier system with a noun class system that marks agreement on nouns according to gender or animacy. Group 1 includes Tucanoan, Zaparoan, Peba-Yaguan and some

[16] Since Katukina does not seem to have a system of nominal classification, or only a system in a preliminary stage (Dos Anjos 2011), Katukina will not be discussed in this chapter. The language will, however, be discussed in Chapter 5, for a comparison between Katukina and Amarakaeri might shed light on the development of the system of nominal classification in Amarakaeri.

Witotoan languages. These are mainly spoken north of the Amazon river, around its tributaries Napo and Putumayo, in Northern Peru and Southern Colombia.

2. *Languages that do not have noun classification or verbal incorporation*
The shared property of the languages which belong to this group is the overall lack of nominal classification systems. This group includes Omagua, Cocama, Tupí-Guaranían languages, and some Witotoan languages. Not surprisingly, these languages are geographically more dispersed; they are spoken all along the Ucayali river in Peru, and south of the Putumayo.

3. *Languages with both noun classification and verbal incorporation*
Languages of this group have noun classification like those in Group 1, but also show a system of verbal incorporation. Many of the verb-incorporated classifiers in these languages seem to be derived from body parts. Group 3 includes the Cahuapanan language Chayahuita, and Preandine Maipuran Arawakan languages, such as Machiguenga, Nomatsiguenga and Amuesha. These are spoken in two main areas: south of the Marañon river and at the southern end of the Ucayali.

4. *Languages that have primarily verbal incorporation*
These are the languages in which a system of verb-incorporated classifiers is dominant. Payne assigns Amarakaeri to this category, and also stresses the derivational function of classifiers in this language. Group 4 also includes some Panoan languages, such as Capanahua. The languages in this group are spread over a large area in Peru, stretching from the Madre de Dios region to the department of Loreto.

As we can see, the typological groups that are established by Payne (1987) are not confined to continuous geographic areas. However, there seems to be some correlation between the groups and their location: Groups 1 and 2 are found mainly north of the Amazon river, while 3 and 4 exist to the south of it. According to Payne (1987: 21), the proximity between languages whose systems of nominal classification share many of the same features suggests a linguistic area, and indicates that there has been significant language contact in the region, long before the colonial period.

The second comparative study on nominal classification in Amazonia is Derbyshire & Payne (1990), which is based on languages from 13 different families or stocks which possess systems of nominal classification. Unlike Payne (1987), Derbyshire & Payne (1990) distinguish three main types of nominal classification systems: numeral, concordial[17] and verb-incorporated. This leads to a more fine-grained typological categorization, into eight groups of languages, according to the combination of systems that they display:

1. *Numeral*
This group consists of languages that show only a numeral classifier system, such as the Tupí language Gavião, and some Yanomaman languages.

[17] Derbyshire & Payne (1990) use the terms 'concordial' and 'concordance' where Aikhenvald (2000a) uses 'agreement'.

2. Concordial

Within the group of languages that display only a concordial system, two subgroups are distinguished: those that have a gender-based system, and those that show both a gender and a nongender system. The former includes some Maipuran Arawakan languages, such as Piro and Apurinã; the latter consists of the Arawan languages Deni, Jaumadi, Madija, Culina and Paumari.

3. Verb-incorporated

Languages in which only a verb-incorporated system is found seem relatively rare in the area; only Pirahã and the Maipuran Arawakan language Terena are mentioned as members of this group.

4. Numeral and concordial

Some of the languages in this group (e.g. Tucano) have a small gender system and a large set of classifiers that may occur in different morphosyntactic loci, such as numerals, nouns, adjectives and demonstratives. These classifiers may also serve a derivational function. In some other languages (e.g. various Witotoan languages), a single set of classifiers has both numeral and nongender concordial functions. The group which combines numeral and concordial systems seems to roughly correspond with that which Payne (1987) denotes as Group 1, consisting of Peba-Yaguan, Tucanoan, Zaparoan and Witotoan languages.

5. Numeral and verb-incorporated

Two languages are mentioned as having a combination of numeral and verb-incorporated systems: the Cahuapanan language Chayahuita, and the unclassified language Waorani. Both systems show a single set of classifiers, mainly derived from body parts, which appear on numerals and may be incorporated into the verb stem.

6. Concordial and verb-incorporated

This group consists of languages which show a combination of concordial and verb-incorporated systems. Derbyshire & Payne (1990: 260), assign Amarakaeri to this group, stating that it has a primarily verb-incorporated system whose classifiers have also developed nongender concordial and derivational functions. The ground for the assumption that the Amarakaeri system finds its origins in verbal incorporation, is the large number of classifiers that seem to have been derived from body parts and have undergone semantic shift into shape classifiers. Hereby, reference is made to the theory that was formulated by Mithun (1986), according to which body parts are some of the most frequently incorporated nouns, and therefore often transform into verb-incorporated classifiers. Derbyshire & Payne (1990: 260) also mention that there might be agreement within the noun phrase in Amarakaeri, but that data to confirm this is lacking. A similar type of system is reported to exist in the Maipuran Arawakan language Parecis.

7. *Numeral, concordial and verb-incorporated*

A combination of all three types of systems is reported to be found in some Maipuran Arawakan languages (Amuesha, Palikur, and the Campa languages), and the Tupí language Mundurukú. Like Amarakaeri, Mundurukú is considered to have a primarily verb-incorporated system, derived from body part nouns, which has developed into a broader nongender agreement system. The Campa languages and Palikur also show a combination of numeral, concordial and verb-incorporated systems, but have an additional gender system. It is not clear whether the systems in these languages also have developed out of a verb-incorporated system (Derbyshire & Payne 1990: 266).

8. *No classifier system*

Four language families are mentioned as lacking systems of nominal classification: Cariban, Gean, Panoan and Tupí-Guaraní. However, some languages of these families are considered to show incipient systems for having verb-incorporation of body part nouns. Such is the case for the Cariban language Hixkaryana and the Panoan language Capanahua.

In addition to a subgrouping of languages according to the types of systems they display, Derbyshire & Payne discuss some of the common characteristics of Amazonian nominal classification systems. First of all, it is stated that one of the main properties of the Amazonian systems is that they cannot be labelled as one of the types that are distinguished by Allan (1977), Dixon (1986) and Mithun (1986), but tend to show a mixture of these types (Derbyshire & Payne 1990: 243). Furthermore, emphasis is placed on the important discourse-anaphoric function that class markers tend to have in Amazonian languages. This trait seems to be related to fact that in many of these languages full nouns are only scarcely used in discourse (Derbyshire & Payne 1990: 243).

Apart from the comparative studies that were presented by Payne (1987) and Derbyshire & Payne (1990), some of the common characteristics of nominal classification systems in Amazonian languages are discussed by Grinevald & Seifart (2004). In an article on the differences and similarities between the highly grammaticalized noun class systems that exist in Niger-Congo languages and the more lexico-grammatical Amazonian systems, the distinctive features of the latter are mentioned. According to Grinevald & Seifart (2004: 279), systems of nominal classification in Amazonian languages are typically large, open systems, with class markers which combine derivational and agreement functions and are often used as anaphora. Many of these systems are semantically transparent, often shape-based, and have class markers with clear nominal origins. Also, sets of class markers are frequently complemented by repeaters. The occurrence of several different types of systems in one and the same language is common, either functioning with the same set of class markers, or with different sets in different morphosyntactic contexts. As has been mentioned before, it is the combination of these features which makes the Amazonian systems difficult to place into typological categories. In summary, as Grinevald & Seifart (2004: 281) put it, "[o]ne is confronted with systems that have the semantics and discourse use of a numeral classifier system, the agreement pattern of a noun class system, and the derivational and compounding productivity of class terms."

3.2 Nominal classification in Amazonian languages: a closer look

In this section, a closer look will be provided at those Amazonian languages whose systems of nominal classification show the most striking or most numerous similarities with the system of Amarakaeri. A discussion of the structure, semantics, functions and complexity of the nominal classification systems in each of these languages will serve as a basis for comparison of my Amarakaeri fieldwork data later on in this thesis. The languages that will be mentioned in this chapter are Miraña (Witotoan), Mundurukú (Tupí), Kwaza (unclassified), and Tariana (Arawakan).

3.2.1 Miraña

Miraña is a Witotoan language that is spoken in Southern Colombia, near the Brazilian border. It is considered a dialectal variant of Bora (Grinevald & Seifart 2004). Miraña has, in many respects, a typical Amazonian system of nominal classification. It has a relatively large set of class markers (around 70), which are distributed over different morphosyntactic contexts, and show different functions and uses. What makes Miraña particularly interesting, is that its class markers combine a strong derivational function with participation in a highly developed agreement system. The class markers function as agreement markers on numerals, adjectives and verbs, as is shown in (12):

(12) *ó -di íhka* ***-:ba*** *tsa* ***-:ba***
1SG -POS COP-**CL:container** one **-CL:container**

múhuː ***-:ba*** *ʔúβiː* ***-:ba***
big **- CL:container** basket **-CL:container**[18]

'I have one big basket.'

(Grinevald & Seifart 2004: 265)

In discourse, the head noun is usually left out after it has been mentioned once, and the class markers serve as anaphora. As is argued by Grinevald & Seifart (2004: 270), this type of agreement pattern is reminiscent of the noun class systems that exist in Niger-Congo languages. However, class markers in Miraña also play an important role in word formation, and often appear in N + CL constructions which derive nouns:

(13) *tubó* ***-íːʔo***
shoot **-CL:little.stick**
'arrow'

(Seifart 2005: 203)

Interestingly, in these constructions the class marker often assumes the role of semantic head, and the preceding noun serves as its modifier (Seifart 2005). This is the case in example (13), in which the word for 'arrow' is formed by a class marker that denotes the type of entity ('stick'), and the preceding noun further narrows down the semantic scope of the class marker ('stick for shooting').

[18] For consistency's sake, I have chosen to apply my standardized glossing system to the examples that are cited in this chapter. A list of glosses can be found in Appendix 1.

In some cases, the noun in this type of compound is a bound root that cannot occur independently. Seifart (2005) calls this type of nouns 'obligatorily classified nouns'. The class marker can still be recognized in these constructions, for it still participates in agreement. This is illustrated in (14).

(14) a. *tsa* ***-ɯ*** *íːʔú* ***-ɯ***
one **-CL:3D.round** egg **-CL:3D.round**
'one egg'

b. * *íːʔú*
Intended: egg(s)

(Seifart 2005: 116)

Another use of class marker is what Seifart (2005) coins the 'absolute' use, in which a class marker is attached to the bound root *tɛː-* in order to form a full noun. In this type of construction, the element *tɛː-* functions as a semantically empty root, enabling the formation of an independent noun which conveys the semantic content of the class marker:

(15) *tɛ́* ***-tohko***
ER **-CL:creek**
'the/ a creek'

(Seifart 2005: 125)

As we will see in examples from other languages, this phenomenon, which I will call 'empty root construction' here, is attested in many of the languages in the region, and shows similarities with the use of the nominalizing prefix *wa-* in Amarakaeri.

Grinevald & Seifart (2004) also argue that class markers in Miraña may have an individuating function. That is, when nouns occur without a class marker, they denote a non-individuated concept, generic term, or biological species, rather than referring to a specific referent. In this context, the use of a class marker makes the noun refer to a specific entity. This is illustrated in (16).

(16) a. *ʔúβi*
basket
'basket(s)'

b. *ʔúβi* ***-ːba***
basket **-CL:container**
'a basket'

(Grinevald & Seifart 2004: 267)

The Miraña class marker inventory and its semantics are interesting from an etymological point of view. Seifart (2005) distinguishes three main groups: *general class markers*, *specific class markers*, and *repeater nouns*. The general class markers constitute a small and strongly grammaticalized gender system which marks animacy, sex and number. The group of specific class markers is less grammaticalized, and mainly classifies the referents of nouns in terms of their shape. Within this

group, the class markers that are monosyllabic tend to have broader semantic values than those that are polysyllabic. This corresponds with the theory which was discussed in Chapter 2, that grammaticalization - in this case, the development of lexemes into class markers - usually goes hand in hand with semantic broadening and phonological reduction. The third group which Seifart mentions, that of repeater nouns, consists of nouns that are used in class marker slots but may also occur independently. These forms can be considered the least grammaticalized within the class marker set, and show the 'openness' of the system with respect to nouns (Seifart 2005: 94). As we see, different stages of grammaticalization may are represented in a single class marker inventory.

3.2.2 Mundurukú

The Tupí language Mundurukú is spoken in the Brazilian states of Pará and Amazonas. The language has a rich system of nominal classification in which class markers are used on nouns, numerals, demonstratives, and incorporated into verbs. The class markers appear in slightly different phonetic forms in these different morphosyntactic contexts (Crofts 1973), but can still be considered to constitute a single set. This makes Mundurukú a typical example of a multiple classifier language (Aikhenvald 2000a). In their verb-incorporated use, the class markers refer to the subject of an intransitive verb, or to the object of a transitive verb. When used on demonstratives and numerals, they show agreement with the noun; in discourse, the full noun is often left out and class markers serve as anaphora (Crofts 1971). Just like in Miraña, the class markers often have a derivative function when being suffixed to nouns. As is illustrated in (17), multiple class markers may be suffixed to the same noun root[19]:

(17) ***yɨk³*** ***-pi³*** ***-da²*** ***-sẽñ²*** ***-pɨ³***
CL:belly -CL:interior -CL:seed -CL:worm -CL:finger
'intestinal worms'

(Crofts 1971: 5)

As we see in (17), class markers may even function as the first part of such compounds, which further blurs the distinction between nouns and class markers, and raises questions about the morphosyntactic status of these elements in Mundurukú. Indeed, Crofts (1971: 4) considers them to be a subtype of noun roots. However, the use of class markers as roots is not very common; they are more frequently found as the second part of compounds, as anaphora, or as agreement markers (Crofts 1973). A very productive use of class markers is attested in N + CL combinations which form part of a set, such as combinations with the noun *ako* 'banana':

(18) *a²ko³* ***-ba⁴***
banana **-CL:arm-like**
'banana fruit'

a²ko³ ***-dip²***
banana **-CL:field**
'banana field'

[19] In the examples from Mundurukú, superscript numbers indicate tone: 1 = high tone, 2 = mid tone, 3 = low tone, 4 = laryngealization (Crofts 1971).

a²ko³ ***-'ip²***
banana **-CL:tree**
'banana tree'

a²ko³ ***-dot²***
banana **-CL:stalk**
'stalk of bananas'

(Crofts 1971: 5)

According to Crofts (1971), this type of sets is limited to plants and trees.

Even though the Mundurukú class markers may occur independently as full nouns, there is also a construction which resembles the empty root construction that exists in several languages of the area. This construction consists of the noun *t̨ɨ²mɨn³* 'thing' plus a reduplicated class marker, as is illustrated in (19) below:

(19) *t̨ɨ²mɨn³* ***-t̨ɨp²*** ***-t̨ɨp²***
thing **-CL:leaf** **-CL:leaf**
'leaf-like object'

(Crofts 1971: 2)

The nouns which are constructed this way have a very generic meaning; they may refer to any item that belongs to this class in terms of its shape.

Most of the class markers in Mundurukú seem to have been derived from body part nouns. This, in combination with their important function as verb-incorporated markers, has led Derbyshire & Payne (1990: 261) to argue that nominal classification in Mundurukú was originally verb-incorporated, and the system has only later spread into other morphosyntactic contexts. Indeed, the Mundurukú class markers show a certain extent of semantic generalization from body part term into general shape marker. For instance, the class marker *-'a²* has the basic meaning 'head', but is also used on nouns that refer to other round things, such as 'mango' and 'potato' (Crofts 1971: 6).

It should also be mentioned that in Mundurukú the origins of the class marker inventory seem to be related to possession type. The language shows a clear distinction between alienably and inalienably possessed nouns; the latter are obligatorily prefixed with a possessive marker. According to Crofts (1971: 4), class markers in Mundurukú can be defined as 'inalienably possessed noun roots which function to classify objects which have similar physical characteristics.' The fact that these class markers are all associated with inalienability is not surprising, for they are mostly based on body part nouns, which are typically inalienably possessed. As we will see, this relation between nominal classification and alienability seems to exist in other Amazonian languages as well.

3.2.3 Kwaza

The unclassified language Kwaza is spoken in the southeast of the Brazilian state of Rondônia, which forms part of the highly linguistically diversified Guaporé-Mamoré region. With only 25 speakers left (Van der Voort 2004), the language can be considered critically endangered. An extensive reference grammar of Kwaza was published by Van der Voort in 2004.

Kwaza has an elaborate system of nominal classification: it displays a single set of about 150 class markers that occur in several different morphosyntactic contexts. When attached to demonstratives, interrogatives and numerals, class markers show agreement with the noun. The noun itself does not have to be overtly expressed, and class markers often have an anaphoric function (Van der Voort 2004: 128). Apart from these contexts, the class markers can also be incorporated into the verb, where they classify one of its arguments:

(20) *awãta* ***-'xy*** *-da* *-ki*
see **-CL:leaf** -1S -DEC
'I'm looking at the paper'

(Van der Voort 2004: 134)

A particular feature of the Kwaza system is that it also employs a semantically neutral class marker, which may act as a substitute for any of the other class markers. When attached to verb stems, this neutral class marker functions as a nominalizer. Apart from marking agreement, class markers in Kwaza often have a derivational function. In combination with nouns, the classifier tends to act as the second member of a compound. An example is given in (21) below:

(21) *jo* ***-'mũ***
manioc**-CL:liquid**
'manioc beer'

(Van der Voort 2004: 134)

In some cases, N + CL constructions have merged; the noun can no longer occur independently, and the combination as a whole has become semantically opaque. This is illustrated in (22):

(22) *kaxa.**'ri***
? .**CL:flat**
'stone file'

(Van der Voort 2004: 135)

Just like in Miraña and Mundurukú, Kwaza has an empty root construction, which derives nouns from class markers. When a class marker is attached to the semantically empty root *e-*, a noun is formed which has the same semantic value as the class marker itself:

(23) *e* ***-'sĩ***
ER **-CL:seed**
'seed'

(Van der Voort 2004: 136)

In Kwaza, the empty root construction often occurs with inalienably possessed items (e.g. body parts). According to Van der Voort, this suggests that the empty root finds its origin in a marker of inalienability. Synchronically, however, the empty root does not seem to be a clear indicator of the distinction between alienable and inalienable nouns, for it may also occur with alienable items, and does not appear on all of the nouns that refer to inalienable items (Van der Voort 2004: 136).

As for the etymology of the class markers, Van der Voort distinguishes several subgroups within the class marker set. Many of the class markers show close phonetic and semantic correspondence to certain nouns, and have a limited distribution and specific reference. These class markers, as Van der Voort argues, seem to be the verb-incorporated or compounded variants of full nouns (Van der Voort 2004: 147).[20] Some other class markers are marked as being etymologically opaque; these are only partially related to certain nouns, and tend to have slightly less specific semantic values. The third group of class markers consists of those that can appear with the empty root *e-* to form nouns. Interestingly, these class markers do not show any relation to nouns; they appear to derive nouns themselves, rather than being derived from nouns. The three subtypes of class markers do not seem to be related to any semantic domain in particular: all three groups contain class markers from several semantic domains, such as shape, substance, function and direction (Van der Voort 2004: 177).

With regard to the relative age of nominal classification in Kwaza, Van der Voort argues it must be very old. Indicators for this are the fact that the class markers occur in many different morphosyntactic contexts, and the extent of grammaticalization of many of the classifiers. However, the system remains quite productive, as it also classifies new nouns, such as loanwords from Portuguese (Van der Voort 2004: 175).

When placing the Kwaza system in a broader typological perspective, one can say that it is one of the more complex and elaborate systems of nominal classification, even among those that exist in Amazonia. Van der Voort mentions the unclassified languages Kanoê and Aikanã as having the most similar nominal classification systems as that of Kwaza, both formally and structurally. This suggests that there has been extensive contact between these languages and Kwaza. In a slightly broader sense, the combination of morphosyntactic contexts in which Kwaza class markers occur, makes the language fall into Derbyshire & Payne's category of languages which have systems that combine numeral, concordial and verb-incorporated classification (Van der Voort 2004: 128). Other languages in this category are Palikur, some Campa languages, and Mundurukú.

3.2.4 Tariana

Tariana is an Arawakan language that is spoken in the north of Brazil. It is part of the Vaupés linguistic area, which is located in the north-west of the Amazon basin and includes languages from different families, such as Tucanoan, Arawakan and Makú. Tariana has an extraordinarily complex system of nominal classification, which is described by Aikhenvald (2000b). Just like Mundurukú, Tariana is considered to have a multiple classifier system. Its class markers are used in a wide range of morphosyntactic loci: they occur on nouns, adjectives, demonstratives, articles, numerals and

[20] In this sense, this group displays similarities with repeater phenomena as were described in section 2.3.2.

possessive constructions, and are incorporated into verbs. All nouns in Tariana carry a class marker. There is a small gender system which distinguishes between animate and inanimate; the group of animate nouns is further divided into feminine and non-feminine, while inanimate nouns are classified according to their shape. The system of shape classifiers is relatively open and makes frequent use of repeaters (Aikhenvald 2000b: 95). The type of nominal classification system that is found in Tariana, which combines gender and shape-based class markers, is similar to the systems of Tucanoan languages, which are also spoken in the Vaupés area.

What makes Tariana particularly interesting with regard to Amarakaeri is its class marker inventory. A list of Tariana class markers, which is provided by Aikhenvald (2000b), shows interesting correspondences with the class marker inventory of Amarakaeri. This is illustrated below:

TABLE 3
SAMPLE LIST OF SHAPE-BASED CLASS MARKERS IN TARIANA
(after Aikhenvald 2000b: 95)

Class marker	**Semantics**
-da	**round**
-ipa	big open space
-ku	extended cloth
-kwema	flat and round
-kha	curved
-maka	clothing
-na	long and vertical
-pa	**largish and long**
-pi	**long, thin and vertical**
-pu	**long and hollow**
-pukwi	round and hollow

As we will see in Chapter 4, the class markers in bold font are also found in Amarakaeri, with almost the exact same forms and semantic values. It seems like these correspondences are the result of direct diffusion, either through direct contact between the two languages, or through other languages in the area. This is rather surprising, considering the geographical distance between Amarakaeri and Tariana.

