

The Diachrony of Classification Systems

EDITED BY
William B. McGregor
Søren Wichmann

JOHN BENJAMINS PUBLISHING COMPANY

THE DIACHRONY OF CLASSIFICATION SYSTEMS

CURRENT ISSUES IN LINGUISTIC THEORY

AMSTERDAM STUDIES IN THE THEORY AND HISTORY
OF LINGUISTIC SCIENCE – Series IV

ISSN 0304-0763

General Editor

JOSEPH C. SALMONS

University of Wisconsin–Madison

jsalmons@wisc.edu

Founder & General Editor (1975-2015)

E.F.K. KOERNER

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin

efk.koerner@rz.hu-berlin.de

Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals that advance our understanding of language, its structure, its function and especially its historical development. CILT offers an outlet for meaningful contributions to current linguistic debate.

A complete list of titles in this series can be found on <http://benjamins.com/catalog/cilt>

Editorial Board

Claire Bowern (New Haven, Ct.)

Sheila Embleton (Toronto)

Elly van Gelderen (Tempe, Ariz.)

John E. Joseph (Edinburgh)

Matthew Juge (San Marcos, Tex.)

Martin Maiden (Oxford)

Martha Ratliff (Detroit, Mich.)

E. Wyn Roberts (Vancouver, B.C.)

Klaas Willems (Ghent)

Volume 342

William B. McGregor and Søren Wichmann (eds.)

The Diachrony of Classification Systems

THE DIACHRONY OF CLASSIFICATION SYSTEMS

Edited by

WILLIAM B. MCGREGOR

Aarhus University

SØREN WICHMANN

Leiden University and Kazan Federal University

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM & PHILADELPHIA



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/cilt.342

Cataloging-in-Publication Data available from Library of Congress.

ISBN 978 90 272 0067 9 (HB)

ISBN 978 90 272 6413 8 (E-BOOK)

© 2018 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

Preface	VII
Abbreviations and conventions	IX
Introduction	1
<i>Søren Wichmann</i>	
The semantic reduction of the noun universe and the diachrony of nominal classification	9
<i>Frank Seifart</i>	
Niger-Congo numeral classifiers in a diachronic perspective	33
<i>Roland Kießling</i>	
Semantic generalization in Ch'orti' Mayan numeral classifiers	77
<i>Kerry Hull</i>	
Diachronic and synchronic aspects of the simplification of grammatical gender in an obsolescent language: The case of Irish	107
<i>Alessio S. Frenda</i>	
Numeral classifier systems in the Araxes-Iran linguistic area	135
<i>Don Stilo</i>	
The diachrony of Oceanic possessive classifiers	165
<i>Frantisek Lichtenberk[†]</i>	
Development and diffusion of classifier systems in Southwestern Amazonia	201
<i>Hein van der Voort</i>	
Nominal and verbal classification: A comparative perspective	241
<i>Walter Bisang</i>	
The diachrony of inflectional classes in four Germanic languages: What happens after transparency is lost?	283
<i>Antje Dammel and Sebastian Kürschner</i>	
The history of verb classification in Nyulnyulan languages	315
<i>William B. McGregor</i>	

Author index	353
Language index	355
Subject index	359

Preface

The origin of this book goes back to a conference, *The Diachrony of Classification Systems*, held at the Netherlands Institute for Advanced Studies in the Humanities and Social Sciences, Wassenaar, The Netherlands on 12–13 March 2009. This conference, convened by the editors, fell under the auspices of the European Cooperation in Science and Technology (COST) Action 31 *Stability and Adaptation of Classification Systems in a Cross-Cultural Perspective*, which ran from 2006 to 2010, and was chaired by Thekla Wiebusch. The editors gratefully acknowledge the financial and logistic contributions of COST to *The Diachrony of Classification Systems* conference. We are particularly grateful to Thekla Wiebusch for all the energy she put into the COST action, and for her support of our conference.