3.2.5 Discussion

The languages that were discussed above all have complex systems of nominal classification, which include relatively large class marker inventories, a distribution of class markers over a wide range of morphosyntactic loci, and a combination of different functions and uses. Class markers display both agreement and derivational functions in all these languages, and have additional functions, such as individuation and nominalization, in some of them (e.g. Miraña). As was mentioned before, the pervasiveness and productivity of the derivational function of class markers is typical of Amazonian languages (Derbyshire & Payne 1990; Grinevald & Seifart 2004), which seems to blur somewhat the distinction between nouns and class markers. In languages like Kwaza, Miraña and Mundurukú, class markers often occur in N + CL constructions which closely resemble N + N compounds and serve to derive nouns. Interestingly, the class marker seems to function as semantic head in these constructions (Van der Voort 2004; Seifart 2005). Moreover, several class markers may be attached to one noun, each adding its own semantic content to the word. As we saw in the case of Miraña and Kwaza, N + CL constructions may also lexicalize and merge, becoming semantically opaque and morphologically inseparable. It is interesting to see that in these cases, the original class marker may retain its agreement properties, while the original host noun can no longer occur independently. Furthermore, Tariana and Miraña make frequent use of repeaters, which may function both as nouns and as class markers; in Mundurukú, class markers are also reported to occur as independent nouns, functioning as the first part of compounds (Crofts 1971). All these properties indicate that in the languages which were discussed in this chapter the distinction between nouns and class markers is far from clear-cut. Along these lines, Weber (2002: 1) argues that the class markers of Bora (which is closely related to Miraña) should be considered bound nouns, for having the same distribution and referential properties as nouns. However, within the framework of the grammaticalization continuum, one could say that the categorization of class markers as either nouns, classifiers or noun classes is more or less a matter of definition, since these elements are somewhere between the lexical and grammatical end of the scale, showing properties of both sides.

In a comparative study of Arawakan languages, Matteson (1972) mentions the widespread use of so-called 'from words' throughout the Arawakan language family[21]. According to Matteson, these elements tend to denote shape and play an important role in word formation, often occurring in compounds, or as independent nouns when combined with some kind of empty root affix. Presently, in the context of the descriptions that have become available of languages such as Kwaza and Mundurukú, it seems that these 'form words' should now be analyzed as 'typical' Amazonian class markers, which display important derivational functions and have noun-like distributions. The shape-basedness of these elements that is signalled by Matteson is an interesting property of many Amazonian class marker inventories, which deserves more attention here. As we saw in the languages that were discussed in this chapter, shape tends to be the dominant semantic domain in the class marker inventories of these languages. This shape-basedness, as is argued by Mithun (1986), is in many cases related to the way in which the nominal classification system originated: many - although not all - of these systems seem to have developed through classificatory noun incorporation involving body part nouns. In the ~~process of grammaticalization,~~ body parts are semantically generalized through extension into the domain of shape (e.g. head → round object). As was mentioned before, Derbyshire & Payne (1990) argue that class markers have taken this

[21] Matteson (1972) classifies the Harakmbut languages as Arawakan, and therefore includes these in her comparison of Arawakan languages.

grammaticalization path in Amarakaeri, Mundurukú, Chayahuita, Waorani and Parecis. Similarly, Kwaza has a large amount of class markers that seem to have body part origins, and which can be incorporated into verbs. Bacelar (2004: 120) argues that Kanoê also belongs to this typological group, for it has a primarily verb-incorporated system of nominal classification similar to that of Mundurukú and Kwaza. In the Arawakan language Palikur, Aikhenvald & Green (1998), distinguish a set of class markers that occur on verbs, and a closed set of body part nouns that may be incorporated into verbs. These two types are distinct with regard to their morphosyntactic behaviour and semantic effects. However, grammaticalization of incorporated body part nouns into verbal class markers does seem to occur, since a body part origin can be established for several of the verbal class markers (Aikhenvald & Green 1998: 455).

Interestingly, as was mentioned before, there seems to be a relation between class markers that have evolved from incorporated body part nouns on the one hand, and the distinction between alienable and inalienable possession on the other. According to Dixon & Aikhenvald (1999), verbal incorporation of only inalienably possessed nouns is an areal feature among Amazonian languages. This corresponds with Mithun's (1986) theory, that body parts - which always belong to someone and are therefore inalienably possessed by definition - are among the nouns that are most frequently incorporated into verbs. Consequently, if mainly inalienably possessed items are incorporated into verbs, these will also serve as the basis for most of the class markers in some languages. Indeed, this seems to be the case in many of the languages that have primarily verb-incorporated systems of nominal classification; their class marker inventories tend to be largely shape-based, often with discernable body part origins. As was mentioned above, Mundurukú class markers are defined as 'inalienably possessed noun roots' (Crofts 1971); the same goes for class markers in Piro (Matteson 1972). The existence of an empty root construction in some languages may be related to this inalienable nature of class markers. Inalienably possessed nouns are bound in the sense that they are usually attached to a possessive marker, or to another noun with which they form a genitive-like construction. In order for these nouns to occur as independent, unpossessed nouns, then, the use of a certain grammatical marker is often required. Matteson (1972) describes this phenomenon for Arawakan languages, stating that 'form words' tend to be inalienably possessed nouns, which either occur with a possessor, or with a 'special suffix': **-tsi* in Proto Piro-Apuriná and Proto-Ashaninka; **-si* in Proto Newiki, and *-e* in Baure. Prefix *wa-* from Harakmbut languages is mentioned as a similar phenomenon (Matteson 1972: 164). As we saw in this chapter, the empty root construction is also attested in Miraña, Kwaza and Mundurukú. Furthermore, Barnes (1990: 283) signals a similar construction, involving a 'dummy noun', in the Tucanoan language Tuyuca, Aikhenvald (2002: 94) observes equivalent phenomena in the Tucanoan languages Desano and Barasano, and Crevels & Van der Voort (2008: 168) mention an empty root construction in Kwaza, Kanoê (both unclassified), Latundê (Nambikwara), and in the Tacanan languages Ese Eja and Cavineña. Although the relation between the empty root construction and inalienable possession is not clearly visible in all of these languages, a comparison by Van der Voort (2004) suggests that this relation may be an areal feature of the Guaporé-Mamoré linguistic area, for it seems to exist in Kwaza, Kanoê, Gavião and Mekens, all of which are spoken in the same region. Gavião, for instance, has a numeral classifier system in which some of the class markers have a corresponding noun stem. These noun stems are always inalienably possessed and must carry a possessive prefix. In order to form a non-possessed variant of the noun, a construction which involves a semantically empty dummy noun, meaning 'thing, possession' must be used (Derbyshire & Payne 248). In Kanoê, the

empty root *i-* is used to derive independent nouns from elements that behave as class markers in other contexts. An example is the independent noun *i-kuta* 'head', which contains the class marker *-kuta* 'head shaped, round' (Bacelar 2004: 119). The vast majority of body part and plant part nouns in Kanoê are formed in this way, using prefix *i-* as a dummy noun (Bacelar 2004: 102); this shows a clear connection between inalienability and the use of the empty root construction. In Kwaza, this correlation seems to be less systematic: although many inalienably possessed noun roots or class markers are prefixed by empty root *e-*, this prefix is also used for some items that cannot logically be considered inalienably possessed. However, Van der Voort (2004: 136) argues that inalienability may be the origin of the empty root construction, and that the distinction may have been lost over time. As we will see later on, a similar process might have taken place in Amarakaeri.

All in all, it seems like the systems in many of the languages that were discussed in this chapter show similar combinations of features. This may be due to the fact that some of the properties that are typical of nominal classification systems in these languages are related to one another. Departing from this idea, two main conglomerates of features can be distinguished in the systems that were discussed here. The first is clustered around the derivational function of class markers. The important role that class markers play in word formation in many languages seems to be related to their noun-like distribution and relatively high degree of semantic transparency. Also, lexicalization of N + CL constructions and the possibility of attaching several class markers to one noun seem to be 'side effects' of this derivational function. This cluster of features is found in languages throughout the Amazonian region. The second cluster is related to classificatory noun incorporation. As we have seen, verbal incorporation of body parts is often associated with shape-basedness of the class marker inventory, and with the existence of an empty root construction that is related to inalienability. The combination of these features seems to be slightly more area-specific, existing mainly in the southern part of the Amazon region and being concentrated around the Guaporé-Mamoré area; it is not as clearly present in more northern language families such as Tucanoan and Witotoan. This roughly corresponds with the geographic distribution of typological groups that was mentioned by Payne (1987), according to which the languages that show classificatory noun incorporation (Groups 3 and 4) mainly exist south of the Amazon river. However, although these clusters may indeed represent areal traits, it must be noted that this distinction is rather tentative, and far from clear-cut, since all languages and systems of nominal classification are different, and many languages also display features from both clusters. More detailed and extensive research would be needed in order to obtain a clearer image of the areal properties of nominal classification systems in Amazonia.

3.3 Origins and development of nominal classification in Amazonian languages

The two main clusters of features that were discussed above seem to be related to the way in which the systems of nominal classification have developed. The systems that present features such as body part incorporation, the empty root construction, and its relation to inalienability, are typically the ones which seem to find their origins in classificatory noun incorporation. Instances of these languages are Mundurukú, Chayahuita, Waorani, Parecis and Kanoê, and, as will be discussed in Chapter 7, probably Amarakaeri. The main argument for considering these languages as having

primarily verb-incorporated systems is that their class markers mainly occur in verb-incorporated positions, and that the class marker inventory is largely composed of class markers that seem to have been derived from body part nouns. According to Derbyshire & Payne (1990: 263), some languages appear to be in a preliminary stage of the process and can be considered to have incipient systems of nominal classification. Examples are the Cariban language Hixkaryana, and the Panoan language Capanahua, which show incorporation of body part nouns into verbs. In the course of time, these systems may also develop other classificatory functions, such as agreement on numerals and demonstratives (Derbyshire & Payne 1990).

However, not all systems of nominal classification seem to originate in classificatory noun incorporation. Many systems in the area are reported to have developed through other grammaticalization paths, such as compounding. In Miraña, for instance, class markers are reported to have entered the system as repeaters, which have undergone semantic broadening and phonological reduction during the grammaticalization process. The syntactic construction in which this grammaticalization occurred was probably that of N + N compounds with genitive-like semantics (Grinevald & Seifart 2004). According to Seifart (2005), the presence of repeaters in Miraña shows a certain degree of openness of the system towards nouns. Other examples of languages that have repeaters are Tuyuca, Yagua and Tariana.

It must be noted, however, that many languages in the area have large and internally diverse class marker inventories, with class markers that show different degrees of grammaticalization and semantic generalization. This indicates that, in many languages, some class markers have undergone grammaticalization earlier, or more quickly, than others. For instance, although Miraña has a core set of highly grammaticalized monosyllabic class markers with a broad semantic value, it still makes frequent use of repeaters, which are semantically specific and less grammaticalized (Seifart 2005: 316). This shows that within one and the same set of class markers, different stages of the grammaticalization process may be represented. Another example is Palikur, which has several distinct class marker sets that occur in many morphosyntactic contexts, such as verbs, nouns, numerals and locatives. While some of these class markers are reported to have developed through a repeater stage, others seem to have originated in classificatory noun incorporation (Aikhenvald & Green 1998: 466). According to Derbyshire & Payne (1990), this points to an independent development of the different classifier sets in Palikur. In any case, it indicates that class markers may not only grammaticalize at different paces and moments in time, but may also take distinct grammaticalization paths.

Something that further strengthens the claim that class marker inventories tend not to have uniform origins, is the fact that some class marker forms are found in many different Amazonian languages. This indicates that some languages have borrowed class markers from other languages. For instance, as we saw in this chapter, some of the class markers that exist in Tariana show striking resemblances with the Amarakaeri shape morphemes. One of these, the form *-pi* 'long, thin, vertical' seems to be especially pervasive throughout the Amazon basin. It is also mentioned by Matteson (1972), who argues it is used to denote rod-shaped and usually flexible objects in many 'Arawakan' languages, such as Piro, Newiki, Apuriná and Amarakaeri. According to Aikhenvald & Dixon (2001), **-pi* was a classifier for long, thin objects in Proto-Arawak, and was derived from the noun **api*, 'snake'. Possibly, the Mundurukú class marker *-pɨ* 'finger' stems from the same source.

In Chapakuran languages, a similar element is used to denote thorns or other sharp, pointed objects on the ground (Adelaar, p.c.). It thus seems that some of the shape-based class markers that occur in Amazonian languages have spread through direct diffusion, rather than having grammaticalized from full noun into class marker in each individual language. The fact that forms such as *-pi* are found in many languages of the area suggests that these are relatively old.

This spread of class marker forms is perhaps not surprising, considering the extensive language contact in the area that was mentioned at the beginning of this chapter. However, as is argued by Aikhenvald (2000a), indirect diffusion between systems of nominal classification is much more common than direct diffusion. Indeed, while the systems of genetically unrelated languages in the area share many structural properties, instances of direct borrowing of class marker forms seem to be relatively scarce. Some instances of extensive structural borrowing between systems of nominal classification are mentioned by Aikhenvald (2002). In the Vaupés area, as a result of contact with Tucano, Tariana class markers have spread into new morphosyntactic environments, such as demonstratives. In the same area, influence from Tucano has affected the semantic organization of nominal classification in Tariana. Moreover, through contact with Bora, class markers in Resígaro have come to be used in new morphosyntactic structures. The contact between these two languages has also resulted in direct diffusion: Resígaro has borrowed a large amount of classifier forms from Bora (Seifart 2007). Concluding, one can say that areal diffusion - whether direct or indirect - may play an important role in the development of nominal classification systems.

As we have seen, many of the characteristics of the systems of nominal classification that have been discussed in this chapter seem to be areal traits, and language contact is often an important factor in the development of these systems. However, in my research on the origins and development of nominal classification in Amarakaeri, I intend to consider not only areal diffusion, but also possible genetic affiliation and inheritance. Therefore, Chapter 5 will be dedicated to Katukina, which is mentioned by Adelaar (2000, 2007) as a possible relative of the Harakmbut language family.

4. Previous works on nominal classification in Amarakaeri

The main works on Amarakaeri that have been published so far are Hart (1963), Helberg (1984), and Tripp (1976, 1995). While Hart has written a short but highly informative article which focuses on the shape morpheme system, Helberg's work is a quite comprehensive reference grammar of the language. Tripp has published an Amarakaeri-Spanish and Spanish-Amarakaeri dictionary in 1995. Moreover, this author has written a number of short articles on the grammar of the language, one of which is also of interest to this research. All these works, and the insights they have provided on the shape morpheme system, will be discussed in this chapter.

4.1 Hart (1963)

Raymond E. Hart, a linguist who is affiliated to the Summer Institute of Linguistics (SIL), conducted fieldwork in one of the Amarakaeri communities on the Colorado river[22] in the period 1958-1960. In his article *"Semantic components of shape in Amarakaeri grammar"*, which is the product of his research, Hart clarifies the Amarakaeri system of shape morphemes. Helberg (1984) states in his grammar that Hart's article is of great importance to the research on this language. However, he also points to the limitations of the article, such as the unclear transcription method. In my view, the paper is very helpful and abounds with interesting examples, but lacks more extensive discussion. In spite of these shortcomings, however, it is still highly relevant to my research.

Hart defines the shape morphemes as 'morphemes which classify objects and actions according to the particular shape or combinations of shapes inherent in the item or action under focus' (Hart 1963: 1). He also states that many of these morphemes are semantically related to body part terms, and that they are quite productive in neologisms. Hart divides the shape morphemes into three sub-classes:

1. Bound forms which occur in their minimal form with prefix *wa-*
2. Bound forms which occur only in compounds
3. Bound linkers which occur in compounds but can also serve as fillers of the pre-stem slots in verb constructions

For each of these subclasses, he mentions the shape morphemes which belong to that class. These are listed in Tables 2, 3 and 4 below[23].

[22] Hart does not specify the name of the community, but given that it was located on the Colorado river, it was probably Puerto Luz or San José Karene.

[23] For consistency's sake, I have chosen to stick to the FENAMAD orthography throughout this thesis, including in the representation of other authors' data and examples.

TABLE 4

SUBCLASS 1 SHAPE MORPHEMES ACCORDING TO HART (1963)

(bound forms which may occur in their minimal form with classifier prefix *wa-*)[24]

shape morpheme	meaning
-'ay	cane, stalk like
-hai'	bone
-hë	hole
-hën	flesh (animate or inanimate)
-hok	pelt
-hön	front part of the neck
-höng	powder
-hot	covering
-hu'	teat, breast
-hun	upper back area
-'i	foot
-'idn	tooth
-'iwit	root
-kangka	chest
-khey	flower
-kidn	small round seed
-kitta'	mouth
-knda	egg
-kmbere	eyebrow
-kmö	shot like object
-kok	front part of cheek
-kpo	eye, eye shape
-ktong	protrusion
-ku	head, crown shape
-kuäkë	channel
-kwë	juice
-kwën	bud
-mba'	hand, hand shape, leaf shape
-mba'agn	shoulder

[24] Hart lists all these morphemes in combination with prefix *wa-*, but I have chosen to leave out the prefix, since these morphemes may also occur without it, for instance when suffixed to nouns.

-më	liver
-mbih	string, string like item
-mïn	intestine
-mbo'	leg
-mbogn	lip, lip shape, edge
-nda	round or oblong shaped fruit
-nëk	belly
-nö'	tongue, tip
-nörë	heart, heart shape
-nöy	forehead
-npä	wound, depressed area
-nduï'	floor, flat area
-'o	arm, branch
-'öh	nose
-'okah	wing
-pa	rod shaped object
-pak	nest
-pe	disk shaped object
-pë	ear (channel)
-pëgn	flat with no definite contour shape
-pi	stick, long narrow object
-pïrï	tendon
-po	round, ball or box like object
-podn	cluster
-pu'	tube shaped object
-sïng	stinger
-sö'	body
-sot	dough like material
-wë	liquid, river
-wih	hair

TABLE 5

SUBCLASS 2 SHAPE MORPHEMES ACCORDING TO HART (1963)

(bound forms which occur only in compounds)

shape morpheme	meaning
-ko	half moon shape
-kogn	hole, pit
-mba	area
-mero	shell
-mëh	hump, curve
-mö	butt, blunt end
-möh	hip
-nö	centre
-öngku	joint
-pih	finger, long flexible object
-si'	peel
-sodn	flappy object
-töng	truck
-wït	pith

TABLE 6

SUBCLASS 3 SHAPE MORPHEMES ACCORDING TO HART (1963)

(bound linkers which occur in compounds but may also occur as fillers of the pre-stem slots in verb constructions)

shape morpheme	meaning
-k-	degree of permanency (uncertain)
-ta'-	against, toward, base
-ti-	extension
-to-	down, upon

With regard to prefix *wa-*, Hart does not give a clear analysis of its exact function and use. He calls it 'classifier prefix' in the context of shape morphemes, but glosses it as 'nominalizer' when it occurs in deverbal nouns. This suggests that in Hart's opinion there are two different prefixes *wa-*.

The main part of Hart's article consists of a discussion of the different morphosyntactic contexts which the shape morphemes may occur in. Three types of context are mentioned: *noun constructions*, *adjective constructions*, and *verb constructions*. Within the group of noun constructions, Hart distinguishes four different subtypes. The first subtype involves combinations of shape morphemes which occur with prefix *wa-*. This type of construction is illustrated in example (24) below:

(24) *wa* ***-'idn*** ***-ta'*** ***-pe***
NMLZ **-SHM:tooth** **-SHM:base** **-SHM:disk.shaped**[25]
'chin'

(Hart 1963: 2)

In the second type of construction, one or more shape morphemes are suffixed to free roots, rendering a compound-like structure:

(25) *äwït* ***-ku***
otter **-SHM:head**
'otter's head'

(Hart 1963: 3)

The third construction type consists of shape morphemes in combination with a nominalized verb, as in example (26). In these constructions, the shape morphemes directly follow prefix *wa-*, and the verb root comes last:

(26) *wa* ***-mba'*** *-oro*
NMLZ **-SHM:hand/leaf** -hang.up
'railing'

(Hart 1963: 3)

Finally, Hart mentions the construction of nouns on the basis of shape morphemes that have undergone semantic shift and have ceased to be bound morphemes. The shape morpheme is the first element of this type of construction, without being prefixed by *wa-*:

(27) ***ku*** ***-ta'*** ***-pe***
SHM:head **-SHM:base** **-SHM:disk.shaped**
'stool'

(Hart 1963: 3)

With regard to the distribution of shape morphemes in adjective constructions, Hart mentions two different subtypes. In the first adjective construction type, the shape morphemes are preceded by the adjective root and followed by enclitic adjective marker *-nda*:

[25] In the examples in this chapter I use my own glosses, in order to avoid confusion.

(28) *wänak* ***-nö*** ***-po*** *-nda*
good **-SHM:centre** **-SHM:round/box.like** -ADJM
'good-hearted'

(Hart 1963: 3)

In the second type of adjective construction, the shape morphemes precede the adjective root; again, the adjective marker comes last. In some of the examples that Hart uses to illustrate this construction, the shape morphemes are preceded by prefix *wa-* (29), while this element is left out in some other cases (30):

(29) *wa* ***-pa*** *-nirit* *-nda*
NMLZ **-SHM:rod.shaped** -prickly -ADJM
'prickly root'

(Hart 1963: 3)

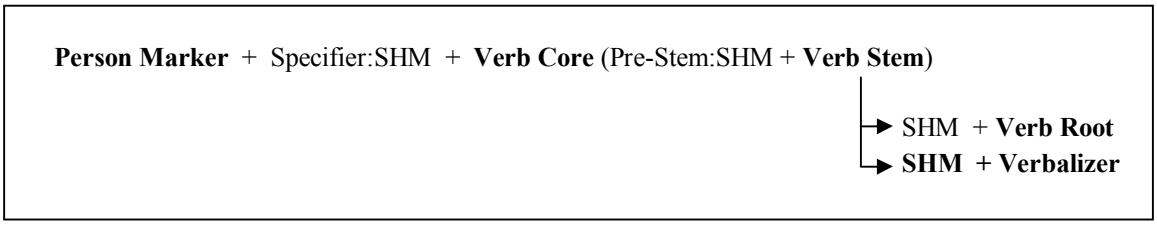
(30) ***mba'*** ***-pe*** *-sërë* *-nda*
SHM:hand/leaf **-SHM:disk.shaped** -bare -ADJM
'bare clipboard'

(Hart 1963: 3)

No explanation for this variation is provided in Hart's article.

As for the occurrence of shape morphemes in verb constructions, Hart distinguishes different subtypes, according to the slot which the shape morpheme fills in the verb. He gives the following formula in order to show which slots the shape morpheme may occur in[26]:

FIGURE 2
OCCURRENCE OF SHAPE MORPHEMES IN VERB CONSTRUCTIONS ACCORDING TO HART (1963)



In example (31), the combination [Person Marker + SHM + Verb Root] is illustrated:

(31) *o'* ***-ku*** *-tiri*
3SG **-SHM:head** -ache
'he has a headache'

(Hart 1963: 4)

In the next example, we see a construction of the type [Person Marker + SHM + Verbalizer]:

[26] Bold font represents obligatory elements, while the other elements are optional; arrows show different options.

(32) *o'* ***-wih*** *-pak*
3SG **-SHM:hair** -VBLZ
'he grows hair'

(Hart 1963: 4)

One of the more complex construction types is shown in example (33), representing the following combination: [Person Marker + Specifier:SHM + Pre-Stem:SHM + SHM + Verb Root].