Six of the papers included in the present volume – those by Seifart, Frenda, van der Voort, Bisang, Dammel & Kürschner, and McGregor – were presented at that conference. To provide a wider geographical and genetic spread of languages, four additional contributions – Kießling, Hull, Stilo, and Lichtenberk – were solicited especially for this volume. We hope that these contributions will stimulate further research into the diachrony of classification systems in the world's languages. We also hope it will stimulate study of the development of other classification systems, including such systems in sign languages and writing systems.

During most of the editorial work Wichmann was supported by an ERC Advanced Grant (MesAndLin(g)k, project number 295918) and, during the final phase of editing, by another ERC Advanced Grant (EVOLAEMP, project number 324246).

We are grateful to the ten colleagues who served as anonymous external referees for the individual chapters in the volume, as well as to the contributors themselves, who served as internal referees, in many instances for more than one paper. Thanks also go to Jeroen Willemsen for preparing the indexes, and to the contributors for their patience and understanding over the book's long gestation and for responding promptly to our queries. We extend particular thanks to the General Editor of this series, Joe Salmons, for careful reading of the manuscript and seeing it through publication. The previous editor, Konrad Koerner, was also supportive of the project. Further acknowledgements for other contributions are given in the individual papers.

William B. McGregor & Søren Wichmann
Århus & Tübingen, November 2017

Abbreviations and conventions

1	first person	CL	noun class
1DIM	one dimensional	CAUS	causative
2	second person	CLF	classifier
2DIM	two dimensional	CM	conjugation marker
3	third person	CN	class noun
3DIM	three dimensional	CNT	continuous
A	agent-like argument; Set A person marker (Mayan languages only)	COIMCH	Consejo Indígena Maya Chorti
AB	absential	COM	comitative
ABL	ablative	COMACH	Coordinadora para el Desarrollo Integral Maya <i>Chortí</i>
ACC	accusative	COMP	comparative
ADV	adverb	CONALFA	El Comité Nacional de Alfabetización
AG	agentive	CONN	connector
ALL	allative	COP	copula
ALIM	alimentary (possessive classifier)	COS	change of state
ALMG	Academia de Lenguas Mayas, Guatemala	CRD	cardinal (case-form of free pronouns)
ASP	aspect marker	CVC	compound verb construction
ASS	associative	DAT	dative
ANIM	animate	DECL	declarative
ANT	anterior	DEF	definite
APPL	applicative	DEM	demonstrative
ART	article	DEP	dependent
AT	associative tone of possessor nouns	DET	determiner
ATT	attributive	DIM	diminutive
AUG	augmented number	DISC	discontinuative
AUGM	augmentative	DIST	distal
AUX	auxiliary	DR	directional
B	Set B person marker	DRINK	drink (possessive classifier)
BDR	bivalent direct	DS	different subject
BE	bound element	DU	dual
BE ₂	verb 'to be', type II	DUR	durative
BEN	benefactive	EMPH	emphatic
BER	beneficiary (nominal case marker)	EN	epenthetic nasal
BR	bound root	EP	epenthetic vowel
CATCH	catch (possessive classifier)		