(33) *yö* ***-npa*** ***-ti*** ***-nö*** *-sik* *-a'*
1SG **-SHM:wound** **-SHM:extension** **-SHM:centre** -darken -CAUS
'I am depressed because of my wound'

(Hart 1963: 5)

Apart from discussing in what morphosyntactic contexts the shape morphemes may occur, Hart also shows that there can be agreement between shape morphemes on different grammatical levels, such as we see in (34):

(34) *wa* ***-pi*** *ih* ***-pi*** *-ka'i*
NMLZ **-SHM:stick.like** 1SG **-SHM:stick.like** -make
'I make an arrow'

(Hart 1963: 5)

While a further discussion of this phenomenon is lacking in the article, some pattern can be found in the examples which Hart mentions. All of these contain a verb construction that includes a shape morpheme; the shape morpheme usually expresses the object of the verb, and is repeated in a noun construction in the same sentence, further specifying the object. This reminds of primarily verb incorporated classifier systems, such as are described by Mithun (1986).

Something that is more amply discussed in the article is the relation between the semantic value of the shape morphemes and their position in a noun construction. Hart argues that when shape morphemes are combined, the morpheme which occurs as the first member of a combination can be considered the head, while those that follow serve as modifiers. The consequence of this is that the shape morpheme which occurs in head position usually has a body part meaning, while those in modifier position tend to have a more general shape meaning. Some body parts occur more frequently as head and some are more commonly used as modifiers, but one and the same shape morpheme may show different semantic values in different positions. An example of this is *-mba'*, which usually means 'hand' in first position (35), but expresses the more general shape meaning 'hand shape / leaf shape' when occurring in modifier position (36):

(35) *wa* ***-mba'*** *-pi*[27]
NMLZ **-SHM:hand/leaf** -SHM:stick.like
'finger'

[27] It must be noted that Hart (1963) distinguishes *-pi* and *-pih* as different shape morphemes; in the original example, *-pih* is used, glossed as 'finger'.

(Hart 1963: 2)

(36) *toket* *-ku* *-ti* ***-mba'***
corn -SHM:head -SHM:extension **-SHM:hand/leaf**
'corn husk'

(Hart 1963: 2)

Finally, Hart mentions that some shape morpheme combinations have undergone considerable semantic shift and have become semantically opaque, like in the combination *wa-nda-po*:

(37) *wa* ***-nda*** ***-po***
NMLZ **-SHM:round/oblong.fruit** **-SHM:round/box.like**
'stomach'

(Hart 1963: 2)

4.2 Helberg (1984)

Heinrich A. Helberg Chávez, a German-Peruvian linguist, started his research on Amarakaeri 1976, and conducted fieldwork in the community of Shintuya. This study resulted in a doctoral dissertation at the University of Tubingen, which was published in 1984. The book is a comprehensive grammar describing the phonology, morphophonology, morphology and syntax of the language. A small section of the book is dedicated to the shape morphemes and their use.

In a chapter on morphology, Helberg distinguishes seven different morpheme classes:

- *nominal morphemes*
- *verbal morphemes*
- *adjectival morphemes*
- *adverbial morphemes*
- *shape morphemes*
- *numeral morphemes*
- *neutral morphemes*

According to Helberg, distinguishing between morphemes that belong to a closed class (affixes) and those that are part of an open class (lexemes) is not always possible, since some morphemes can function either as affixes or as lexemes. These constitute the class of neutral morphemes. Although Helberg considers the shape morphemes to be a separate class, he also defines them as 'neutral', since they can function both as lexemes and as affixes. He furthermore defines the shape morphemes as semantic classifiers, and signals the same semantic ambivalence that was mentioned by Hart (1963) before, i.e. the different semantic values of shape morphemes in different positions of the word.

In his discussion of the shape morphemes, Helberg cites Hart (1963) and seems to largely agree with this author's analysis. He gives a list of shape morphemes which is very similar to Hart's. The

only differences are that Helberg adds the shape morpheme *-mbuh* 'stem of young plant' to the list, and that he analyzes Hart's *-pi* 'stick, long narrow object' and *-pih* 'finger' as one and the same morpheme. Helberg mentions the same subclasses of shape morphemes as Hart, but also establishes a fourth subclass. This class consists of those shape morphemes that can be used as nominal affixes, apart from their occurrence in combination with prefix *wa-*. Helberg argues that these morphemes function as lexemes when combined with *wa-* (38), but that they can also function as affixes when combined with certain nouns or proper names (39):

(38) *wa* ***-pi***
NMLZ **-SHM:stick.like**
'stick'

(Helberg 1984: 247)

(39) *wa* *-iri* ***-pi***
NMLZ -person.of.respect **-SHM:stick.like**
'slim person of respect'

(Helberg 1984: 247)

According to Helberg, this subclass consists only of the shape morphemes *-po* and *-pi*.

Concerning the use of shape morphemes in verb constructions, Helberg further expands on and gives examples of the fact that shape morphemes *-to-* and *-ta'-*[28] are frequently used in the pre-stem, serving to narrow down the semantic value of the verb. Morpheme *-ta'-* adds the meaning 'on, towards', while *-to-* means 'along with, under, on', as we see in (40):

(40) *chiak* 'to come'
***to**-chiak* 'to bring along, to pick up'

(Helberg 1984: 400)

Just like Hart, Helberg also signals that some combinations of shape morphemes have become semantically opaque, and that these should be considered semantic units. He analyzes the combination *-si'-po* (which means 'child' when attached to prefix *wa-*) as a derivative suffix which functions as a diminutive, and illustrates this with the following example:

(41) *wa* *-mä'mbuy* ***-si'*** ***-po***
NMLZ -brother **-SHM:peel** **-SHM:round/box.like**
'younger brother'

(Helberg 1984: 446)

Concerning prefix *wa-*, Helberg provides a slightly different analysis than Hart. According to Helberg, the prefix has two main functions: *nominalizer* and *prospective infinitive*. In its function of nominalizer, the prefix serves to form nouns on the basis of elements such as shape morphemes. In its function as prospective infinitive, *wa-* is mainly used on verbs in final subordinated clauses. As Helberg argues, it seems that Amarakaeri speakers attribute this same 'final' meaning to the prefix

[28] These two shape morphemes belong to the group which Hart defines as 'bound linkers'.

when it functions as a nominalizer, leading to an added semantic value of 'thing for V-ing', in many cases. In its function of nominalizer, the prefix has three allomorphs: *wa-*, *a-*, and *-o-*. When it functions as a prospective infinitive, however, the only form of the prefix is *wa-* (Helberg 1984: 189).

Helberg also states that, when combined with certain kinship terms, *wa-* can be considered a definite article:

(42) *mämä* 'grandmother' (vocative)
***wa**-mämä* 'the grandmother'
(Helberg 1984: 435)

As Helberg notes, this function of *wa-* corresponds with the translation of the prefix as definite article 'the' by Matteson (1972).

4.3 Tripp (1976, 1995)

Robert Tripp, who is affiliated to the SIL, started research on Amarakaeri in 1959. Tripp collected most of his data in the community of Puerto Luz, which is situated at the Colorado river. His studies resulted in a bilingual dictionary (1995), and a number of articles on the structural characteristics of Amarakaeri, one of which (1976) is also relevant to this research.

The dictionary consists of the sections Amarakaeri-Spanish and Spanish-Amarakaeri, and a few appendices, one of which contains a short grammar sketch. The lexicon includes quite a large number of entries, but seems to be slightly conservative. Some words for new cultural items (e.g. 'generator', 'television') are lacking, and since the aim of the dictionary was to present Amarakaeri words (Tripp 1995: 14), most loanwords from Spanish and Quechua are also left out. Although justifiable, this choice seems to make the dictionary less representative of actual Amarakaeri speech. In the dictionary, prefix *wa-* (which is represented as *hua-* in Tripp's spelling) is considered part of the word, which means that all words starting with *wa-* are listed together. In his introduction Tripp mentions that complex nouns, as well as body part nouns, appear in the dictionary with prefix *wa-*.

In the grammar sketch, Tripp distinguishes 6 types of nouns:

- *common nouns*
- *proper nouns*
- *vocative nouns*
- *composed nouns*
- *complex nouns*
- *derived nouns*

It is not clear what exactly defines the class of common nouns, but they seem to be morphologically simple; Tripp gives *apetpet* 'jaguar' as an example. Proper nouns are people or place names. Vocative nouns, as Tripp argues, are often kinship terms, which appear in a special form when functioning as a vocative, such as *pag'* 'father'. The class of composed nouns seems to be a somewhat diverse group, which contains compounds and several types of N + Adj combinations. The nouns which may be formed with prefix *wa-* are the complex and derived nouns. Complex nouns are those which contain one or more shape morphemes and carry prefix *wa-*; Tripp gives the examples *wa-ku-chi-po* 'thigh' and *wa-mba'-pi* 'finger'. It must be noted that the distinction between complex nouns and body part nouns remains unclear; it seems like Tripp uses both terms interchangeably. The class of derived nouns is also quite diverse, and includes nouns which are derived from verbs and are prefixed by *wa-*.

Interestingly, in a section on the possessive construction, Tripp notes that body part nouns can appear either with or without *wa-* in this type of construction, as is illustrated in examples (43) and (44):

(43) *ndo -edn wa **-mba'***
1SG-POS NMLZ **-SHM:hand/leaf**
'my hand'

(Tripp 1995: 196)

(44) *ndo -edn **-mba'***
1SG-POS-**SHM:hand/leaf**
'my hand'

(Tripp 1995: 196)

Tripp does not further discuss this phenomenon.

In a section on verbal morphology, some examples of verbs containing shape morphemes are given, and Tripp argues that the incorporated shape morpheme may refer to either the complement, location or subject of the verb. An instance of the latter is given in (45) below:

(45) *e' **-po** -bit*
INF **-SHM:round/box.like** -shrink
'to disinflate (subj.: ball)

(Tripp 1995: 220)

In the grammar sketch, the shape morphemes are defined as 'affixes or bound roots which indicate shape or figure' (Tripp 1995: 192, my translation); in the context of verb formation, however, Tripp coins them 'shape prefixes/classifiers' (Tripp 1995: 219, my translation). Prefix *wa-* is coined 'prefijo clasificador', which seems to be a translation of the term 'classifier prefix' used by Hart (1963).

A more elaborate definition of the shape morphemes can be found in Tripp's article "Los verbos Amarakaeri" (1976). In this work, which provides an analysis of the complex structure of the

Amarakaeri verb from a tagmemic perspective, the shape morphemes are mentioned as the components which form the 'classifier root'.[29] They form a large class of shape morphemes which may occur either alone or in groups or combinations, and form the core of the classifier noun[30] (Tripp 1976: 8), which may also occur in verb constructions.

All the works that have been discussed in this chapter are of great importance to the present research, and will be used in comparison with my own fieldwork data.

[29] Tripp distinguishes three non-verbal roots here: the nominal, adjectival, and classifier root.
[30] In this work, the nouns which are a combination of prefix *wa-* and one or more shape morphemes are coined 'classifier nouns' instead of 'complex nouns'.

5. Katukina

Katukina is a language family which is geographically situated in the western part of the Brazilian state Amazonas. According to Dos Anjos (2011), the Katukina family consists of two languages: (1) Katukina-Kanamari, which can be further subdivided in the dialects Kanamari and Katukina do Bia, and (2) Katawixi. Very little is known about Katawixi, and it remains unclear whether this language is still spoken or has become extinct. Speaker numbers for Katukina-Kanamari are estimated around 1250 (Dos Anjos 2011). Previous works on Katukina-Kanamari include Groth (1985), Queixalós & Dos Anjos (2007), and Queixalós (2004, 2006). A reference grammar of Katukina-Kanamari was published by Dos Anjos in 2011. Queixalós & Dos Anjos (2007) characterize Katukina-Kanamari as a morphologically simple language with a basic Agent-Verb-Patient constituent order. Other features of the language are a relatively isolating morphology, ergative alignment patterns, and noun incorporation (Queixalós 2004, 2006).

A for the classification of the Katukina language family, Adelaar (2000) proposes a genetic affiliation with the Harakmbut languages, based on a comparative study of (mainly) Katukina-Kanamari and Amarakaeri. The most striking resemblances between the two languages are found in the lexicon, and in particular, in the core vocabulary (cf. 5.1). On the basis of these cognates, systematic correspondences between the vowel inventories of the two languages are also established. In contrast, relatively few grammatical correspondences can be found. Instances of these are the prefix for the first person singular subject and the locative suffix (Adelaar 2000: 222). The comparative study of Katukina-Kanamari and Amarakaeri is complemented with data from Katawixi in Adelaar (2007). The apparent genetic affiliation between the Katukina and Harakmbut language families suggests that the Harakmbut languages were originally spoken in Brazil, and that their speakers have migrated to southern Peru at some point in history. The low number of lexical borrowings from Quechua in these languages also points towards a relatively late appearance of Harakmbut groups in the Madre de Dios region (Adelaar 2000: 234).

In the light of the proposed genetic affiliation between Katukina and Harakmbut languages, a closer look at Katukina is interesting for the present research, since it may provide new insights into the development of nominal classification in Amarakaeri. In this chapter, therefore, some of the features of Katukina languages and their correspondences with Amarakaeri will be discussed in more detail. Thereby, I will mainly rely on works about Katukina-Kanamari, since information on Katawixi is very scarce.

5.1 Cognates of Amarakaeri shape morphemes

As is shown by Adelaar (2000) a relatively large number of words or morphemes in Amarakaeri seem to have cognates in Katukina languages. Many of these are shape morphemes in Amarakaeri and full nouns (often body part terms) in Katukina. Whereas the comparison in Adelaar (2000) is mainly based on correspondences between Amarakaeri and Kanamari, data from the language

Katawixi is discussed in Adelaar (2007). The lexical correspondences between these three languages that involve Amarakaeri shape morphemes are listed in Table 7 below.[31]

TABLE 7
PROPOSED COGNATES OF AMARAKAERI SHAPE MORPHEMES IN KATUKINA-KANAMARI AND KATAWIXI
(after Adelaar 2000, 2007)

Amarakaeri	**Katukina-Kanamari**	**Katawixi**
-hai 'bone'	*[bara]hai* 'flesh' (*bara* = 'hunt, animal')	
-hën 'flesh'	*[bara]hai* 'flesh' (*bara* = 'hunt, animal')	*hãñ* 'meat'
-i' 'foot'	*i* 'foot'	
-idn 'tooth, tooth shaped'	*i* 'tooth'	
-kpo 'eye, sphere'	*iko* 'eye'	
-ku 'head'	*ki* 'head' *baki* 'wrist'	*icanga, [tu]kaẽ* 'head'
-mba 'hand, leaf-shaped'	*ba* 'hand' *baki* 'wrist'	
-mba'agn 'shoulder'	*pan* 'arm'	*upa, upáhe* 'arm'
-më' 'liver'	*ma* 'stomach'	
-mïn 'intestine'	*min* 'belly'	
-nëk 'belly'	*naki* 'inside'	
-nö' 'tongue'	*no[ko]* 'tongue' *no[naki]* 'mouth'	*no, nokú* 'tongue'
-öh 'nose'	*o(h)pak* 'nose'	*uhi* 'nose'

[31] As Adelaar (2000: 223) signals, some of the cognates that are proposed in this list are more certain than others, and some may not survive critical examination in future studies.

-pi 'stick-shaped object'		*pi* 'spine'
-po 'round object'	*[takara]pu* 'egg' (*takara* = 'chicken')	
-wë 'river, liquid'	*wa(h)* 'river' *watahi* 'water'	*wahã* 'river' *wa-* 'water'
-wih 'hair'	*poi* 'hair' *[ki]poih* 'hair'	*[örada]wihi* 'eyelashes'

Apart from correspondences with Amarakaeri shape morphemes, cognates of nouns in Amarakaeri are also mentioned, such as the word for 'house': *hak* (Amarakaeri) - *hak* (Katukina) - *kah* (Katawixi). The vast majority of all these cognates form part of the basic lexicon, which includes words that refer to body parts, kinship, hunting, agriculture, and natural phenomena. Considering this, the cognates make a strong case for the proposed genetic affiliation between Harakmbut and Katukina languages (Adelaar 2000). Interestingly, the lexical correspondences between Amarakaeri and Katukina seem more numerous than those between Katukina and Katawixi. This suggests that Katawixi has undergone a process of relexification, under influence of other languages (Adelaar 2007: 167). It must also be mentioned that a few of the Harakmbut-Katukina cognates are also found in Panoan languages such as Shipibo and Cashinahua, which are geographically located in between the Harakmbut and Katukina languages.

Examining the proposed cognates in Table 7, it appears that the Amarakaeri terms often have a broader semantic value than their counterparts in Katukina and Katawixi, e.g. *-po* 'round object' in Amarakaeri, as opposed to *pu* 'egg' in Katukina. According to Adelaar (2007: 165), this suggests that semantic specialization has taken place in Katukina languages. However, it could also be argued that these elements have undergone semantic generalization in Amarakaeri, during a process in which they grammaticalized into shape morphemes. The main question here is, therefore, whether the cognates were full nouns in the proto-language and have become grammaticalized in Amarakaeri, or whether they already had class marker properties in the proto-language and have been upgraded to full nouns in Katukina. What supports the former scenario is the supposed unidirectionality of the grammaticalization process, which is mentioned by Heine and Kuteva (2009); an argument for the latter is that Amarakaeri is, in many respects, more conservative than Katukina (Adelaar, p.c.). This issue will be further addressed in Chapter 7.

5.2 Incipient nominal classification

As is noted by Queixalós & Dos Anjos (2007) and Dos Anjos (2011), Katukina-Kanamari has a small set of lexico-grammatical elements that resemble class markers. These markers appear

suffixed to nouns, indicating the shape or substance of the referents of the nouns they occur with. A list of markers is given below:

an	'long and thin'
ba	'flat'
hai	'fibre'
kon	'round'
hi	'liquid'

The morphosyntactic status of these markers remains a point of discussion: while Queixalós & Dos Anjos (2007 denominate them 'classifiers', they are defined as 'generic nouns' in Dos Anjos (2011). As an argument for considering them as a subtype of nouns, Dos Anjos (2011: 140) mentions the fact that they occur as the second part of compounds, a slot that is usually filled by a noun and forms the head of the compound. Some instances are given in examples (46)-(48):

(46) *kana* ***-an***
sugar.cane (Pr.) **-long.thin**
'sugar cane'

(Dos Anjos 2011: 138)

(47) *yok -hi:k taro* ***-ba***
1SG -see leaf **-flat**
'I saw the leaf'

(Dos Anjos 2011: 138)

(48) *ma -o don* ***-hi***
3PL -drink fish **-liquid**
'They drink the fish soup'

(Dos Anjos 2011: 139)

It must be mentioned, however, that the function of class markers as head of NPs is attested in many Amazonian languages, such as we saw in languages like Miraña and Amarakaeri, where class markers have a strong derivational function and are difficult to distinguish from nouns. Indeed, Dos Anjos signals that the shape markers in Katukina-Kanamari are reminiscent of class markers, and that they seem to be undergoing a process of grammaticalization (Dos Anjos 2011: 140). This development is believed to be relatively recent, as most of the shape markers still have a full noun equivalent. A list of the markers and their corresponding nouns is given below:

an	'long and thin'	<	*an*	'leg'
ba	'flat'	<	*ba*	'hand'
hai	'fibre'	<	*hai*	'meat'
kon	'round'	<	*kon*	'seed'
hi	'liquid'	no corresponding noun		

As we can see, the semantic value of these markers is clearly related to the meaning of their full noun equivalents. It was mentioned in Chapter 2 that a clear an identifiable nominal origin of class markers in a language suggests a relatively early stage of grammaticalization. In this light, Katukina-Kanamari can be said to display an incipient system of nominal classification. It should also be noted that the system is quite productive, as the shape markers are often used in noun derivation and also appear on lexical borrowings, such as the noun *kana* which stems from Portuguese, as we saw in (46) (Dos Anjos 2011).

Defining the shape markers in Katukina-Kanamarí as incipient class markers seems realistic, considering the fact that shape-based systems of nominal classification are an areal trait in the region. It is not clear, however, how exactly this system has originated, and whether its development involves classificatory noun incorporation, as seems to be the case in Amarakaeri and languages like Kwaza and Mundurukú. In any case, it seems more likely that the system in Katukina-Kanamari has developed independently of Harakmbut, possibly under influence of neighbouring languages, than that it already existed in the proto-language of the proposed Harakmbut-Katukina family. Reasons to assume this are the fact that the system of nominal classification in Harakmbut languages involves a large set of class markers, and that it seems to be much more established than the small and incipient system that is found in Katukina-Kanamari. Moreover, *ba* 'flat' is the only Katukina-Kanamari marker that is also found in Amarakaeri (i.e. *-mba'* 'hand, leaf-shaped'); the other shapes and substances are expressed by different markers in the two language families.

5.3 Possible indications of the origins of Amarakaeri prefix *wa-*

Apart from finding cognates of the Amarakaeri shape morphemes in Katukina languages, it is interesting to look for elements that may correspond with the nominalizing prefix *wa-* from Amarakaeri, which is ubiquitous in word formation and often occurs in combination with shape morphemes.

A possible candidate is the generic relational noun *wa* in Katukina-Kanamari. Like many other Amazonian languages, Katukina-Kanamari displays a clear distinction between alienably and inalienably possessed nouns (Groth 1985, Queixalós 2006, Dos Anjos 2011). The group of inalienably possessed nouns mainly consists of kinship terms and body parts (both human and non-human), but also contains some other nouns that logically related to a possessor, such as the word for 'name' (Groth 1985: 2). The grammatical distinction between the two subclasses of nouns is based on their different behaviour in the genitive construction. Inalienable nouns simply carry a personal prefix in order to be possessed. This is illustrated in (49):

(49) *a -bakon*
3SG -finger
'his finger'

(Queixalós 2006: 14)

For alienably possessed nouns, the genitive construction is formed in a different way. Instead of being attached to the noun, the personal prefix is attached to the generic relational noun *wa*, followed by the full noun, such as we see in (50):

(50) *a* ***-wa*** *hak*
3SG **-GRN** house
'his house'
(Queixalós & Dos Anjos 2007: 16)

This generic relational noun is reported to have the very broad semantic value 'thing' (Dos Anjos 2011). According to Queixalós (2006: 15), it serves as a grammatical possessum in the genitive construction, and the relation between *wa* and the noun resembles that between a subject and its predicate nominal. It can therefore be argued that *wa* serves as a dummy noun, which fulfils the syntactic function of possessed noun but is semantically empty, and that it is specified semantically by the full noun which follows it. When a very general meaning is implied, the construction can also be used independently, without being followed by a full noun. This is illustrated in (51).

(51) a ***-wa***
3SG **-GRN**
'that (thing) of his'
(Queixalós & Dos Anjos 2007: 16)

Although it may well be unrelated to the generic relational noun *wa* that was described here, the antipassive prefix *wa-* should also be mentioned in this context. Dos Anjos (2011) describes antipassive *wa-* as one of the valency-changing mechanisms that exist in Katukina-Kanamari, and, more specifically, as an intransitivizing element. It replaces the verbal object, thereby turning transitive verbs into intransitives (52):

(52) a. *tatamtih nhama ma -pu bari*
there then 3PL -eat banana
'There then they ate bananas.'

b. ***wa*** *-pu nhama nuk*
AP -eat then they
'Then they ate.'
(Groth 1985: 20)

Groth (1985), defines *wa-* as a 'dummy object', since it does not refer to any patient in particular and is thus semantically empty. This raises the question whether the antipassive *wa-* may have the same origin as the generic relational noun, which serves as a semantically empty dummy noun.
The occurrence of the form *wa* as various types of 'dummy element' in Katukina-Kanamari is of particular interest to the discussion about the origin of prefix *wa-* in Amarakaeri. Since this prefix often occurs as a semantically empty 'dummy noun' in Amarakaeri, which serves to transform shape morphemes into full nouns[32], it could be argued that it might have the same origin as the

[32] Cf. the 'empty root construction' that was mentioned in Chapter 3.

generic relational noun and/or the antipassive prefix in Katukina-Kanamari. This is, however, a very tentative hypothesis, and it must be noted that while the generic relational noun in Katukina-Kanamari occurs with alienable nouns, prefix *wa-* in Amarakaeri seems to be used more frequently with inalienable nouns.

Another interesting clue for the origin of prefix *wa-* in Amarakaeri is provided by Adelaar (2000: 222), who signals that the noun *wadik* 'name' in Katukina-Kanamari seems to contain the same initial element *wa-* as many nouns in Amarakaeri. This idea is further sustained by the fact that the word for 'name' in Amarakaeri is *wa-ndik*, consisting of prefix *wa-* and the bound root *-ndik*, which may also occur without the prefix. This suggests that the element *wa* already existed in the proto-language which the Harakmbut and Katukina languages may stem from. Considering the fact that so far no other nouns have been found in Katukina-Kanamari that have an initial *wa-*, it could be argued that the prefix has disappeared from Katukina languages, while being retained in the Harakmbut family. Another possibility is that the prefix has been reanalyzed as a generic relational noun in Katukina languages. All this will be further discussed in Chapter 7.

5.4 Noun incorporation

Considering that the Amarakaeri system of nominal classification is claimed to have originated in classificatory noun incorporation (Derbyshire & Payne 1990), it is interesting to note that there is noun incorporation in Katukina-Kanamari as well. While in most Amazonian languages that incorporate nominals into the verb this construction is limited to inalienable nouns (Dixon & Aikhenvald 1999), Katukina-Kanamari also incorporates alienably possessed nouns. However, as we will see, these two subgroups of nouns are incorporated in very different manners.

As is described by Queixalós (2006), inalienably possessed nouns are simply prefixed to the verb in order to be incorporated. In this operation, the participant which is incorporated into the verb makes place for another participant, thereby granting it a more salient position in the sentence. This is illustrated in the example (53); while the focus lies on *(yo)ki* in (a), it is shifted to *adu* in (b), through incorporation of *ki* into the verb:

(53) a. *ti:k* *yo* ***-ki***
be.black 1SG **-head**
'I have a lot of hair' (lit.: my head is black)

b. ***ki*** *-ti:k* *adu*
head -be.black 1SG
'I have a lot of hair' (lit.: I am black-headed')
(Queixalós 2006: 11)

In the following example we see how noun incorporation may occur in a transitive verb. In (54), the original possessor (the child) is raised to the object position, resulting in a sentence that could be literally translated as 'the mother arm-scratched the child':

(54) *nyama -na=* ***pan*** *-unyuk opatyin*
mother -CASE **arm** -scratch child
'The mother scratched the child's arm'

(Queixalós 2006: 11)

As was mentioned before, alienably possessed nouns are incorporated in a different way. It is considered ungrammatical to just prefix these to the verb; the applicative prefix *ok-* has to be used to make the construction grammatically acceptable (Dos Anjos 2011).

(55) *Hayo -na=* ***ok*** ***-poako*** *-hu:na adu*
Hayo -ERG **APPL** **-oar** -hit 1SG
'Hayo hit my oar'

(Dos Anjos 2011: 355)

Although incorporation of alienable nouns is possible in Katukina-Kanamari, it is much less common than incorporation of inalienable nouns, especially in intransitive verbs. Queixalós (2006: 14) gives a possible explanation for this difference, arguing that the fact that inalienable nouns tend to occur in combination with their possessor causes an increased pressure on the predicate to incorporate the object noun, so that the possessor can be raised to the object position. This may lead to a more frequent incorporation of inalienably possessed nouns.

While noun incorporation in Amarakaeri is reported to have induced the development of a system of nominal classification (Derbyshire & Payne 1990), it remains uncertain whether the incipient system of nominal classification in Katukina-Kanamari, that was mentioned in 5.2, is related to noun incorporation as well. It must be noted, however, that body part incorporation - which is regarded a frequent source for nominal classification (Mithun 1986, Derbyshire & Payne 1990) - is frequently attested in Katukina-Kanamari, as we saw, for instance, in examples (53) and (54). In spite of the similarities between noun incorporation in Amarakaeri and Katukina-Kanamari, it is difficult to tell whether noun incorporation existed in the proposed proto-language, or whether these systems have developed independently, at a later stage.

In this chapter, some interesting similarities - both lexical and structural - between Katukina languages and Amarakaeri were discussed. Considering the fact that these languages may well belong to the same linguistic family, the signalled correspondences are highly relevant to this research. They will therefore play an important role in the formation of hypotheses about the origins and development of nominal classification in Amarakaeri, later on in this thesis. Moreover, they will serve as a frame of reference for my fieldwork data, which will be presented and discussed in the following chapter.

6. Fieldwork data

In this chapter my fieldwork data of the Amarakaeri shape morpheme system will be presented. During fieldwork, I focused on the group of shape morphemes which may occur independently when combined with prefix *wa-* (subclass 1 shape morphemes according to Hart). I first checked with my consultants whether all the shape morphemes that are listed by Hart (1963) and Helberg (1984) exist in the Boca Inambari language community. This was indeed the case, although some of the morphemes seemed to have slightly different meanings than recorded in previous works (for instance, to my consultants, *-mbogn* means 'mouth' rather than 'lip, edge'). Some other shape morphemes still exist, but do not seem to be the preferred term for the object or body part they refer to (e.g. *-hudn* exists, but *watapi* is more commonly used for 'upper back area'). Two of the shape morphemes mentioned by Hart (1963) and Helberg (1984) were not confirmed by my consultants: *-'ay* 'stalk like' and *-kwë* 'juice' (according to my consultants, only the more general *-wë* 'liquid, river' is used for juice). However, the vast majority of the shape morphemes as listed in previous works were confirmed by the speakers of Boca Inambari. Having established this, I further examined their semantics, uses and occurrences, which will be discussed in this chapter. I investigated their functions and distribution in different morphosyntactic contexts (6.1), the degrees of semantic generalization that they display (6.2), the different degrees of lexicalization they show in combination with nouns (6.3), and the ways in which the shape morphemes are used in neologisms (6.4) and with loanwords (6.5). Finally, I examined the functions and distribution of prefixes *wa-* and *e'-* (6.6). All this will be of use in the discussion about the origins and development of the system.

6.1 Functions and contexts of use of the shape morphemes

In this section, the functions and distribution of the Amarakaeri shape morphemes will be described. Apart from creating a clearer image of the contexts of use of these morphemes, this description will also serve to map the differences and similarities between shape morphemes and full nouns. This is highly relevant, since shape morphemes in some Amazonian languages are considered to be very similar to nouns. For instance, Weber (2002) argues that Bora class markers have the structural status of nouns, for they may occur independently, have the same referential properties as nouns and are distributed in a very noun-like fashion. Similarly, Amarakaeri shape morphemes share their referential properties with nouns and play an important role in word formation. What distinguishes them from nouns? The data that is presented in this section might shed light on this question.

6.1.1 Functions of the shape morphemes

As was noted by Payne (1987) and Grinevald & Seifart (2004), class markers in Amazonian languages typically combine derivational and agreement functions. The same seems to hold true for the Amarakaeri shape morphemes. Both functions, as found in my data, will be discussed below.

6.1.1.1 Derivation

Class markers in Amazonian languages such as Miraña, Kwaza and some Tucanoan languages have a strong derivational function and often occur as meaning-bearing elements when attached to nouns. Similarly, many words in Amarakaeri consist of compound-like structures that are combinations of noun roots with shape morphemes. In these structures, the shape morpheme is the right hand element and could be considered the semantic head, specified by the noun that precedes it. The N + SHM combination thus denotes an entity of the type that is denoted by the shape morpheme:

(56) *watawa**-'i***
chicken-**SHM:foot**
'chicken foot'

In this respect, the shape morphemes are very similar to nouns, for the right hand slot of compounds may also be filled by a noun:

(57) *siro* *-koso*
metal -pot
'metal pot'

However, in some of the N + SHM constructions it is less clear whether the shape morpheme functions as a semantic head, since it may be left out in some cases. For instance, in the case of *apoare* 'papaya', the word *apoare* itself can be enough to refer to papaya fruits, trees, juice, etc. According to my consultants, whether the shape morpheme can be left out in combinations like *apoare-po* depends on the context; if there is no possible ambiguity in reference, then using the noun alone is enough. Otherwise, the shape morpheme is needed in order to specify the reference (examples 59-61).

(58) *apoare*
papaya
'papaya (e.g. tree, fruit, etc.)'

(59) *apoare* ***-po***
papaya **-SHM:round/box.like**
'papaya fruit'

(60) *apoare* ***-wë***
papaya **-SHM:liquid**
'papaya juice'

(61) *apoare* ***-kidn***
papaya **-SHM:seed/small.round**
'papaya seed'

Seifart (2005) notes a similar phenomenon in Miraña and argues that some N + CL combinations do not clearly denote an entity of the kind that is denoted by the classifier, but could be argued to be

the intersection of the denotations of their parts. For instance, example (62) does not necessarily denote an 'oblong object of the banana type', but is just as much a 'banana of the oblong type'; therefore, one could say there is no hierarchical internal structure between the noun and the class marker (Seifart 2005: 202):

(62) *úɻhɨ* ***-ʔo***
banana **-CL:3D.oblong**
'banana fruit'
(Seifart 2005: 202)

In Amarakaeri, this type of semantic intersection is also found in some loanwords that are combined with shape morphemes (cf. 6.5).

Another important type of word formation in Amarakaeri, already mentioned by Hart (1963) and Helberg (1984), consists of just the shape morpheme, attached to prefix *wa-*. The result is a noun whose semantic value is the same as the core meaning of the shape morpheme itself. Prefix *wa-* therefore seems to function as a dummy noun, a semantically empty root that lends independent noun status to the shape morpheme. This construction is highly reminiscent of the empty root construction that was mentioned to exist in languages such as Kwaza, Kanoê and Gavião. An example of this construction in Amarakaeri is given in (63):

(63) *wa* ***-ku***
NMLZ **-SHM:head**
'head'

These combinations of *wa-* + SHM seem to function as full nouns. They can also be used to replace more specific terms in discourse. That is, the term *waku* comprises all kinds of heads (jaguar head, dog head, bird head, human head), and can be used as a more general term. For instance, my consultants would say 'All animals have a *waku*'. In order to further specify the semantic value of a *wa-* + SHM combination, more shape morphemes can be attached to it (64).

(64) *wa* ***-'öh*** ***-pa***
NMLZ **-SHM:nose** **-SHM:rod.shaped**
'trunk'

Overall, the shape morphemes play an important role in word formation, as is also noted by Hart (1963) and Adelaar (2000). An examination of Tripp's (1995) dictionary shows that many Amarakaeri words contain one or more shape morphemes. The same holds true for neologisms, which will be discussed in 6.4.

6.1.1.2 Agreement

While their derivational use makes them look quite similar to nouns, the shape morphemes also display agreement functions, which are more reminiscent of class markers. In my data, two types of agreement can be distinguished: agreement between a verb-incorporated shape morpheme and a

noun, and agreement at the NP level. Of these two, verb-incorporated agreement seems to be the most pervasive type, and has been extensively described by Hart (1963). The existence of this type of agreement is confirmed by my own data. Examples include:

(65) *(Kẽn) o* ***-ku*** ***-wih*** *-tegn.*
(he) 3SG **-SHM:head** **-SHM:hair** -cut
'He cuts his (head) hair.'

(66) *Pera* ***-po*** *o* ***-po*** *-yakay.*
rubber **-SHM:round/box.like** 3SG **-SHM:round/box.like** -kick
'He kicks the ball.'

(67) *Mbe* ***-hai*** *-yok.*
IM.SG→SG **-SHM:bone** -give
'Give me a bone.'

(68) *Num -da mbe* ***-pi*** *-yok wa-mandoya-****pi****.*
other-INT IM.SG→SG **-SHM:stick.like** -give NMLZ-write **-SHM:stick.like**
'Give me another pen.'

As we can see in examples (66) and (68), a noun may be added to specify the argument of the verb which is referred to by the shape morpheme. This creates agreement between the verb-incorporated shape morpheme and the overtly expressed noun. However, as is shown in (65) and (67), adding a noun is not always obligatory. The conditions that make the noun obligatory or optional do not remain entirely clear from the data, but it seems likely that shape morphemes which express only the object's shape, such as *-po* and *-pi* in examples (66) and (68), are in more need of semantic specification than those that already have a rather specific meaning, such as *-ku-wih* (65) and *-hai* (67)[33]. As is argued by Mithun (1986), agreement develops from noun incorporation when some of the V + N compounds have become so institutionalized that they cease to fulfil their initial task as referent tracking devices. It thus becomes necessary to overtly express the argument that the incorporated noun originally refers to, in order to clarify the reference. Institutionalization of V + SHM compounds indeed seems to have taken place in Amarakaeri. Some of the shape morphemes are often incorporated into the same verb; although still semantically transparent, these combinations could be considered fixed and are included as separate entries in Tripp's (1995) dictionary. Some of these are:

*e'**po**yakai'*	'to kick a round object'	<	*e'yakai'*	'to kick'
*e'**po**aga'*	'to inflate a ball'	<	*e'ag'*	'to inflate'
*e'**pid**tuka'*	'to punch in the ribs'	<	*e'tuka*	'to punch'
*e'**ba**tihuekea'*	'to dry/wash the hands'	<	*e'tihueke'*	'to clean'
*e'**kid**ene'*	'to sow'	<	*e'ene'*	'to disperse'
*e'**pi**tuk*	'to plant a stick'	<	*e'tuk*	'to plant'
*e'**da**putuj*	'to grab fruits from a tree'	<	*e'putuj*	'to bring down'

[33] Cf. 6.2 for a discussion on the different degrees of semantic generalization within the shape morpheme inventory.

The formation of this kind of verb, incorporating a shape morpheme, could be considered a type of derivation. Following Mithun's theory, one could thus argue that in this case, derivation may indirectly give rise to agreement.

The second type of agreement, which takes place within the NP level, is not specifically mentioned by Hart (1963), Helberg (1984) and Tripp (1995). Derbyshire & Payne (1990) do argue that Amarakaeri might have a nongender concordial system besides its verb-incorporated shape morpheme system, but they note that data is lacking in order to determine whether there is agreement on the NP level. In my data, a few instances of such agreement occur. These consist of ADJ + SHM constructions plus a modifying element (numeral, possessive pronoun, etc.). In example (69) the shape morpheme appears both on the numeral and in combination with the adjective:

(69) *mbot* ***-mba'*** *wamboru'* ***-mba'*** *-da*
two **-SHM:hand/leaf** big **-SHM:hand/leaf** -ADJM
'two big hands'

In example (70) we find a similar kind of agreement, where the shape morpheme is used both with the adjective and in the possessive construction, which is part of the same NP:

(70) *ndo-edn* ***-mba'*** *wamboru'* ***-mba'*** *-da*
1SG-POS **-SHM:hand/leaf** big **-SHM:hand/leaf** -ADJM
'my big hand'

This agreement within the NP looks similar to agreement in some Amazonian classifier languages that were discussed in Chapter 3, such as Miraña and Mundurukú. However, shape morphemes in those languages show a much stronger concordial function than shape morphemes do in Amarakaeri, since only a few instances of this type of agreement occur in my data. It must be noted, however, that my data set mainly consists of word elicitation and grammaticality judgements, and that it contains relatively few larger sentences. A different type of data set might well show more instances of agreement within the NP. This deserves attention in future research. It is well possible, however, that the scarcity of examples indicates that the nongender concordial function of Amarakaeri shape morphemes still finds itself in a preliminary stage.

6.1.2 Distribution of shape morphemes in other morphosyntactic environments

In the previous section, we saw that the shape morphemes resemble nouns in some respects and classifiers in others, depending on the different functions they may fulfil. In order to further determine the structural status of shape morphemes, this section will provide an overview of the ways in which they occur in combination with other linguistic elements, such as possessives, numerals and adjectives. As we will see, the use of nominalizing prefix *wa-* is obligatory in some of these morphosyntactic contexts, but optional or redundant in others. This is particularly interesting, since *wa-* is usually the element that grants independent noun status to the shape morphemes. Its degree of obligatoriness may thus shed light on the morphosyntactic status of these morphemes.

6.1.2.1 Possessive construction

As was already noted by Tripp (1995), the possessive construction may take two different forms when it involves shape morphemes: it can appear either as a possessive phrase including prefix *wa-*, or the possessive marker may be directly attached to the shape morpheme, a construction in which *wa-* is left out. In my data the same two possibilities occur: a possessive construction with or without prefix *wa-* (71, 72):

(71) *ndo -edn wa **-ku***
1SG-POS NMLZ **-SHM:head**
'my head'

(72) *ndo -edn **-ku***
1SG-POS **-SHM:head**
'my head'

For all of the shape morphemes both possessive constructions seem to be grammatical. Two consultants were individually asked to translate possessive phrases like 'my hand' and 'my nose' into Amarakaeri; in many cases, the two consultants prompted different constructions for the same shape morpheme. When asked for a grammaticality judgement about the construction which they had not mentioned themselves, the consultants marked these as being grammatical as well, without being able to mention any semantic difference between the two. This suggests that the two constructions can indeed be used interchangeably. Full nouns, such as *hak* ('house'), are always possessed in the following way:

(73) *ndo -edn hak*
1SG-POS house
'my house'

As we can see, in the general possessive construction the possessive is followed by a noun. In this light, the fact that the shape morphemes may be directly attached to the possessive, without being preceded by *wa-*, suggests a noun-like status. However, since the construction with *wa-* is also used, it could be argued that the shape morphemes are, in this respect, somewhere in between full nouns and bound roots.

6.1.2.2 Interrogatives

Just like in the possessive construction, shape morphemes may appear with or without *wa-* when combined with interrogatives. This is illustrated in examples (74) and (75), both of which were judged grammatical:

(74) *men **-'i***
INTERR **-SHM:foot**
'which foot?'

(75) *men -wa **-'i***
INTERR -NMLZ **-SHM:foot**
'which foot?'

6.1.2.3 Demonstratives

When used with demonstratives, the shape morphemes seem to behave as full nouns. Just like nouns (76), the shape morphemes directly follow the demonstrative (77):

(76) *nu hak*
DEM2 house
'that house'

(77) *nu **-ku***
DEM2 **-SHM:head**
'that head'

Insertion of *wa-* in this construction was judged ungrammatical (78).

(78) **nu wa-ku*
Intended: 'that head'

6.1.2.4 Numerals

Also when combined with numerals, the shape morphemes behave like nouns. Just like full nouns (79), shape morphemes are prefixed by the numeral (80), and the insertion of prefix *wa-* into this construction is judged ungrammatical (81):

(79) *mbot tak*
two candle
'two candles'

(80) *mbot **-mba'***
two **-SHM:hand/leaf**
'two hands'

(81) **mbot wa-mba'*
Intended: two hands

However, *wa-* is obligatory when it belongs to a word that is more complex than just a *wa-* + SHM construction, such as *wa'idnkoyo*:

(82) *mbot **wa** -'idn -koyo*
two **NMLZ** -SHM:tooth -clean(V)
'tooth brush'

This may be due to the fact that in the lexeme *wa'idnkoyo* prefix *wa-* nominalizes the verb rather than the shape morpheme.[34]

6.1.2.5 Adjectives

When shape morphemes are combined with adjectives, two types of construction are possible. In the first construction type, the shape morpheme is preceded by the adjective and followed by adjective marker *-nda* (83). In the second type, the shape morpheme precedes the adjective (84):

(83) *uru* ***-ku*** *-nda*
beautiful **-SHM:head** -ADJM
'beautiful head'

(84) ***öh*** *-mbedn*
SHM:nose-green
'green nose'

This corresponds with Hart (1963: 3), who notes that the adjective root may either precede or follow the shape morpheme. The fact that the shape morpheme can function as the first part of a construction, without being prefixed by *wa-*, is somewhat remarkable if we assume that the shape morphemes are bound roots. Just like in some other constructions, shape morphemes seem to behave more as nouns than as shape morphemes here. However, not all shape morphemes may appear without *wa-* in this construction, and for others the prefix is optional, as we see in examples (85) and (86), both of which are judged grammatical. This may indicate an ongoing change, in which some of the shape morphemes have advanced more than others.

(85) ***'idn*** *-sik*
SHM:tooth -black
'black tooth'

(86) *wa* ***-'idn*** *-sik*
NMLZ **-SHM:tooth** -black
'black tooth'

Interestingly, the combination of adjectives with full nouns is constructed in a different fashion. While the shape morphemes can be placed in between the adjective and the adjective marker, as was shown in (86), these two elements seem to be inseparable when combined with a full noun:

(87) *hak* *uru* *-nda*
house beautiful -ADJM
'beautiful house'

[34] The *wa* + SHM + V construction will be further discussed in 6.4.

(88) *uru* *-nda* *hak*
beautiful -ADJM house
'beautiful house'

6.1.3 Concluding remarks

In sum, the Amarakaeri shape morphemes resemble nouns in many respects. First of all, they display a strong derivational function and may serve as semantic heads in compound-like structures, just like nouns. Moreover, they have the same distribution as full nouns in many morphosyntactic contexts. On the other hand, shape morphemes do not seem to occur entirely independently, as nouns do. Apart from a few lexicalized cases, which will be discussed in 6.3, they only appear prefixed by nominalizer *wa-*, or attached to other elements, such as nouns or adjectives. Moreover, the agreement functions which the shape morphemes display suggest a class marker status.

On the basis of the present data, it thus seems like the shape morphemes find themselves at an intermediary stage on the grammaticalization scale, somewhere in between full nouns and class markers. The fact that the use *wa-* is optional in some environments and obligatory in others, could also be an indication of this.

Finally, it must be noted that agreement within the NP seems very limited in Amarakaeri, in comparison with other Amazonian classifier languages. This confirms the idea by Derbyshire & Payne (1990), that Amarakaeri should be considered a primarily verb-incorporated classifier language.

6.2 Semantic scope and generalization of shape morphemes

As we saw in Chapter 2, there is general agreement on the idea that nouns are the most common source ofclass markers. The process of grammaticalization which nouns undergo when transforming into class markers tends to go hand in hand with semantic change (Aikhenvald 2000a, Seifart 2005). Most commonly, although not always, these nouns undergo semantic generalization, becoming applicable to more referents than before. The word for 'eye', for instance, often becomes a classifier for small round objects (Aikhenvald 2000a: 405). Departing from this idea, combined with Mithun's (1986) theory that most verb-incorporated classifiers are derived from body part nouns, it seems well possible that the Amarakaeri shape morphemes have been derived from nouns, mainly body part terms, and that they have gradually extended their meanings towards shape, similarly to the type of development that seems to have occurred in some other Amazonian languages such as Mundurukú. This idea is supported by the fact that several of the Amarakaeri shape morphemes seem to have cognates in Katukina which display less broad semantic scopes than their Amarakaeri counterparts. This suggests that nouns from the proto-language may have grammaticalized and semantically extended in Amarakaeri, while remaining the same in Katukina. On the other hand, Adelaar (2007) argues that the difference in semantic scope between the cognates in the two languages might be due to semantic specialization in Katukina (Adelaar 2007: 164). Possibly, the shape morphemes already had a class marker status in the proto-language, and were upgraded to full

nouns in Katukina. The fact that the Katukina languages are in many respects more innovative than those of the Harakmbut family (Adelaar, p.c.), seems to strengthen this claim. However, if we presume that grammaticalization and semantic broadening often occur simultaneously, the occurrence of one of these two processes could indicate that the other is also taking place. In this section, therefore, I will examine the semantic scope of the shape morphemes, and compare the semantics of their full noun cognates in Katukina with their meanings in Amarakaeri, both in neologisms and in more traditional words. In this way, I will try to determine to what extent these morphemes have undergone semantic bleaching, and where each of the morphemes can be placed on a scale of semantic generalization. This, in its turn, may provide clues about the development of the shape morpheme system.

During fieldwork, my main method for gathering data on the semantic scope of the shape morphemes was elicitation. First, I verified the core meaning of each shape morpheme as noted by Hart (1963) and Helberg (1984), by asking my consultants to translate *wa* + SHM combinations and by taking the first (or most frequently) prompted translation of each of these combinations as the core meaning. I would then come up with items that I thought might be included in a certain shape morpheme's semantic scope, and ask my consultants whether this would be a correct reference, e.g. 'Can a banana be a *wanda*?' or 'Is the seed of an avocado a *wahai* too?' For every shape morpheme I also asked my consultants to tell me what kind of items it could refer to, and whether there was anything else, apart from its most obvious meaning, that it could be used for. The resulting data indicates that the shape morphemes display very different degrees of semantic generalization. I therefore divided them into three main groups: shape morphemes with a specific meaning, shape morphemes with a core meaning and semantic extensions, and shape morphemes that express only shape. These three groups will be discussed below.

6.2.1 Shape morphemes with a specific meaning

Many of the shape morphemes display a very specific meaning. That is, it seems that these shape morphemes can only be used to denote their core (usually body part) meaning, and that they do not have a broader semantic value related to shape, texture or function. Therefore, it could be argued that this subset of shape morphemes is semantically the most noun-like, since these morphemes have a very specific reference.[35]

An example of a specific shape morpheme is *-nda*, which is only used to refer to medium-sized round or oblong fruits (89, 90). Bigger fruits of that shape, such as papaya, are classified as *-po* instead of *-nda* (91), and also differently shaped fruits, such as banana (92), are excluded from the semantic scope of this shape morpheme:

(89) *wa* ***-nda***
NMLZ **-SHM:round/oblong.fruit**
'round/oblong fruit'

[35] It must be noted, however, that the fact that no semantic extensions of these shape morphemes appear in my data does not necessarily mean that these shape morphemes have no semantic extensions at all; I might simply not have encountered them.

(90) *kumewi* ***-nda***
guava **-SHM:round/oblong.fruit**
'guava (fruit)'

(91) *apooare* ***-po***
papaya **-SHM:round/box.like**
'papaya (fruit)'

(92) *aroy* *-Ø*
banana -Ø
'banana'

It can therefore be concluded that the semantic scope of *-nda* has not been extended to the general meaning 'fruit'. Moreover, there seem to be no shape-based extensions for this shape morpheme, since other medium-sized round or oblong objects (e.g. stones) cannot be classified with *-nda* either.

Another instance of a specific shape morpheme is *-nörë*, 'heart'. Although it can refer to the human heart and the heart of any type of animal, it does not seem to have any other applications. It does not display the more general meaning 'centre', nor can it be used to refer to any other internal organs of the body. More instances of specific shape morphemes are *-hai* 'bone', *-min* 'bowel', *-no* 'tongue' and *-kok* 'face'. All of these seem to be semantically restricted to a very specific referent. It must be noted, however, that even these specific body part shape morphemes can be used to refer to these body parts in animals as well; they are thus not restricted to the human body.

6.2.2 Shape morphemes with a core meaning and semantic extensions

Some shape morphemes have a clear core meaning, but also show semantic extensions. These extensions can be related to either shape, function, position, texture, material, or a combination of those. An example is shape morpheme *-mba'* 'hand/leaf'. It seems to be related to the full noun *ba* in Katukina, which combines the same two meanings (Adelaar 2000, 2007). This suggests that a shape-based extension from 'hand' to 'leaf' already took place in the proto-language. In Amarakaeri, the shape morpheme has acquired an even broader meaning, as will be shown below.

The morpheme has retained 'hand' as its core meaning, when attached to dummy noun *wa-*:

(93) *wa* ***-mba'***
NMLZ **-SHM:hand/leaf**
'hand'

In this quality, *-mba'* also appears in compound-like derivational constructions with nouns. Not only can it refer to human hands, it is also used for animal feet. However, it is restricted to the animal's front feet, since the back feet are always classified with *-i'* ('foot'). This indicates that the core of the analogy *human body part* → *animal body part* lies in the function or position of body parts, rather than their shape. That is, animals whose front feet do not resemble the shape of human hands, such as the tapir, are still considered to have *wamba*:

(94) *keme* ***-mba'***
tapir **-SHM:hand/leaf**
'front foot of tapir'

The semantic extension from 'hand' to 'leaf' seems to be based on the resemblance in shape between hands and some leafs, like the tobacco leaf. However, the originally shape-based extension seems to have been extended further, since *-mba'* can now refer to leaves in all shapes, such as the oblong-shaped banana leaf:

(95) *aroy* ***-mba'***
banana **-SHM:hand/leaf**
'banana leaf'

Another interesting manifestation of *-mba'* is in the noun *e'mba'* (96), which is a combination of *-mba'* with infinitive marker *e'-*, which occurs as an infinitive marker. The usual strategy for converting a shape morpheme into a noun is by adding prefix *wa-* to it. However, the noun *wamba'* already exists, meaning 'hand', as we saw in (93). This suggests that the use of infinitive marker *e'-* as a nominalizer for shape morphemes may be used as a second option when prefix *wa-* is already taken. [36]

(96) *e'* ***-mba'***
INF **-SHM:hand/leaf**
'leaf'

Shape morpheme *-mba'* has undergone more semantic extensions; it displays the semantic value 'sheet of paper' in some neologisms. This can be illustrated by its use in words for 'bank note' in examples (97) and (98). This semantic extension from 'leaf' to 'sheet' seems to be shape-based.

(97) *yögnki* ***-mba'***
money **-SHM:hand/leaf**
'bank note'

(98) *bota -diez* ***-mba'***
two -ten(Sp.) **-SHM:hand/leaf**
'twenty (soles) note'

Finally, a metonymic extension (*pars pro toto*) has taken place in the word for 'book', where *-mba'* refers to a collection of paper instead of a single sheet:

(99) *e'* *-mandoya'* ***-mba'***
INF -write **-SHM:hand/leaf**
'book'

[36] The use of infinitive marker *e'-* as a nominalizer will be discussed in 6.6.

In sum, shape morpheme *mba-* has become semantically extended from 'hand' to 'leaf', 'sheet' and 'paper'. This shows that the shape morpheme has undergone semantic bleaching towards the general shape-based meaning of 'flat and flexible'. The fact that this process is also attested in neologisms, like *yögnkimba'* and *e'mandoya'mba'*, suggests that the broadening of the semantic scope of this shape morpheme might be an ongoing process, although more extensive data would be needed to confirm this.

Another interesting case of semantic extension is shape morpheme *-kpo*, which according to Hart (1963) and Helberg (1984) has the semantic value 'eye/sphere'. It appears to have the noun equivalent *iko* in Katukina, meaning 'eye' (Adelaar 2000), which also suggests that this is the core meaning of the present shape morpheme. Moreover, the combination of *-kpo* with prefix *wa-* still has the semantic value 'eye':

(100) *wa* ***-kpo***
NMLZ **-SHM:eye/light.sphere**
'eye'

The shape morpheme is also found in compound-like constructions, where it can refer to the eye of any animate being, as is illustrated in examples (101) and (102):

(101) *harakmbut -o* ***-kpo***
human -EUPH **-SHM:eye/light.sphere**
'human eye'

(102) *biwi -o* ***-kpo***
snake -EUPH **-SHM:eye/light.sphere**
'snake eye'

Apart from the meaning 'eye', however, morpheme *-kpo* is also found in words where it has the meaning 'light-emitting sphere', such as *ñokpo* 'sun' and *shokpo* 'star'[37]. A similar meaning is found in neologisms. For instance, the front part of a flash light, which is round and emits the light, is called *wambachayokpo*:

(103) *wambachay -o* ***-kpo***
enlighten -EUPH **-SHM:eye/light.sphere**
''eye' of flash light'

Also, the word for 'light bulb' is made up of the word for generator plus shape morpheme *-kpo* (104). Again, the morpheme is used here to designate a light-emitting round object.[38]

[37] As will be discussed in Section 6.3, in these words the shape morpheme has merged with the nominal root.
[38] The direct connection between a generator and a light bulb might not seem very obvious, but makes perfect sense in the environment of the Boca Inambari community, where generators are the only source of electricity, and where light bulbs automatically switch on whenever the generator is turned on. This illustrates how culture-specific semantic extensions tend to be.

(104) *tuktuk* *-po* ***-kpo***
generator -SHM:round/box.like **-SHM:eye/light.sphere**
'light bulb'

In sum, it seems like the semantic value of *-kpo* has been extended from 'eye' towards 'light-emitting round object'. It must be noted that, contrary to the conclusions of Hart (1963) and Helberg (1984), none of the cases in my data point to extensions towards the more general shape-based meaning 'sphere'. It seems to me that *-po* rather than *-kpo* is the default shape morpheme for spherical items. It is interesting to note that all of the 'extended' uses of *-kpo* seem to be related to light. The connection between 'eye' and 'light' may be that both are closely related to sight. When I asked Manuel why a light bulb was considered a *wakpo*, he said: 'Because it is like an eye: it shines.'

Other instances of shape morphemes with semantic extensions that deviate from their core meaning are shown in Table 1 below.

TABLE 8
SEMANTICALLY EXTENDED SHAPE MORPHEMES

Amarakaeri shape morpheme	**Proposed cognates in Katukina-Kanamari (Ka) and Katawixi (Kw) (Adelaar 2000, 2007)**	**Core meaning in Amarakaeri**	**Semantic extensions in Amarakaeri**
-ën	*[bara]hai* 'flesh' (Ka) *hãñ* 'meat' (Kw)	'meat'	• wood
-'i	*i* 'foot' (Ka)	'foot'	• wheel
-'idn	*i* 'tooth' (Ka)	'tooth'	• tooth shaped objects (e.g. small banana) • objects which pierce (e.g. arrow tip, tree spine)
-kidn	-	'small round seed'	• other small round objects (e.g. beads, little lumps of gold) • large seeds (e.g. avocado seed) • oblong seeds (e.g. cucumber seed)
-kpo	*iko* 'eye' (Ka)	'eye'	• light-emitting sphere (e.g. sun, light bulb)

-ku	*ki* 'head' (Ka) *[tu]kaẽ* 'head' (Kw)	'head'	• upper part of something (e.g. tongue tip)
-mba'	*ba* 'hand' (Ka)	'hand, leaf-shaped'	• flat and flexible objects • paper → book etc. • differently shaped front feet of animals (e.g. hoofs)
-mbo	-	'leg'	• table leg
-mbogn	-	'lips'	• other mouth-shaped objects (e.g. wound, ravine)
-o(pi)	-	'arm'	• branch of tree
-wë	*wa(h)* 'river' (Ka) *watahi* 'water' (Ka) *wahã* 'river' (Kw) *wa-* 'water' (Kw)	'river'	• other liquids (e.g. juice, rain, gasoline)

Examining this table, we see that many of the extensions are not purely shape-based, but correspond to a certain 'body part analogy'. That is, animal body parts and the different parts of trees or plants tend to be denoted by the same shape morphemes as human body parts. For instance, most trees are made up of *-opi* 'arm' (branches) and *-mba'* 'hand' (leaves) and may also have *-'idn* 'tooth' (spines). This analogy also applies, in some cases, to components of man-made objects. For example, a vehicle's wheels are referred to as *-'i* (foot), and a table leg is *-mbo* (leg).

While some of the extensions are clearly shape-based, such as *-mbogn* 'lips' for mouth-shaped objects, others seem to be based on the position of the referent in the entity it is part of, such as *-ku* 'head' for the upper part of an entity. Other extensions, like *-ën* 'meat' for wood, are primarily based on texture. My consultant indicated that it is the tough, fibrous structure of wood which makes it an instance of *waën*, while the 'flesh' of fruits is too soft to belong to this group. Still other extensions, such as *'idn* 'tooth', seem to be based on both function and shape.

In sum, there is no single line along which semantic extensions are made. The largest common denominator of these extensions is that they depart from (usually) a body part meaning, and are based on one or more properties of that body part (shape, texture, function, position).

6.2.3 Shape morphemes that express only shape

The third group that I distinguished consists of shape morphemes that only denote shape, in any context. Not for all of these shape morphemes can a body part (or otherwise nominal) origin be found. Other than is the case for the shape morphemes that were discussed in the previous section, the semantic scope of these morphemes does not show any restrictions based on properties such as function, texture or position; they are the default shape morphemes for objects of a certain shape. This group mainly consists of the shape morphemes which have a *-pV* structure: *-pa* 'rod shaped', *-pe* 'disk shaped', *-pi* 'stick-like', *-po* 'round, box-like', and *-pu* 'tube shaped'. Furthermore, shape morpheme *-mbih* 'string-like' seems to belong to this group, since it refers to a certain shape rather than to anything else. The fact that these shape morphemes do not only have the broadest semantic scope but are also phonologically the simplest of the shape morphemes, is interesting in the light of the hypothesis as mentioned by Seifart (2005), that semantic generalization and grammaticalization are often parallel to phonological simplification. On the other hand, it must be noted that several of the shape morphemes from this group (*-pi*, *-pa* and *-pu*) have classifier counterparts in Tariana with the same semantic value. This suggests that these morphemes have entered Amarakaeri through direct diffusion rather than being the result of grammaticalization and semantic broadening.

As we will see in section 6.4, the shape morphemes that denote only shape are frequently used in the formation of neologisms, presumably because of their broad semantic scope. This is already noted by Hart (1963), who gives the following examples of N + SHM neologisms which are formed with the noun *siro*, 'metal, glass, plastic':

siro-pi	'small nail, needle'
siro-pa	'large nail, metal rod'
siro-po	'tin can'
siro-pu	'glass bottle, metal tube'
siro-mbih	'wire, plastic fish line'

(Hart 1963: 1)

Probably the most generalized and most frequently used shape morpheme of this group is *-pi*, which can be used to refer to any long and thin, 'stick shaped' object, as is illustrated in the following examples:

(105) *wa* ***-pi***
NMLZ **-SHM:stick.like**
'stick-like object'

(106) *öwey* *-ën* ***-pi***
tree -SHM:flesh **-SHM:stick.like**
'wooden stick'

(107) *Jessica* ***-pi***
Jessica **-SHM:stick.like**
'(skinny) Jessica'

(108) *wa* *-mandoya'* ***-pi***
NMLZ -write **-SHM:stick.like**
'pen'

(109) *lapiz* ***-pi***
pencil (Sp.) **-SHM:stick.like**
'pencil'

As we can see in these examples, *-pi* is also used with proper names (107), in the formation of neologisms (108), and to 'amarakaerianize' Spanish loanwords (109).[39] As was mentioned in Chapter 3, the use of *-pi* or similar elements is widespread in Amazonian languages, and it also exists as a classifier for long, thin and vertical objects in Tariana (Aikhenvald 2000b). Adelaar (2007) also mentions a possible cognate *pi* 'spine' in Katawixi, but the uses of shape morpheme *-pi* in my data do not indicate any relation to this body part meaning. It is possible that a word **pi* 'spine' existed in the proto-language of Harakmbut and Katukina, and that it has undergone extensive semantic generalization in Amarakaeri. However, it also seems likely that *-pi* has been borrowed from a neighbouring language, instead of having been derived from a body part noun.

Shape morpheme *-pe*, in contrast, does show a clear body part origin. When combined with dummy noun *wa-*, it has the meaning 'cheek'. It may also refer to animal's cheeks, such as the dog's:

(110) *wa* ***-pe***
NMLZ **-SHM:cheek/disk.shaped**
'cheek'

(111) *kuhua* ***-pe***
dog **-SHM:cheek/disk.shaped**
'side of dog's face'

However, as we can see in the following examples, the use of this shape morpheme in neologisms indicates that it has undergone a process of semantic extension, and that it now has the more general semantic value 'flat round object' (112, 113) or even simply 'flat object' (114).

(112) *wa* *-chawaya'* ***-pe***
NMLZ -watch **-SHM:cheek/disk.shaped**
'satellite receiver'

(113) *wa* *-ku* ***-pe***
NMLZ -SHM:head **-SHM:cheek/disk.shaped**
'coin'

(114) *wa* *-okah* ***-pe***
NMLZ -SHM:wing **-SHM:cheek/disk.shaped**
'wing of airplane'

[39] The use of shape morphemes with loanwords will be discussed in more detail in Section 6.5.

Considering this development, *-pe* seems to represent a transitory stage between the group of shape morphemes which have a core (often body part related) meaning and semantic extensions (such as *-kpo*), and those that have become so generalized that they denote only shape. This strengthens the idea that there is a general movement from specific body part terms towards morphemes that denote only shape.

Another interesting case is *-po* 'round, box like'. Adelaar (2007) proposes *pu* (Katukina) and *po* (Katawixi), which mean 'egg', as possible cognates. It must be noted that *-po* cannot be used in this sense in present-day Amarakaeri, but it seems well possible that the term existed in the proto-language and has been semantically generalized in Amarakaeri. Its semantic value is now mainly shape-based. While there are a number of more specific shape morphemes for round things (e.g. *-nda* for round or oblong fruits, *-kpo* for light-emitting spheres, *-knda* for eggs, and *-kidn* for small round things), *-po* seems to be the default shape morpheme for spherical objects. Items which are considered *wapo* by my consultants include footballs, melons, papayas, metal barrels and chubby people; virtually anything - or anyone - that is quite large and either roundish or box shaped. Interestingly, however, in its use in neologisms this shape morpheme seems to have acquired a new, function-based semantic value, which slightly deviates from its purely shape-based meaning. In neologisms, shape morpheme *-po* can be used for any object that is some kind of machine or electronic device. The word for 'generator', for instance, is formed by an onomatopoeic word combined with shape morpheme *-po* (115).

(115) *tuktuk* ***-po***
generator **-SHM:round/box.like**
'generator'

When I asked my consultant why the generator was named like that, he answered 'because it is a machine (and it makes that sound)'. This is one of the indications that *-po* has acquired the meaning 'machine'. In addition to that, the 'pequepeque'-engine (long tail engine), which can hardly be considered round or box-like, is also formed by attaching *-po* to an onomatopoeia:

(116) *pequepeque* ***-po***
pequepeque **-SHM:round/box.like**
'pequepeque engine'

Moreover, all kinds of electronic devices are formed with *-po*. While the use of *-po* for 'television' (117) can be justified by its box-like shape, the shape connection is much less obvious for 'mobile phone' (118).

(117) *wa* *-chawaya'* ***-po***
NMLZ -watch **-SHM:round/box.like**
'television'

(118) *wa* *-mana'pak* ***-po***
NMLZ -talk **-SHM:round/box.like**
'mobile phone'

Clearly, this supposed 'machine' meaning cannot be the origin that shape morpheme *-po* was derived from, since, by definition, it only occurs in modern objects. This suggests that there is not only a movement towards semantic generalization, but that even the most generalized shape morphemes may, in the course of time, develop more specialized meanings alongside their core meaning.

6.2.4 Concluding remarks

In this section, it has been shown that the shape morphemes display very different levels of semantic generalization: while some have a highly specific meaning, others show semantic extensions from their core meaning, and some shape morphemes can be used to denote a certain shape in any context. For some of the shape morphemes which denote only shape, no clear nominal origin can be found. It seems well possible that these were borrowed from neighbouring languages, particularly since some of them occur in Tariana with the same form and semantic scope. Apart from this small group, many of the shape morphemes seem to have nominal origins in the proto-language. The fact that these morphemes (e.g. *-mba'*, *-kpo*) display semantic extensions in Amarakaeri, suggests that they have undergone a process of semantic generalization. It seems likely that the shape morphemes have evolved from body part nouns to morphemes with broader semantic scopes, and that the three types of shape morphemes that were mentioned in this section represent different points on a scale of semantic generalization. A confirmation of this can be found in shape morphemes that seem to function on the edge of two of the categories that were established in this chapter, such as *-pe* (which still has a body part meaning, but denotes only shape in neologisms), and *-kidn* (which mainly means 'seed', but can also be used as a more general shape morpheme for anything that is small and round). Whether this presumed process of semantic broadening is ongoing, is difficult to tell, but it must be noted that some of the shape morphemes show their highest degree of semantic generalization when used in neologisms. Interestingly, however, at least one of the shape morphemes also shows semantic specialization in neologisms, namely *-po* 'machine, electronic device'.

In sum, comparing the shape morphemes' cognates in Katukina with their semantic extensions that are displayed in my data, it seems likely that the Amarakaeri shape morphemes have been derived from nouns and that they have broadened their semantic scope through time, parallel to a process of grammaticalization. This is in line with claims by Mithun (1986), Aikhenvald (2000a) and Seifart (2005) about the nominal origins of classifiers and the parallel occurrence of grammaticalization and semantic broadening. However, it is not impossible that this hypothesis is incorrect, and that an opposite development (i.e. semantic specialization in Katukina) has taken place instead, as is argued by Adelaar (2007). More data from Katukina would be needed in order to investigate this possibility.

6.3 Semantic conventionalization, lexicalization and merging

As was shown in 6.1, the Amarakaeri shape morphemes often occur in combination with nouns. The choice and use of shape morphemes in these combinations show different degrees of conventionalization in different words. That is, in some cases the decision to use a certain shape morpheme can be made 'on the spot', while some other N + SHM combinations are rather fixed and seem to have become lexicalized, thereby losing their semantic transparency. Furthermore, in some cases the noun root seems to have merged with the shape morpheme. These processes will be discussed here.

6.3.1 Spontaneous use of shape morphemes

In some of the N + SHM combinations in my data, the choice for a certain shape morpheme seems spontaneous. For instance, when speaking about the possible uses of *-pi*, my consultant commented that it could be added to my name, because of my 'stick-like' body shape. Had I been shorter and heavier, Manuel stated, he would have called me *Jessica-po*:

(119) *Jessica-**pi***
Jessica-**SHM:stick.like**
'(skinny) Jessica'

(120) *Jessica-**po***
Jessica-**SHM:round/box.like**
'(chubby) Jessica'

In this case, the choice of the shape morpheme depends entirely on the shape of the specific referent which the noun refers to. This can, therefore, be considered the least conventionalized use of the shape morphemes. Helberg (1984) distinguishes the shape morphemes that can be used in this way as a separate group: 'shape morphemes that can be combined with prefix *wa-* and can be used as nominal affixes' (Helberg 1984: 247, my translation). According to Helberg, this group consists only of the shape morphemes *-po* and *-pi*. However, I would argue that these are not the only shape morphemes that can be used as such. Although one could say that in combinations like *Jessica-pi* the shape morpheme can easily be left out, and therefore has a much less noun-like function than shape morphemes in words such as *watawa-'i* 'chicken foot', the distinction does not seem to be as clear-cut as is suggested by Helberg. As was discussed in 6.1, shape morphemes can be left out in some N + SHM combinations; it could be argued that in these cases, the shape morpheme functions as a nominal affix rather than having a noun-like function. Such combinations are not only formed with *-pi* and *-po*, but may just as well involve other shape morphemes.

Data from neologisms may also shed light on this matter, since these are in fact descriptions of novel objects, and can therefore be expected to be less conventionalized than nouns that have existed in the language for longer. As will be discussed in 6.4, not all neologisms are agreed upon yet within the Amarakaeri speech community, and I found variety in neologisms between different speakers. For instance, the word for 'satellite receiver' can be formed by using either *-pe* or *-po*:

(121) *wa -chawaya' **-pe***
NMLZ -watch **-SHM:cheek/disk.shaped**
'satellite receiver'

(122) *wa -chawaya' **-po***
NMLZ -watch **-SHM:round/box.like**
'satellite receiver'

This suggests that also the use of shape morphemes in some neologisms is still somewhat free and spontaneous, and that it involves a wider range of shape morphemes than just *-pi* and *-po*. Seifart (2005) describes a 'Shape Classifier Task' which was carried out in Miraña; an elicitation task in which speakers have to describe novel objects to one another. It would be interesting to perform such a task in Amarakaeri in future research, in order to learn more about the spontaneous use of shape morphemes in the language.

6.3.2 Semantic conventionalization

In other cases, combinations of certain shape morphemes with certain nouns show a higher degree of semantic conventionalization. A case in point is the use of shape morpheme *-po* with the noun *apoare* 'papaya'. While most fruits (orange, carambola, guava) are formed by using shape morpheme *-nda* (round/oblong fruit), papaya is formed with *-po*. My consultants explained that the papaya fruit is too big to be a *wanda*, and that it is therefore considered a *wapo*. However, papaya fruits are obviously not all equal in size. I therefore asked my consultants whether a *small* papaya could be referred to as *apoare-nda*, but this combination was immediately rejected (123). Instead, a combination which involves the diminutive (124) was mentioned by my consultants:

(123) **apoare **-nda***
papaya **-SHM:round/oblong.fruit**
Intended: 'small papaya'

(124) *apoare **-po** -sipo*
papaya **-SHM:round/box.like** -DIM
'small papaya'

This suggests that any mature papaya is inherently a *wapo* and can never be a *wanda,* for however small a specific member of the group may be; therefore, the combination *apoare-po* can be considered fairly conventionalized.

6.3.3 Lexicalization

In the case of *apoare-po-sipo*, the choice of shape morpheme *-po* seems to be based on a prototype (the average papaya) rather than a specific referent. This shows that it has a higher degree of conventionalization than combinations such as *Jessica-pi*, which are chosen on the spot. However, although conventionalized, the combination *apoare-po-sipo* is still semantically transparent. In some other cases, N + SHM combinations have not only become fixed, but have also lost their semantic transparency. This is illustrated in example (125):

(125) *wa* ***-ku*** ***-chi*** ***-po***
NMLZ **-SHM:head** **-SHM:extension** **-SHM:round/box.like**
'thigh'

The semantic opacity of this word lies in the fact that its meaning is not the direct result of the sum of its different parts. Moreover, shape morpheme *-pi*, may be attached to it, which renders the somewhat contradictory shape morpheme sequence *-po-pi-pi*:

(126) *wa* ***-ku*** ***-chi*** ***-po***
NMLZ **-SHM:head** **-SHM:extension** **-SHM:round/box.like**
-pi ***-pi***
-SHM:stick.like **-SHM:stick.like**
'skinny legs' (derogatory)

This corroborates the idea that the combination of shape morphemes in *wakuchipo* has become highly conventionalized and that it can be considered lexicalized. Interestingly, Hart (1963) mentions some instances of shape morpheme combinations that have undergone considerable semantic shift and may occur without prefix *wa-*, such as *kusi'pe* 'canoe':

(127) ***ku*** ***-si'*** ***-pe***
SHM:head/crown **-SHM:peel** **-SHM:disk**
'canoe'

(Hart 1963: 3)

This independent occurrence is of particular interest, since it suggests that some combinations of shape morphemes may not only become semantically opaque, but can also acquire full noun status. It must be noted, however, that independent use of lexicalized shape morpheme combinations does not seem to be widespread in the language. Hart only mentions a few instances of this use, most of which have shape morpheme *-ku* as a first member.

6.3.4 Merging

As was shown in the previous section, combinations of shape morphemes become fixed and semantically opaque when lexicalization occurs. A similar, but slightly different process is merging. In this process, a morpheme combination does not only become fixed, but its parts also lose the ability to occur independently. In my data of Amarakaeri, some instances of this process are found. For instance, in the word *ñokpo* 'sun', which clearly seems to contain shape morpheme *-kpo*, the element **ño* cannot occur independently, and seems to have become a bound root:

(128) **ño*

(129) *ño* ***-kpo***
? **-SHM:eye/light.sphere**
'sun'

The merging and semantic conventionalization of *ñokpo* is further confirmed by the fact that the combination *ñokpo-kpo* appears in the data:

(130) *ño* ***-kpo*** ***-kpo***
? **-SHM:eye/sphere** **-SHM:eye/sphere**
sun -SHM:eye/sphere
'sun (sphere)'

The occurrence of such constructions is interesting; the fact that the reduplication of a shape morpheme is not considered redundant here indicates that the N + SHM combination has completely merged. It also shows that the use of a shape morpheme (the second *-kpo*, in this example) may be optional, and that it does not necessarily contribute semantic content to the noun, but may serve to 'reconfirm' the shape of the referent (just like in combinations such as *Jessica-pi*). Interestingly, however, in some cases of merging the original shape morpheme is still recognized grammatically, as it can be repeated in agreement. An instance of this appears in Hart (1963):

(131) *O* ***-kpo*** *-harak* *ño**kpo**-ta*
3SG **-SHM:eye** -kill.3SG sun -OBJ
'He eye-kills the sun'
(Hart 1963: 5)

The reason for this may be that words like *ñokpo* are still somewhat semantically transparent: while *ño-* has lost its independent semantic value, *-kpo* can still be recognized as a shape morpheme with the meaning 'light-emitting sphere', since the entity it refers to (i.e. the sun) is indeed a sphere which emits light. An interesting question is whether semantically opaque morphemes such as *-po* in *wakuchipo* are also repeated in agreement. If this is not the case, one could argue that lexicalization and merging are fundamentally different processes, since merging does not necessarily imply loss of semantic transparency, while lexicalization does. This, however, will have to remain for future research.

6.3.5 Concluding remarks

As we have seen in this section, the shape morphemes show different degrees of semantic conventionalization. Whereas some N + SHM combinations are formed spontaneously and are semantically transparent, others have become semantically opaque. Some combinations have become so conventionalized that noun and shape morpheme have merged: the noun has become a bound root, although the shape morpheme may still bear an agreement function. It was shown that different degrees of semantic conventionalization exist even for one and the same shape morpheme, such as *-po*. The fact that each N + SHM combination has its own place on the conventionalization scale indicates that different combinations have originated at different times, and that some of them are used more frequently, and are thus more prone to semantic conventionalization than others. The high degree of lexicalization and merging of some N + SHM combinations suggests that the shape morphemes have been around for a while. It must be noted, however, that the system is still productive, since shape morphemes can still be used spontaneously for new objects. This notion will be further elaborated on in the next section.

6.4 Use of shape morphemes in neologisms

In Amarakaeri, just like is the case in many other Amazonian languages, not lexical borrowing but the creation of neologisms is the most popular way of referring to new cultural items. The majority of new items and objects, from TV to fish hook, and from calendar to mobile phone, are denoted by newly constructed Amarakaeri words rather than by loanwords. These neologisms are very interesting in the light of language change and development, since they show the current state of affairs in word formation in the language. An examination of the ways in which Amarakaeri neologisms are formed may thus provide a clearer image of both the productivity and the morphosyntactic properties of the shape morphemes.

Even though there is no central authority for the formation of neologisms in Amarakaeri, newly formed words are spread from one Amarakaeri community to the other, whenever there is contact.[40] It seems, however, that many of the neologisms are not very well established yet, since not all speakers agree on all of them. Some neologisms even seemed to be made up on the spot. In order to find neologisms, I compiled a list of Spanish words for some 60 relatively modern items, and asked my consultants to translate these into Amarakaeri. The vast majority of these words indeed had equivalents in Amarakaeri, although a few items were only denoted by Spanish loanwords (cf. 6.5). Interestingly, the majority of the recorded neologisms involve the use of shape morphemes and prefix *wa-*. This proves that the derivational use of the shape morphemes is still highly productive in the language.

The following basic morpheme structures occur in my list of neologisms:

N + N
N + SHM
wa- + SHM + SHM
wa- + V
wa- + V + N
wa- + V + SHM
wa- + SHM + V
e'- + V + SHM

Each of these will be discussed in more detail below. Apart from the words that were formed on the basis of these constructions, there were also a small number of words which I was not able to analyze morphologically, like *korärä* 'airplane' and *yudnta* 'shirt'. Some of these terms are probably older Amarakaeri words whose semantic value has been extended in order to cover new cultural items. An example of this process is the word *mbëgn*, which was formerly used to refer to the traditional stoneware bowls of the Harakmbut, but has now acquired the broader meaning of 'plate, vessel', thus referring to the shallow plates that are currently used in the community. These words, although very interesting in their own right, will not be further discussed here.

[40] Although some of the communities are quite far apart, Amarakaeri speakers regularly travel between them, either to visit relatives, to exchange goods, or to attend celebrations.

6.4.1 N + N

Just like in many of the world's languages, two nouns may be combined in order to form a new word in Amarakaeri. Many of these compounds consist of combinations in which the first element denotes a certain substance or material, and the second part refers to the item which is made of that material:

(132) *siro -koso*
metal -pan
'aluminium pan'

In these cases, thus, we see a modifier-head construction which is semantically headed by the right hand constituent.

6.4.2 N + SHM

As was already noted by Hart (1963), a frequently used combination for neologisms is that of material names with shape morphemes. This construction closely resembles the N + N construction that was discussed above. Example (133) is mentioned by Hart (1963: 1) and was confirmed by my consultants:

(133) *siro* ***-pi***
metal **-SHM:stick.like**
'(small) nail'
(Hart 1963: 1)

Apart from the construction 'material + shape', many other semantic values of N + SHM combinations occur in my data. Some instances are:

(134) *yögnki* ***-mba'***
money **-SHM:hand/leaf**
'bank note'

(135) *korärä* ***-'i***
airplane-**SHM:foot**
'wheel of airplane'

It could be argued that in each of these examples the shape morpheme designates the type of entity, while the noun modifies it in some way. However, this modifier-head relationship is not completely obvious and unambiguous in *yögnki-mba'*, which can be said to be 'paper of the money-type', but could just as well be described as 'money of the paper-type' (cf. 6.1.1.1). The data also indicates that a N + SHM combination can be further extended by adding more shape morphemes. In these cases, the first N + SHM combination practically functions as a noun which modifies the final shape morpheme:

(136) *öwey* ***-ën*** ***-pi***
tree **-SHM:flesh** **-SHM:stick.like**
wood -SHM:stick.like
'(small) wooden stick'

6.4.3 wa- + SHM + SHM

Some neologisms are formed by adding a shape morpheme to an already existing *wa-* + SHM combination:

(137) *wa* ***-'i*** ***-ot***
NMLZ **-SHM:foot** **-SHM:covering**
'shoe, boot'

Since *wa-* + SHM combinations function as nouns, this type of construction can be considered similar to the N + SHM construction. The analogy is particularly evident when comparing examples (138) and (136), which can be used interchangeably to denote the same referent:

(138) *wa* ***-ën*** ***-pi***
NMLZ-**SHM:flesh** **-SHM:stick.like**
'(small) wooden stick'

The *wa-* + SHM + SHM construction may also be extended by adding more shape morphemes. In example (139) we see that also in this type of construction the rightmost element functions as the semantic head:

(139) *wa* ***-mbo*** ***-pa*** ***-ot***
NMLZ **-SHM:leg** **-SHM:rod.shaped** **-SHM:covering**
'trousers'

Interestingly, however, this does not always seem to be the case in words that are not neologisms. As was noted by Tripp (1963: 3) and later confirmed by my consultants, in the word *wanö'kunö'* 'tongue tip', shape morpheme *nö'* is repeated at the end of the word. This second *nö'*, does not seem to be the semantic head of the word, nor does it appear to add a strong semantic value to it.

(140) *wa* ***-nö'*** ***-ku*** ***-nö'***
NMLZ **-SHM:tongue** **-SHM:head** **-SHM:tongue**
'tongue tip'
(Hart 1963: 3)

However, no other instances of this type of appear in my data.

6.4.4 wa- + V

Many neologisms are formed by combining prefix *wa-* with a verb root. This combinati
on results in a noun with the meaning 'thing for V-ing'. This observation is in line with the characterization of prefix *wa-* as a 'prospective infinitive' by Helberg (1984: 189).

(141) *wa -tai'*
NMLZ -sleep
'bed'

(142) *wa -poto*
NMLZ -shoot
'rifle'

6.4.5 wa- + V + N

When a noun is added to the *wa-* + V construction, it specifies the entity which is involved in the activity denoted by the verb. The use of nominalizer *wa-* seems to be obligatory in this type of construction.

(143) *wa -mawea' -hak*
NMLZ -cook(V) -house
'kitchen'

6.4.6 wa- + V + SHM

One of the most productive morpheme combinations in the formation of neologisms is that of prefix *wa-* plus verb, followed by a shape morpheme. Just like in the *wa-* + V + N constructions, these neologisms usually denote the item (specified by the shape morpheme) which is involved in the activity expressed by the verb:

(144) *wa -hapak* ***-po***
NMLZ -speak **-SHM:round/box.like**
'telephone, radio'

Comments by my consultants suggest that in this construction the shape morpheme mainly functions to narrow down the entity that is introduced by *wa-* and further specified by the verb. That is, the use of the shape morpheme is not obligatory, and the combination of *wa-* + V may be used to refer to the same item. Compare the following examples:

(145) *wa -chawaya'*
NMLZ -watch
'thing for watching' (e.g.: television, satellite receiver)

(146) *wa* *-chawaya'* ***-pe***
NMLZ -watch **-SHM:cheek/disk.shaped**
'satellite receiver'

6.4.7 wa- + SHM + V

Another frequently used construction for neologisms is *wa-* + SHM + V. When we compare this construction to the previously mentioned *wa-* + V + SHM combination, it becomes clear that the individual morphemes have different functions in the two constructions. Whereas *wa-* + V + SHM results in a word of the type 'SHM for V-ing', the construction *wa-* + SHM + V designates a 'thing for V-ing SHM'. That is, in the former, the shape morpheme serves to specify the entity that is introduced by *wa-*, while in the latter the shape morpheme refers to the object of the verb.

(147) *wa* ***-'idn*** *-koyo*
NMLZ **-SHM:tooth**-clean
'tooth brush'

(148) *wa* ***-sö*** *-chiwekea'*
NMLZ **-SHM:body**-dry
'towel'

Interestingly, it seems like this construction features mainly the more specific shape morphemes, while the *wa* + V + SHM construction uses mainly the shape morphemes that denote only shape. This could be explained by the fact that in the latter, the shape morpheme functions only to further specify the referent in terms of its shape, while in the former it has to fully denote the object of the verb.

6.4.8 e'- + V + SHM

In the formation of some neologisms not *wa-* is used as a nominalizer, but prefix *e'-*. The main function of this prefix in the language is that of an infinitive marker (Helberg 1984). However, in some neologisms it seems to have a function similar to that of *wa-*, carrying the meaning 'thing', rather than just nominalizing the verb. This is illustrated in example (149):

(149) *e'* *-mandoya'* ***-mba'***
INF -write **-SHM:hand/leaf**
'book'

The use of prefix *e'-* instead of *wa-*, and the differences between these two prefixes, will be further discussed in Section 6.6.

6.4.9 Concluding remarks

Considering the ways in which neologisms are created in the language, we can conclude that both prefix *wa-* and the shape morphemes are still very productive in word formation. This supports the idea that the shape morphemes have a strong derivational function. It must also be noted that some shape morphemes appear to be much more productive in this context than others. Particularly those shape morphemes that purely denote shape (*-pi*, *-po*, *-pu*, etc.) occur in many new nouns, while many other shape morphemes do not appear in the data at all. This is not very surprising, since the shape morphemes that have the most generalized semantic value are usable in more contexts, and thus in more neologisms, than the semantically more specific shape morphemes. As was demonstrated in 6.4.7, different types of shape morphemes are used in different types of constructions, in correspondence with the function that the shape morpheme serves in each construction type. It is also important to note that the use of shape morphemes in all of the recorded neologisms is semantically transparent, which indicates that any semantic opacity in the use of shape morphemes in older words is related to lexicalization and semantic conventionalization.

Furthermore, the list of possible morpheme combinations suggests that all new nouns have a basic construction that starts with either a noun or with prefix *wa-* or *e'-* (which can be said to serve as dummy nouns). No neologisms in the data start with a shape morpheme; this indicates that the shape morphemes do not have noun status when it comes to forming new nouns. Also, no new nouns appear to start with a verb. All this is illustrated with examples (150)-(152). When comparing (150) and (151), we see that the use of the shape morpheme is not obligatory in the formation of this word, and that it only serves to narrow down its semantic scope. In contrast, leaving out *wa-* in this construction would render a grammatically incorrect word (152):

(150) *wa* *-mandoya'* ***-pi***
NMLZ -write **-SHM:stick.like**
'pen'

(151) *wa* *-mandoya'*
NMLZ -write
'thing for writing (e.g.: pen, crayon)

(152) **mandoya'* ***-pi***
write **-SHM:stick.like**
Intended: 'pen'

Finally, it must be noted that only the formation of nominal neologisms was studied in this research. The creation of verbal neologisms is another interesting topic that deserves attention in future research.

6.5 Use of shape morphemes with loanwords

Although the creation of neologisms is the most common strategy to denote new cultural items in Amarakaeri, my data also shows some instances of Spanish loanwords. Interestingly, the majority of Spanish loans in my data are combined with shape morphemes. Most of these combinations are very similar to the other N + SHM combinations in the language, in which the shape morpheme functions as the semantic head and the noun serves as a specifier. This is illustrated with examples (153) and (154):

(153) *naranja* ***-kidn***
orange (Sp.) **-SHM:small.round/seed**
'orange seed'

(154) *naranja* ***-nda***
orange (Sp.) **-SHM:round/oblong.fruit**
'orange (fruit)'

Interestingly, the loan *naranja* also occurs in the data without shape morpheme *-nda*. According to my consultants both uses of the word are correct, but the difference between *naranja-nda* and *naranja* is that the former is considered Amarakaeri and the latter is Spanish. This indicates that shape morphemes can be used to 'amarakaerianize' loanwords, since loanword + SHM combinations are considered to be Amarakaeri words. Also, some Amarakaeri words have loanword equivalents, as we can see in examples (155) and (156):

(155) *watawa* ***-'i***
chicken **-SHM:foot**
'chicken foot'

(156) *pollo* ***-'i***
chicken (Sp.) **-SHM:foot**
'chicken foot'

According to the consultant, these two words have the same meaning and can be used interchangeably. It seems likely that the choice for one or the other mainly depends on extralinguistic factors.

While in most loanword + SHM combinations the shape morpheme functions as a semantic head, in some cases it seems to contribute little semantic value. This is illustrated in examples (157) and (158). In (157) the shape morpheme is the semantic head of a compound, whereas it seems to be semantically redundant in (158):

(157) *vaca* ***-ën***
cow (Sp.) **-SHM:flesh**
'beef'

(158) *tabla* ***-ën***
plank (Sp.)-**SHM:flesh**
'plank'

This is interesting, since it shows different functions of shape morphemes: a noun-like function in (157) and a more classifier-like function in (158). The latter is reminiscent of the use of shape morphemes in instances like *Jessica-pi*, where it can quite easily be left out without changing the meaning of the noun. It might be the case that shape morphemes are added to loans simply through analogy with native N + SHM combinations, and that their main function is to 'amarakaerianize' these loans.

Finally, it must be noted that none of the Spanish loanwords which occur in the data carry prefix *wa-*. This seems logical, since the loanwords in my data are nouns in Spanish, and are adopted as such into Amarakaeri, which makes the use of a nominalizing prefix redundant.

6.6 Prefixes *wa-* and *e'-*

An important element related to both the shape morphemes and word formation in Amarakaeri is prefix *wa-*, which was already mentioned in Chapter 4. The prefix seems to be obligatory in some morphosyntactic contexts and optional in others. In this section, the different functions of *wa-* will be discussed, as well as its contexts of use. Prefix *e'-*, which seems to serve a similar function in word formation, will also be mentioned here.

6.6.1 Functions and contexts of use of wa-

As we saw in Chapter 4, previous works do not agree on the function and nature of prefix *wa-*. Hart (1963) and Tripp (1995) denominate *wa-* a 'classifier prefix'. The exact meaning of this term remains unclear, as it is not explained by either of the authors, but the term suggests that the use of prefix *wa-* is related to nominal classification. According to Matteson (1972), the prefix is a definite article. Helberg (1984) proposes a more general function of *wa-*, arguing that it functions both as a nominalizer and a 'prospective infinitive'.

On the basis of my data, Helberg's analysis of the function of *wa-* seems correct. The more general function of prefix *wa-* appears to be that of a nominalizing element. It plays an important role in word formation, as it serves to derive nouns from both verbs and shape morphemes, and from all possible combinations that start with one of these elements, as was demonstrated in the section on neologisms. However, when used on verbs, *wa-* not only serves to nominalize the verb, but also adds the value of 'thing for V-ing', which corresponds with the 'prospective infinitive' function mentioned by Helberg. This is illustrated in example (159), where the combination of *wa-* with the verb for 'to sleep' means 'bed', i.e. an object that is used to perform the action which is expressed

by the verb. This shows the added semantic value of prefix *wa-* in this context, since a more semantically neutral nominalization of this verb could be expected to render the noun 'sleep'.

(159) ***wa*** *-tai*
NMLZ -sleep
'bed'

This seems to be related to the '*wa*-infinitive' that is mentioned by Helberg (1984), which is used in dependent final clauses such as in example (160):

(160) *wëy -mëy -töne' mbo' -yok -në madera -ën*
plant-PL -elder 3PL→1PL -give-AF wood(Sp.)-POS

mba -ku -ta hak ***wa*** *-ka'*
SHM:hand -SHM:head -SHM:base house **NMLZ** -make

'The trees give us their wood and leaves to build houses.'

(Helberg 1984: 455)

In contrast, when attached to shape morphemes the prefix is semantically empty. Its sole function in this context seems to be that of nominalizer, since nouns that are derived from shape morphemes only possess the semantic properties of the shape morpheme itself. This is illustrated in (161):

(161) ***wa*** *-ku*
NMLZ -SHM:head
'head'

This adds to the argument that *wa-* serves as a dummy noun when it is prefixed to shape morphemes. Its function in this context is very similar to that of the empty root elements that exist in other Amazonian languages, such as Miraña, Kwaza, Kanoê and Mundurukú (cf. Chapter 3): a semantically empty element that grants individual noun status to classifiers. Its semantic emptiness was confirmed by my language consultants. When asked for their intuitions about the meaning of *wa-*, they stated that it had no meaning at all, and that it was just an element which was used in word formation. Interestingly, when I later asked one of them to name a few words that started with *wa-*, he came up with a list of words that started with a double *wa-*, such as *wawa* 'wasp' and *wawaknda* 'very big'. This suggests that, according to speakers' intuitions, prefix *wa-* cannot be distinguished as a separate part of the noun. This may be related to the fact that many nouns start with the prefix.

The obligatoriness of prefix *wa-* in different contexts involving shape morphemes was already discussed in 6.1. It can be concluded that its use is obligatory in some constructions (e.g. when shape morphemes are used as independent nouns), ungrammatical in others (e.g. with numerals and demonstratives), and that it can also be optional (e.g. in the possessive construction). While it may be left out in some of the contexts in which it serves as a dummy noun with shape morphemes, the

use of *wa-* always seems to be obligatory in its role of verb nominalizer. This is illustrated in (162)-(164). Whereas *wa-* is left out in the construction DEM + SHM (162), it is obligatorily used in combination with nominalized verbs (163) and in *wa-* + V + SHM constructions (164):

(162) *in* *-ku*
DEM1 -SHM:head
'this head'

(163) *in* ***-wa*** *-tai*
DEM1 -**NMLZ** -sleep
'this bed'

(164) *in* ***-wa*** *-chawaya* *-po*
DEM1 -**NMLZ** -watch -SHM:round/box.like
'this television'

6.6.2 Use of e'- with the same function as wa-

As was already mentioned in the section on neologisms, prefix *e'-* is sometimes used in word formation instead of *wa-*. While it mainly functions as an infinitive marker (Helberg 1984), it is also used to derive nouns from shape morphemes, just like *wa-*. Interesting is the question how the choice for *e'-* or *wa-* is determined in these cases. In my data, the following pairs appear:

***wa**mba'*	'hand'
***e'**mba'*	'leaf'
***wa**pidn*	'bone, rib'
***e'**pidn*	'spine'
***wa**we*	'river'
***e'**we*	'rain'

A survey of words starting with *e'-* in Tripp's dictionary shows that only a handful of these are nouns; the vast majority represent verb infinitives. In Tripp (1995) I found the following pairs of nouns:

***hua**bia'og*	**'**yeast'
***e'**bia'og'*	**'**yeast'
***hua**bij*	'rope'
***e'**bij*	'liane'
***hua**borog*	'heron'
***e'**borog*	'rooster crowing'

hua*buei'*	'dead'
e'*buei'*	'dead'
hua*hui'ok*	'winter, rainy season'
e'*huiokpak*	'start of winter'
hua*itake'*	'meeting point'
e'*itaka'*	'crowd'
hua*tiai'*	'smoking'
e'*tiai*	'dry clothes'
hua*toe*	'spouse'
e'*toepak*	'wedding, marriage'

It is difficult to find a pattern in the semantics of these pairs. However, there seems to be a slight tendency for the *e'*- nouns to be more specific than their *wa*- counterparts. Furthermore, it must be noted that far more nouns carry *wa*- than *e'*- in the language, and that most of the *e'*- nouns have an equivalent that starts with *wa*-. This is confirmed by my data on neologisms, in which the vast majority of new nouns are formed with *wa*-. This indicates that prefix *wa*- may be the default nominalizer, and that *e'*- is used to form another noun, related in meaning, when *wa*- is already taken. It might be the case that *e'*- was originally just an infinitive marker, but that its use has been extended towards a more general nominalizer, through analogy with *wa*-. Interestingly, some of the nouns that are nominalized with *e'*- seem to have undergone verbalization first (166), since they carry verbalizing suffix *-pak*. In these cases, prefix *e'*- has two functions: that of infinitive marker and that of more general nominalizer (167).

(165) ***wa*** *-toe*
NMLZ -spouse
'spouse'

(166) *-toe* ***-pak***
-spouse **-VBLZ**
'marry'
(Helberg 1984: 345)

(167) ***e'*** *-toe* ***-pak***
INF -spouse **-VBLZ**
'to marry (V); wedding, marriage (N)'

6.6.3 Concluding remarks

As was shown in this section, prefix *wa-* serves as a nominalizer and prospective infinitive, in correspondence with claims by Helberg (1984). Furthermore, considering its prevalence in neologisms, it can be concluded that the prefix is highly productive in word formation. Interestingly, infinitive prefix *e-* is sometimes used to replace *wa-* in its nominalizing function.

It must be noted that, while the current functions of prefix *wa-* have been established in this section, the question as to why so many words in Amarakaeri carry the prefix, remains unanswered. An attempt at answering this question will be made in the next chapter, which will provide a closer look at the possible origins and development of prefix *wa-* and the shape morpheme system.

7. Discussion

In this chapter, all the information from previous chapters - the theoretical and typological framework, comparative data from Katukina and other Amazonian languages, previous works on Amarakaeri, and my fieldwork data - will be knit together, in order to form hypotheses about the nature, origins and development of the Amarakaeri shape morpheme system and prefix *wa-*. First of all, in 7.1, the Amarakaeri shape morpheme system will be placed into a typological framework. In 7.2, the (mixed) origins of the Amarakaeri shape morpheme system will be discussed; the possible origins of prefix *wa-* are the topic of 7.3. In 7.4, several scenarios for the development of the shape morpheme system and the role of prefix *wa-* will be discussed. Finally, in 7.5, I will briefly mention in what ways the shape morpheme system seems to indicate areal influence from other Amazonian languages.

7.1 Typological and comparative considerations

How can the Amarakaeri shape morphemes be categorized typologically? Should they be considered class markers or nouns, and if they are class markers, what type are they? As was discussed in Chapter 2, different typological frameworks have been proposed from which nominal classification can be viewed. Whereas Dixon (1986) has established two main types of systems, i.e. noun classes and noun classifiers, Aikhenvald (2000a) and Grinevald (2002) use a 'continuum approach', in which each phenomenon of nominal classification is believed to represent a certain point on a scale of grammaticalization, ranging from class terms at the lexical end to noun classes at the grammatical end. As was already noted by Derbyshire & Payne (1990), most Amazonian systems of nominal classification do not fit into either of the categories that were established by Dixon (1986), and show properties of both types. It is therefore not surprising that also the Amarakaeri shape morpheme system cannot be described by means of Dixon's categorization. Viewing the shape morphemes from a continuum perspective seems to be more fruitful, for it allows a more 'custom-made' description of the system's properties. Moreover, the continuum approach takes into account that systems of nominal classification may evolve and move along the grammaticalization in the course of time, which makes it especially suitable for the present research.

In order to categorize the shape morphemes typologically, we first need to look at what distinguishes them from nouns. A summary of the noun and class marker properties of the shape morphemes, which have been discussed in Chapter 6, is presented in Table 9 below.

TABLE 9
NOUN PROPERTIES AND CLASS MARKER PROPERTIES OF THE AMARAKAERI SHAPE MORPHEMES

Noun properties	Class marker properties
strong derivational function	well established agreement between noun and verb-incorporated shape morpheme; (proposed) incipient agreement within the NP
same distribution as nouns in certain contexts (e.g. with numerals and demonstratives)	different distribution than nouns in certain constructions (e.g. with adjectives and in the possessive construction)
narrow semantic scope of some of the shape morphemes (e.g. *-min* 'bowel')	broad semantic scope of some of the shape morphemes (e.g. *-pi* 'stick like item')
	inability to occur fully independently

As was discussed in Chapter 3, the distinction between nouns and class markers is far from clear-cut in many Amazonian languages. For instance, Weber (2002) argues that Bora class markers should be considered bound nouns, and according to Crofts (1971), class markers in Mundurukú are a subtype of noun roots. The same could be said of the Amarakaeri shape morphemes, for they display many noun properties. However, the morphemes could just as well be coined class markers, since they have agreement functions and are not full nouns. In the end, this is mainly a matter of definition; any label we put on complex phenomena such as the Amarakaeri shape morphemes will not do justice to their complexity. This is why systems of nominal classification should be viewed as systems in development, representing points along the grammaticalization continuum. In this context, the Amarakaeri shape morphemes seem to be situated somewhere close to the lexical end of the scale, because of their noun-like behaviour and their relatively little developed agreement functions. It must be noted that there are strong differences within the shape morpheme inventory, as was shown in 6.2: some of the morphemes have broad semantic scopes and denote only shape, whereas others have a very specific meaning. This indicates that the inventory is not uniform, and that different shape morphemes may, in fact, represent different points along the grammaticalization continuum. This corresponds with Aikhenvald (2000: 375), who states that different stages of grammaticalization may coexist in the nominal classification system of a single language.

As for the type of classifiers, the shape morphemes are most similar to verbal classifiers, as was already noted by Derbyshire & Payne (1990). The main motivation for this characterization is that most of the agreement takes places between verb-incorporated shape morphemes and corresponding nouns. However, since some agreement at the NP level was also recorded in my data, it could be argued that the shape morphemes are developing other functions as well. In this sense, Amarakaeri could be considered a 'multiple classifier language', a term which Aikhenvald (2000: 204) explains

as a language in which a single set of classifiers appears in different morphosyntactic environments. In my view, however, this would be too strong a characterization for the Amarakaeri shape morpheme system, in which the vast majority of agreement still seems to take place with verb-incorporated morphemes. As was shown in 6.1, the shape morphemes may occur in many different morphosyntactic loci (e.g. with demonstratives, adjectives, interrogatives, on nouns, etc.). It is essential to note, however, that the shape morphemes do not have an agreement function in most of these contexts, and that they practically function as nouns in many cases. The occurrence of shape morphemes in many morphosyntactic environments is therefore not an argument for categorizing the language as a multiple classifier language; rather, it stresses the noun-like behaviour of the shape morphemes. All in all, the categorization of the shape morpheme system by Derbyshire & Payne (1990: 260), who describe it as a 'primarily verb incorporated system', seems to be the most appropriate one.

It must also be noted, once more, that the Amarakaeri shape morphemes play an important role in word formation. In my data, particularly the strong presence of the shape morphemes in neologisms is striking. It indicates that speakers of Amarakaeri often recur to the use of shape morphemes when confronted with new objects, and that there is a strong tendency to describe objects in terms of shape. Comparing the Amarakaeri system with systems from other Amazonian languages, the strong derivational function of the shape morphemes seems to fall into place. As was discussed in Chapter 3, it is quite common for Amazonian languages to have a system of nominal classification that combines agreement and derivation; such a combination is considered one of the typical characteristics of Amazonian systems by Grinevald & Seifart (2004), and in Chapter 3 of this thesis it was shown to exist in various Amazonian languages, such as Miraña and Mundurukú.

Apart from their important role in word formation, there are more aspects on which the Amarakaeri shape morphemes fit in with 'typical' Amazonian classifier systems. These include the semantic transparency of the system, and its shape-based semantics. The existence of an empty root construction, used to grant independent noun status to classifiers, is mainly found in a more specific sub-region of the Amazon basin: the Guaporé-Mamoré linguistic area. According to Crevels & Van der Voort (2008), a similar construction exists in the languages Kwaza, Kanoê, Cavineña, Latundê and Ese Ejja. These belong to different linguistic families, but are all believed to form part of the Guaporé-Mamoré region. The occurrence of the empty root construction in these languages is even considered to be one of the defining characteristics of the linguistic area (Crevels & Van der Voort 2008: 168). This interesting correspondence will be further elaborated on in 7.3.

The relative age of the system is difficult to establish. However, according to Aikhenvald (2000a), semantic transparency and clear lexical origins of class markers are often found in younger systems. Indeed, the use of the Amarakaeri shape morphemes is semantically transparent (although N + SHM combinations may become lexicalized), and many of the shape morphemes seem to have been derived from nouns. In comparison with their noun cognates in Katukina, the morphemes do not seem to have undergone extensive phonological simplification; this is another indicator that the system is relatively young. As for the system's dynamism, it was shown in this thesis that the shape morphemes are still productive, both in derivation and in agreement: they are extensively used in the formation of neologisms, are still being incorporated into verbs, and they even seem to be developing more agreement functions. Moreover, the semantic generalization of the shape

morphemes does not seem to have come to a halt, as neologisms display some of the most semantically generalized uses of the shape morphemes. This suggests that the shape morphemes may still be undergoing a process of grammaticalization, slowly moving towards the grammatical end of the continuum, although they are still situated at the more lexical half of the scale.

7.2 Possible origins of the shape morphemes

Where do the shape morphemes come from, and how did they develop? As was mentioned in Chapter 2, classifiers are often derived from nouns (Mithun 1986, Aikhenvald 2000a). Typical sources of classifiers are generic nouns, unit counters, kinship terms, and body part nouns. The latter often develop into verbal classifiers through classificatory noun incorporation (Mithun 1986), as seems to have happened in Amazonian languages like Mundurukú, Chayahuita, Waorani and Parecis (Derbyshire & Payne 1990). Considering the fact that the Amarakaeri shape morphemes are primarily verbal classifiers, development through classificatory noun incorporation seems to be the most likely of the grammaticalization paths that were discussed in Chapter 2. The other two ways in which lexemes tend to develop into class markers - through compounding or repeaters - seem less probable. As was shown in 2.3, grammaticalization through compounding usually involves generic-specific pairings (such as the combination 'animal wallaby' in the Australian language Yidiny), which gradually evolve into CL + N combinations. However, I found no indication that such pairings ever existed in Amarakaeri. The other grammaticalization path, which leads through repeaters, tends to involve a stage in which repeaters exist alongside their independent noun counterparts, often in truncated form (such as is the case in Miraña). Since the Amarakaeri shape morphemes do not appear as independent nouns in any context, we would have to assume that the system has already passed this intermediary stage, and that the repeaters have already become class markers, if we are to adopt the repeater scenario. Again, however, there are no indications that we should do so, whereas there are several reasons to believe that the system has taken the path of noun incorporation.[41]

The origins of the shape morphemes seem to be mixed. As was shown by Adelaar (2000, 2007), several of the shape morphemes appear to have full noun cognates in Katukina. I would argue that this suggests that the shape morphemes were independent nouns in the proto-language, and that they have grammaticalized in Amarakaeri. The final stage of the process of classificatory noun incorporation, as Mithun (1986) describes, is when verb-incorporated classifiers come to indicate qualities instead of entities. As we saw in 6.2, this seems to be the case for some of the shape morphemes, such as *mba'-*, 'hand', which may simply mean 'flat and flexible' in some contexts, or *-kidn* 'seed', which is sometimes used for 'small and round' objects. The fact that not all shape morphemes have reached this stage, many of them displaying very specific semantics, indicates that the process of grammaticalization has not been 'completed' yet.

However, for some of the shape morphemes no nominal origins can be found. Certain morphemes, such as *-pi* 'stick like' and *-pu* 'tube shaped' do not seem to have been derived from body part

[41] This will be discussed in more detail in 7.4.

nouns. Interestingly, these shape morphemes, which form a small subset of the inventory, have the broadest semantic scopes, are phonologically the most simple ones, and are more frequently used in neologisms than other shape morphemes. It could be argued that these morphemes have become so grammaticalized that they have undergone extensive semantic bleaching and phonological simplification, since these processes often occur simultaneously. This could be the reason why no nominal origins can be traced for these shape morphemes. However, another - and perhaps more likely - possibility is that some of the shape morphemes have been borrowed from neighbouring languages. As was mentioned in Chapter 3, shape morphemes *-nda*, *-pa*, *-pi* and *-pu* are also found in the Arawakan language Tariana, being class markers that are almost identical, both in form and in semantic scope, to the Amarakaeri shape morphemes. This strongly supports the idea that these morphemes have entered Amarakaeri through direct diffusion. However, the geographical distance between Amarakaeri and Tariana, which is spoken in the Vaupés area of Northern Brazil, is rather large. One would thus expect to find the same morphemes in Arawakan languages that lie in between, such as Yine (also known as Piro), Iñapari and Machiguenga. Indeed, Parker (1995: 189) mentions *-pi* as a very productive nominal classifier in Iñapari, which is suffixed to nouns referring to objects that are long, thin and round, like a snake. Examples include the words for 'nail', 'spine', 'rib', and different types of snakes. Most strikingly, the word for 'finger' is formed in a similar way as in Amarakaeri:

(168) *muyú* ***-pi*** *-ti*
hand **-CL:long.thin-**NPOS[42]
'finger'

However, the only other nominal classifier that Parker mentions is *-ri*, which appears on nouns that refer to handmade objects. The other classifiers that Amarakaeri and Tariana have in common, (*-nda*, *-pa* and *-pu*), are not reported to exist in Iñapari, nor does an examination of a word list of the language indicate their existence. The same goes for Yine and Machiguenga. While a suffix *-pi* is not explicitly mentioned in the grammar notes that accompany the Yine dictionary by Nies (1986), several words for long, thin items in the word list seem to contain it. An example is the word for pencil, *yonawapi*: since *yonawa* means 'to write', one could assume that the word was formed in the same fashion as Amarakaeri *wamandoyapi* 'pen'. In the dictionary of Machiguenga (Snell 1998), fewer occurrences of *-pi* can be found, but there are indications that it does exist in the language, e.g. in the word for 'arrow', *chakopi*. It must be noted, however, that *-pi* for long and thin objects is widespread throughout the Amazon basin, as was already mentioned in Chapter 3. It might stem from the Proto-Arawakan noun **api* 'snake' (Aikhenvald & Dixon 2001), and is also reported to exist in Newiki, Apuriná, and Chapakuran languages. Therefore, the existence of *-pi* in Tariana, Iñapari, Yine and Machiguenga is less surprising than the correspondences between other classifiers in Tariana and shape morphemes in Amarakaeri. More detailed studies are needed in order to discover how these correspondences have come into being.

For now, we can conclude that it is very likely that shape morpheme *-pi* has entered Amarakaeri through direct diffusion from neighbouring languages, and that it is well possible that the same has happened with some of the other shape morphemes. The Amarakaeri shape morpheme inventory thus seems to have mixed origins: many of the morphemes appear to have been derived from full

[42] The use of non-possessive marker *-ti* in Iñapari will be further discussed in 7.3.

nouns, while others may have been borrowed from other languages. This could serve to - partially - explain the very different degrees of semantic generalization that the shape morphemes display.

7.3 Possible origins of prefix *wa-*

The strong presence of prefix *wa-* is, without a doubt, one of the most particular and fascinating characteristics of the Amarakaeri language. The Amarakaeri vocabulary can be divided into two main groups of nouns: those that do, and those that do not carry prefix *wa-*. In Tripp's (1995 dictionary, prefix *wa-* is considered part of the noun, resulting in a very long section of words starting with the letters HU (the prefix is spelt *hua-* in Tripp's orthography. What is the reason why so many Amarakaeri words carry this prefix? Several possible explanations will be discussed here.

In Chapter 6, it was established that *wa-* is highly productive in word formation, functioning both as a nominalizer and as a prospective infinitive. This may partially explain the issue. As was demonstrated in 6.4, neologisms in Amarakaeri are often composed of morpheme combinations of which the first part is either a verb or a shape morpheme. As all these combinations require the use of prefix *wa-* in order to form full nouns, the result is a large amount of new words in the language that display '*wa*-initial' constructions such as *wa-* + V + N or *wa-* + SHM + V. It is not unlikely that this type of word formation has existed in Amarakaeri for a long time, filling the vocabulary with a large amount of *wa*-words. This idea corresponds with Hart (1995), who distinguishes *common*, *proper* and *vocative* nouns from *composed*, *complex* and *derived* nouns, stating that the latter three types are usually formed with prefix *wa-*. Indeed, most of the *wa-* words that occur in Tripp's dictionary seem to be composed or derived. However, some of the nouns which carry *wa-* are in fact very short, phonologically simple words that can hardly be analyzed any further than to *wa-* + N combinations, such as *waka'* 'tool' and *wadu* 'worm'. It therefore seems too simplistic to say that noun derivation alone is the cause of the strong presence of prefix *wa-* in the language.

Another possible explanation for the prevalence of *wa-* in Amarakaeri vocabulary is provided by Gray (1996), who argues that the *wa*-nouns tend to be generic, while other nouns refer to individual species and entities (Gray 1996: 5). Gray does not further explain or exemplify this idea. A look at Tripp's (1995) dictionary indicates that, while it is true that most names of species do not carry the prefix, the distinction is not very clear-cut. Moreover, Gray seems to ignore the fact that many words that carry *wa-* are composed or derived nouns, in which the prefix functions as a nominalizer.

A different explanation can be formed on the basis of comparison with some of the other Amazonian languages that were discussed in Chapter 3. In languages such as Kwaza, Kanoê and Mundurukú, as we saw, a distinction seems to exist - or have existed - between alienably and inalienably possessed nouns (Van der Voort 2004, Bacelar 2004, Crofts 1971). In Kanoê and Kwaza, this distinction is reflected in the use of an empty root construction, in which a semantically empty prefix is used to create non-possessed versions of inalienably possessed nouns. This has resulted in a group of '*i*-nouns' in Kanoê (Bacelar 2004) and '*e*-nouns' in Kwaza (Van der Voort 2004), which mainly include body part terms and other inalienable items. The construction is illustrated in example (169), from Kanoê:

(169) ***i*** *-taw*
ER -tongue
'tongue'

Similar subtypes of nouns are reported to exist in Latundê, Ese Eja and Cavineña (Crevels & Van der Voort 2008: 168). In Kwaza, the distinction between alienability and inalienability is not very clear-cut, and some alienable items may also carry the empty root. However, Van der Voort (2004: 136) argues it is well possible that the distinction used to exist in the language, but that it has become blurred in the course of time.

As was mentioned in Chapter 5, the alienability distinction also exists in Katukina-Kanamari, which seems to be genetically related to the Harakmbut languages. This certainly makes a case for the existence of such a distinction in the proto-language. However, no semantically empty root appears to exist in Katukina or Katawixi. An alienability distinction is also reported to exist in the Arawakan languages that are adjacent to Amarakaeri. Although these languages do not attach inalienably possessed nouns to a semantically empty root, they do use a suffix that grants independent (i.e. non possessed) status to these nouns. Iñapari, for instance, has a non-possessive suffix *-ti*, which must be used on inalienably possessed nouns when they appear without possessor (Parker 1995):

(170) *no -iñáre*
1SG-cheek
'my cheek'
(Parker 1995: 179)

(171) *iñáre* ***-ti***
cheek -**NPOS**
'cheek'
(Parker 1995: 179)

The suffix is mainly used with body part nouns and kinship terms, but also appears with some other nouns that refer to inalienable things, such as the words for 'language', 'name' and 'house'. A similar suffix exists in Machiguenga. Snell (1998) mentions that 'unspecified possessor suffix' *-tsi*/*-ntsi* is used on body part nouns and other inalienable items when the noun appears without possessor prefix. Interestingly, the same suffix, or a suffix with exactly the same form, is also used as a verb nominalizer in the language (Snell 1998):

(172) *gito* ***-tsi***
head-**NPOS**
'head'
(Snell 1998: 25)

(173) *tima* ***-tsi***
live -**NPOS**
'he who lives'
(Snell 1998: 30)

This is worth mentioning, since prefix *wa-* in Amarakaeri shows the same combination of functions: an affix that grants independent status to body part nouns, and a (verb) nominalizer.

All in all, it seems safe to say that a distinction between alienably and inalienably possessed items, and the use of a certain affix on inalienable nouns when occurring without possessor, is an areal trait. It is therefore interesting to consider whether the same distinction could lie at the base of the distinction between *wa*-nouns and independent nouns in Amarakaeri. As we have seen, most body part nouns in the language carry the prefix, since they are usually *wa-* + SHM combinations. Moreover, most kinship terms (when not in vocative) and the word for 'name' start with *wa-*. This indicates that inalienability could indeed be the original reason for the use of *wa-*. In this context, the formation of the possessive construction, which was mentioned in Chapter 6, should be kept in mind. If *wa-* is indeed used as an empty root or a non-possessive affix, such as is the case with *-ti* in Iñapari and *-tsi*/*-ntsi* in Machiguenga, one would expect it to be left out in the possessive construction, as is the case in (174). Interestingly, however, the use of *wa-* in this construction is also judged grammatical, as is shown in (175):

(174) *ndo -edn* ***-ku***
1SG -POS -**SHM:head**
'my head'

(175) *ndo -edn wa* ***-ku***
1SG -POS NMLZ -**SHM:head**
'my head'

The fact that both constructions are considered grammatically correct, might indicate that a shift is taking place, in which the alienability distinction is losing ground, and *wa-* + SHM combinations are becoming increasingly lexicalized. This, however, is a rather tentative explanation.

Considering the areal presence of alienability distinctions, and the nature (i.e. body part, kinship term) of most of the *wa*-nouns that do not seem to be derived, it seems well possible that *wa-* was originally a non-possessive prefix or semantically empty root for inalienable items. Possibly, the prefix has extended its function over time, becoming a nominalizer, just like seems to have happened in Machiguenga. Consequently, in its use as nominalizer, it has been employed to form many more words, resulting in a large amount of *wa-* nouns in the Amarakaeri vocabulary. It must be noted, however, that some of the *wa*-nouns that seem to be too short to be composed or derived, are not necessarily inalienable, such as the already mentioned forms *waka'* 'tool' and *wadu* worm. These forms are difficult to explain. Another complicating issue is the fact that some nouns can be used both with and without *wa-*, such as (*wa-*)*dagn* 'path', although examples of this phenomenon are not very numerous in Tripp (1995). It might be the case that *wa-* has lost its value as a non-possessive prefix at some point, causing these 'inconsistencies' in the system, although this explanation is rather speculative. The same may have happened in Kwaza, as is argued by Van der Voort (2004: 136). Finally, the possibility that two distinct prefixes *wa-* exist in the language - a non-possessive marker and a nominalizer - should also be kept in mind.

Although *wa-* may be the reflection of an areal feature, it does not correspond in form with similar elements in any of the previously mentioned languages. This is not surprising, since Amarakaeri is not genetically related to any of these languages. In Katukina-Kanamari, the only plausible relative of the Harakmbut family (Adelaar 2000), no similar prefix can be found. However, as we saw in Chapter 5, Katukina has a generic relational noun *wa* which serves as a dummy element in the possessive construction, functioning as a grammatical possessum in combination with alienably possessed nouns (Dos Anjos 2011). Although it must be noted that this function is different from its presumed function in Amarakaeri (replacing the possessum in Katukina, as opposed to replacing the possessor in Amarakaeri), the fact that a dummy noun *wa* exists in Katukina is interesting in this context. Furthermore, as is noted by Adelaar (2000: 222), the word for 'name' in Katukina is *wadik*, which is cognate to Amarakaeri *wandik* 'name'. This is an interesting fact, since it suggests that prefix *wa-* may have existed in the proto-language and later disappeared from Katukina, having left a trace in the word *wadik*. An argument for this theory is the clear distinction between alienable and inalienable nouns that exists in Katukina; it is therefore not unlikely that the language used to have an empty root or non-possessive element to form generic nouns with inalienable roots, just like many languages in the area. One could wonder whether, instead of having disappeared from the language, the prefix may have developed into something else in Katukina, such as the already mentioned generic relational noun *wa*. However, this will be difficult to investigate.

7.4 Scenarios for the development of the shape morpheme system

As was discussed in Chapter 3, several Amazonian languages (e.g. Kwaza, Kanoê, Mundurukú) show a certain cluster of properties: a distinction between alienable and inalienable nouns, an empty root construction, classificatory noun incorporation, and a relatively large set of shape-based class markers. It can be argued that these characteristics are related to one another: body part nouns are inalienably possessed by definition; they therefore tend to occur with a semantically empty noun formative root when used independently. In many Amazonian languages, inalienably possessed nouns are the only nouns that can be incorporated into verbs (Dixon & Aikhenvald 1999), and body part nouns are some of the most frequently verb-incorporated items among these (Mithun 1986), often developing into shape-based class markers.

The shape morpheme system in Amarakaeri may be a product of a similar combination of traits: it displays classificatory noun incorporation and a large number of shape-based class markers; moreover, the presence of prefix *wa-*, which is used as a semantically empty noun-formative root with the shape morphemes, suggests that an alienability distinction may have existed in the language. The most probable scenario for the development of the system, therefore, seems to be as follows:

- The proto-language displayed a distinction between alienably and inalienably possessed nouns; the latter had to carry prefix *wa-* in order to function as non-possessed nouns.
- In Amarakaeri, some of these inalienably possessed nouns, mainly those that denoted body parts, were frequently incorporated into verbs (as is often the case with body part nouns).

- Some of these V + N compounds became so institutionalized that the incorporated nouns lost their functions as referent tracking devices.
- Corresponding full nouns had to be used in order to fulfil this referential task, thereby giving rise to agreement.
- The frequently incorporated body part nouns grammaticalized into shape morphemes: they developed agreement functions and underwent semantic generalization, becoming applicable to a wider range of referents (e.g. 'cheek' → 'flat, round object').
- However, while developing agreement functions, the shape morphemes retained many of their noun properties. They still have a strong derivational function and behave like nouns in many contexts, just like the class markers of many other Amazonian languages.
- Because of their important role in word formation, and frequent appearance in N + SHM combinations, the shape morphemes also undergo processes such as lexicalization and merging.
- At some point, the alienability distinction in the language became weaker, and prefix *wa-* developed into a nominalizer and prospective infinitive.

However, while it seems likely that the shape morphemes were full nouns in the proto-language, the possibility that these morphemes already had classifier properties in Proto-Harakmbut-Katukina should not be excluded. In this scenario, the shape morphemes would have been upgraded to full nouns in Katukina, while remaining classifier-like morphemes in Harakmbut languages. Although grammaticalization is usually unidirectional (Heine & Kuteva 2009), systems of nominal classification may be reduced or even lost under influence of neighbouring languages (Aikhenvald 2000a). According to this scenario, the presumed semantic generalization of the shape morphemes in Amarakaeri should instead be viewed as semantic specialization of their noun cognates in Katukina. Adelaar (2007) adopts this view, stating that Katukina is, in many respects, more innovative than Amarakaeri, which would make semantic change more likely to occur in Katukina than in Harakmbut (Adelaar, p.c.).

Moreover, it is possible that there was no alienability distinction in the proto-language, and that the shape morphemes were full nouns in the proto-language and remained to be full nouns in Katukina. In Amarakaeri, then, some nouns could have lost their full noun status in the process of grammaticalization. Having become bound roots, the shape morphemes then needed a nominalizing prefix in order to occur independently: prefix *wa-*. According to this scenario, *wa-* has simply always been a nominalizer in Amarakaeri, but its existence or absence in the proto-language cannot be traced back. What makes this scenario possible is the fact that *wa-* currently functions as a nominalizer, and that there is no clear-cut distinction between alienable and inalienable nouns in Amarakaeri vocabulary. However, this scenario does not provide an explanation for the presence of the noun *wadik* in Katukina, nor does it explain why short, non-composed words, such as many kinship terms, carry prefix *wa-*. All in all, the first scenario, according to which *wa-* was originally a non-possessive marker for inalienably possessed nouns, seems more solid, for it seems to provide a more sound explanation for the strong presence of prefix *wa-* throughout the Amarakaeri vocabulary.

7.5 Signs of areal influence on the Amarakaeri shape morpheme system

Finally, some attention should be paid to the implications of the present research in the context of areal and historical linguistics in Amazonia. As was shown in this thesis, and also noted by Adelaar (2000: 223), the Amarakaeri shape morpheme system corresponds, in many respects, with the classifier systems in languages such as Kwaza, Mundurukú and Kanoê, all of which belong to the Guaporé-Mamoré linguistic area. Interestingly, Pozzi-Escot (1998: 93) also notes grammatical correspondences between Harakmbut languages and Ese Eja, which is one of the languages from the Guaporé-Mamoré area[43]. This points towards areal influence from these languages, which are mainly spoken in the Brazilian state Rondônia and the Bolivian department Beni. It would therefore be interesting to investigate whether there are more correspondences between Amarakaeri and the languages of this linguistic area.

It must be noted, however, that no instances of *direct* diffusion of classifier forms from these languages have been found in Amarakaeri. In contrast, some highly interesting correspondences have been discovered between the Amarakaeri shape morpheme inventory and a subset of classifiers in the Arawakan language Tariana. The fact that some of these are almost identical in both languages, both in form and in meaning, is remarkable, considering the geographical distance between the two languages. An examination of Arawakan languages such as Iñapari, Piro/Yine and Machiguenga, which are spoken closer to the Amarakaeri territory, suggests that at least classifier *-pi*, has a similar function in these languages. More extensive comparative research would be needed in order to investigate how exactly the classifier correspondences between Amarakaeri and Tariana originated. This falls beyond the scope of the present research, but it deserves to be explored in future studies. A comparative study of Amarakaeri and Tariana may tell us more about language contact and migration patterns in the Amazon basin; a topic about which very little is known to date.

[43] The correspondences that Pozzi-Escot mentions are the existence of an infinitive prefix *e-*/*e-'*, and the use of an adjective marker *-nda*.

Conclusion

The main goal of this research was to learn more about the nature, origins and development of the Amarakaeri shape morpheme system. In terms of its nature, I have concluded that the system seems to be a primarily verb incorporated system with relatively noun-like class markers. As expected, the shape morphemes do not fit into traditional categories such as 'noun classes' or 'noun classifiers', and should be viewed from the perspective of a grammaticalization continuum. Within this framework, it seems that the Amarakaeri shape morpheme system is situated somewhere near the lexical end of the continuum, since the shape morphemes resemble nouns in many respects, and agreement is relatively limited in the system.

My fieldwork data has shed some light on several features of the system. An examination of the semantic scope of the shape morphemes has shown that the shape morpheme inventory is internally diverse and that it may have mixed origins. I have also shown how shape morphemes undergo processes such as lexicalization and merging, due to their derivational function. Data from neologisms has confirmed that both the shape morphemes and prefix *wa-* are highly productive in word formation, and that the use of shape morphemes in neologisms is semantically transparent. Furthermore, I have shown that the shape morphemes are used to 'amarakaerianize' Spanish loan words in the language.

Placing the system into its areal context has provided new insights into the origins and development of the shape morphemes and prefix *wa-*. Several other Amazonian languages display the same cluster of features that is found in Amarakaeri: classificatory noun incorporation, an empty root construction, incorporation of body part nouns, and a large set of shape-based class markers. Since most of these languages make a distinction between alienably and inalienably possessed nouns, it is not unlikely that this distinction has also existed in Amarakaeri, and that prefix *wa-* was originally an empty root that granted a non-possessed status to inalienably possessed nouns. Furthermore, based on existing theories, cross-linguistic comparison and my own data, I have concluded that the shape morphemes have probably grammaticalized through noun incorporation. As expected, however, no solid conclusions can be drawn about the origins and development of the system. More data, also from Katukina, would be needed in order to further investigate these issues.

As is often the case with research, the present study has raised many new questions, all of which deserve to be explored in future studies. With regard to the Amarakaeri shape morpheme system, more extensive research is needed in order to investigate its discourse-pragmatic functions. Also, the incipient agreement at the NP level, which was recorded in my data, should be further investigated, for it may provide new insights into the current development of the system. With regard to language contact and areal influence, possible links between Harakmbut languages and the Guaporé-Mamoré linguistic area, and the discovered correspondences between Amarakaeri and Tariana, deserve to be further explored. All these topics will have to remain for future research.

References

Adelaar, W.F.H. (2000): 'Propuesta de un nuevo vínculo genético entre dos grupos linguísticos indígenas de la Amazonía occidental: Harakmbut y Katukina', in Miranda Esquerre, L. (ed.) *Actas del I Congreso de Lenguas Indígenas de Sudamérica* 2: 219-236. Lima: Universidad Ricardo Palma.

Adelaar, W.F.H. (2007): 'Ensayo de clasificación del katawixí dentro del conjunto harakmbut-katukina', in Romero Figueroa, A., Fernández Garay, A. and Corbera Mori, A. (eds.), *Lenguas indígenas de América del Sur: Estudios descriptivo-tipológicos y sus contribuciones para la linguística teórica*: 159-169. Caracas: Universidad Católica Andrés Bello.

Aikhenvald, A.Y. (2000a): *Classifiers: a typology of noun categorization devices*. New York: Oxford University Press.

Aikhenvald, A.Y. (2000b): 'Unusual classifiers in Tariana', in Senft, G. (ed.) *Systems of nominal classification.*

Aikhenvald, A.Y. (2002): *Language contact in Amazonia*. New York: Oxford University Press.

Aikhenvald, A.Y. and Dixon, R.M.W. (2001): *Areal diffusion and genetic inheritance*. New York: Oxford University Press.

Aikhenvald, A.Y. and Green, D. (1998): 'Palikur and the typology of classifiers', in *Anthropological Linguistics* 40(3): 429-480.

Allan, K. (1977): 'Classifiers', in *Language* 53: 285-311.

Bacelar, L. (2004): *Gramática da lingua Kanoê.* Doctoral dissertation, University of Nijmegen.

Barnes, J. (1990): 'Classifiers in Tuyuca', in Payne, D.L. (ed.) *Amazonian linguistics: studies in lowland South American languages*: 273-92. Austin: University of Texas Press.

Barriales, J. and Torralba, A. (1970): *Los Mashcos*. Lima: Santiago Valverde.

Booij, G. (2005): *The grammar of words. An introduction to morphology.* Norfolk / New York: Oxford University Press.

Crevels, E.I. and Van der Voort, H.G.A. (2008): 'The Guaporé-Mamoré region as a linguistic area', in Muysken, P. (ed.), *From linguistic areas to areal linguistics*: 151–179.

Crofts, M. (1971): *Repeated morphs in Mundurukú.* Summer Institute of Linguistics. <http://www.sil.org/americas/brasil/publcns/ling/MUMorphs.pdf>

Crofts, M. (1973): *Gramática mundurukú.* Série Linguística 2. Summer Institute of Linguistics. <http://www.sil.org/americas/brasil/publcns/dictgram/MUGram.pdf>

Derbyshire, D. C. and Payne, D.L. (1990): 'Noun classification systems of Amazonian languages', in Payne, D.L. (ed.), *Amazonian linguistics: studies in lowland South American languages,* 243-71. Austin: University of Texas Press.

DINEIBIR-DEIB (2006): *Resolución directorial.* <http://www2.minedu.gob.pe/digeibir/lenguas/2006-09-08-Lengua-HARAKBUT.pdf>

Dixon, R.M.W. (1986): 'Noun classes and noun classification in typological perspective', in Craig, C. (ed.) *Noun classes and categorization* 105-112. Philadelphia: John Benjamins.

Dixon, R.M.W. and Aikhenvald, A.Y. (1999): *The Amazonian Languages.* Cambridge: Cambridge University Press.

Dos Anjos, Z. (2011): *Fonologia e gramática Katukina-Kanamari*. LOT Dissertation series 281.

Fabre, A. (2005): *Diccionario etnolinguístico y guía bibliográfica de los pueblos indígenas sudamericanos*. <http://butler.cc.tut.fi/~fabre/BookInternetVersio/Dic=Harakmbet.pdf>

Gray, A. (1996): *The Arakmbut: mythology, spirituality and history in an Amazonian community*. Providence/Oxford: Berghahn.

Greenberg, J. (1987): *Language in the Americas*. Stanford: Stanford University Press.

Grinevald, C. (2000): 'A morphosyntactic typology of classifiers', in Senft, G. (ed.) *Systems of nominal classification:* 11-49. Cambridge: Cambridge University Press.

Grinevald, C. (2002): 'Making sense of nominal classification systems: noun classifiers and the grammaticalization variable', in Diewald, G. and Wischer, I. (eds.) *New reflections on grammaticalization:* 259-275. Amsterdam/Philadelphia: John Benjamins.

Grinevald, C. and Seifart, F. (2004): 'Noun classes in African and Amazonian languages: towards a comparison', in *Linguistic Typology* 8(2): 243-285.

Groth, C. (1985): 'Syntax of the phrase types in canamarí', in Fortune, D.L. (ed.), *Porto Velho Workpapers* 93-129. Brasilia: Summer Institute of Linguistics.

Hart, .E. (1963): 'Semantic components of shape in Amarakaeri grammar', in *Anthropological Linguistics* 5(9): 1-7.

Heine, B. and Kuteva, T. (2009): 'The genesis of grammar: on combining nouns', in Botha, R. and De Swart, H. (eds.) *Language evolution: the view from restricted linguistic system:* 139-177.

Helberg Chávez, H.A. (1984): *Skizze einer Grammatik des Amarakaeri*. Doctoral dissertation, University of Tubingen.

Instituto Nacional de Estadística e Informática (2008): *II Censo de comunidades indígena de la Amazonía Peruana.* <www1.inei.gob.pe/biblioineipub/bancopub/Est/Lib0860/tomoI.pdf>

Lakoff, G. (1987): *Women, fire, and dangerous things: what categories reveal about the mind.* Chicago: University of Chicago Press

Matteson, E., Wheeler, A., Jackson, F.L., Waltz, N.E. and Christian, D.R. (1972): *Comparative studies in Amerindian languages*. The Hague / Paris: Mouton.

Mithun, M. (1986): 'The convergence of noun classification systems', in Craig, C. (ed.) *Noun classesand categorization*: 379-397. Philadelphia: John Benjamins.

Moqui, E. (1974): *Naturaleza y vida social No. 2: Manual para los cursos de Naturaleza y Vida Social y Praáctica del Castellano*. SIL publication <http://www.sil.org/americas/peru/show_work.asp?id=20654>

Moseley, C. (ed.) (2010): *Atlas of the world's languages in danger*. Paris: UNESCO. http://www.unesco.org/culture/en/endangeredlanguages/atlas.

Nies, J. (1986): *Diccionario Piro (Tokanchi gikshijikowaka-steno)*. Serie Lingüística Peruana 22. Yarinacocha: Ministerio de Educación and Instituto Lingüístico de Verano.

Parker, S.G. (1995): *Datos de la lengua iñapari.* Documento de Trabajo 27. Yarinacocha: Ministerio de Educación and Instituto Lingüístico de Verano.

Payne, D. L. (1987): 'Noun classification in the Western Amazon', in *Language Sciences* 9(1): 21-44.

Plaster, K. and Polinsky, M. (2007): 'Women are not dangerous things. Gender and categorization'. *Harvard Working Papers in Linguistics 12*. http://scholar.iq.harvard.edu/files/scholar/uploads/11/Dyirbal.pdf

Pozzi-Escot, I. (1998): *El multilingüismo en el Perú*. Cuzco: CBC.

Queixalós, F. (2004): 'Split transitivity and coreference in Katukina', in Queixalós, F. (ed.) *Ergativity in Amazonia III*: 175–188. Paris: CELIA-CNRS.

Queixalós, F. (2006), 'Incorporation nominale en sikuani et en katukina-kanamari', in *Amerindia* 31: 61-86.

Queixalós, F. and Dos Anjos, Z.G.S. (2007): 'A língua Katukina-Kanamari', in *Liames* 6: 29-60.

Seifart, F. (2005): *The structure and use of shape-based noun classes in Miraña* (*North West Amazon)*. Nijmegen: Max Planck Institute for Psycholinguistics.

Seifart, F. (2007): 'The prehistory of nominal classification in Witotoan languages', in *IJAL* 73(4): 411-445.

Seifart, F. (2009): 'Multidimensional typology and Miraña class markers', in Epps, P. and Arkhipov, A. (eds.) *New challenges in typology: transcending the borders and refining the distinctions* 365-388. Berlin/New York: Mouton de Gruyter.

Senft, G. (2000): 'What do we really know about nominal classification systems?', in Senft, G. (ed.) *Systems of nominal classification:* 11-49. Cambridge: Cambridge University Press.

Snell, B. A. (1998): *Pequeño diccionario machiguenga-castellano.* Documento de Trabajo 32. Lima: Instituto Lingüístico de Verano

Tripp, R. (1976): 'Los verbos Amarakaeri', in *Datos etno-linguísticos: colección de los archivos del ILV* (33). Lima: Instituto Linguístico de Verano.

Tripp, R. (1995): *Diccionario amarakaeri-castellano.* Yarinacocha: Ministerio de Educación and Instituto Linguístico de Verano.

Urteaga Crovetto, P. (2007): 'The social construction of ethnic groups and indigenous peoples in the Southeastern Peruvian Amazonia', in *Anales* 9-10: 73-106.

Voort, H. van der (2004): *A grammar of Kwaza*. Berlin: Mouton de Gruyter.

Weber, D. (2002): 'The structural status of Bora classifiers' *Work Papers of the Summer Institute of Linguistics*, North Dakota Session 46. http://www.und.edu/dept/linguistics/wp/2002.htm

Wise, M.R. (1999): 'Small language families and isolates in Peru', in Dixon, R.M.W. and Aikhenvald, A.Y. (eds) *The Amazonian Languages.* Cambridge: Cambridge University Press.