ERG	ergative	OBJ	object
EV	evidential	OBJMK	object marker
EX	existential	OBL	oblique
F	feminine	OF	out-of-focus marker
FNS	final nominal suffix	OM	masculine singular oblique
FOC	focus	OP	plural oblique case
FOOD	food (possessive classifier)	OSII	oblique singular, type II
FRUST	frustrative	P	prefix
FUT	future	PERS	personal (suffix)
GEN	genitive	PFV	perfective
GER	gerundival	PL	plural
HABIT	habitat (possessive classifier)	PLFM	Proyecto Lingüístico Francisco Marroquín
HNC	human numeral classifier	PN	pronoun
IMP	imperative	PO	plural oblique case
INF	infinitive	POC	Proto-Oceanic
IMM.PST	immediate past	PORTION	portion (possessive classifier)
INAN	inanimate	POSS	possessive
IND	indicative	POT	potential
INF	infinitive	PP	past participle
INT	interrogative	PPĪ	past participle, type I
INTL	intentional	PRED	predicate
IPVF	imperfective	PRET	preterite
IRR	irrealis	PRF	perfect
IS	indefinite subject	PRFLX	perfective & reflexive
IV	inflecting verb	PRO	bound pronominal
LINK	NP linker, ezafe, etc.	PROG	progressive
LK	linker	PROX	proximal
LOC	locative	PRS	present
LV	linking vowel	PSHWNG	Proto-South Halmahera– West New Guinea
M	masculine	PST	past
MIN	minimal number	PTCP	participle
MOV	moving	PU	punctual
N	neuter	PURP	purposive
NC	numeral classifier	PV	preverb
NEG	negation	QUOT	quotative
NEGV	negative	REAL	realis
NEUT	neutral aspect	REF	referential
NMLZ	nominalization	REL	relativizer
NOM	nominative	REM	remote past
NONSG	non-singular	RP	repeater
NONSOLID	nonsolid oral consumption (possessive classifier)	S	suffix
NUM	numeral		

SBJ	subject	TREE	tree/tall plant (possessive classifier)
SBJV	subjunctive	TV	transitive
SCM	specific class marker	UNC	universal numeral classifier
SG	singular	VAL	valuable (possessive classifier)
SHELTER	shelter (possessive classifier)	VCL	verbal classifier
STAT	stative	VD	directional verb
SUB	subordination	VOL	volitive
SUBR	subordinator	∅	zero morpheme, semantically empty morpheme (in glosses)
SVC	simple verb construction	-	morpheme break
TAM	tense aspect mood	=	clitic break
THEME	thematic formant	#	word boundary
TRC	truncated element		

Introduction

Søren Wichmann

Leiden University and Kazan Federal University

1. Background

Rather little is known about the development of linguistic classification, whether in the verbal or nominal domains. These lacunae have motivated the present volume. The last couple of decades have seen much work directed at the study of systems of classification in languages (e.g., Aikhenvald 2000; Senft 2000; Schulze-Berndt 2000; McGregor 2002; Grinevald & Seifart 2004; Seifart & Payne 2007; Janda et al. 2013), and the history of research in this area has also become better documented (Kilarsky 2013, 2014). Most of this work, however, has focused on typological and synchronic aspects of these systems. This volume is unique in that it deals specifically with the diachrony of linguistic systems of classification.

The selection of papers aims to provide both a thematic and a geographical spread as regards the diachrony of classification systems. Collectively they revolve around three general topics: origins, developments, and diffusion of classifier systems. The following overview summarizes how each of these three topics are addressed.

2. Origins

The term ‘classifier’ is actually somewhat misleading, since it emphasizes an aspect of the phenomenon which is merely a byproduct of what the phenomenon is designed to achieve. Example (1), taken from Seifart’s contribution, shows two classifiers in action, *ix* “female non-kin” and *ixim* “(object consisting of) corn”. When (1a) is viewed in isolation from its discourse context, these two classifiers serve no other purpose than to redundantly specify some components of the meanings of *ix* “girl” and *b’itx* “corn bread”, which might be viewed as classification for the sake of classification. When the context (1b) is included, however, it becomes clear that the classifiers really serve to make their referents available for anaphoric reference.

- (1) Jacalteco noun classifiers (Mayan, Guatemala; Craig 1986: 264)
- a. *swatx' ix ix ixim b'itx*
 made CL girl CL tamale(corn_bread)
 “The girl made the tamales (corn bread).”
- b. *xtsoñ ix ixim yiñ how-eb' sentavo*
 sold CL:FEMALE_NON_KIN CL:CORN for five-PL cents
 “She sold them for five cents.”

Seifart makes the point in his contribution that a major function of classifiers is to enable anaphoric reference and suggests that their origin relates to this function.

Related to the question of origins is the question why some types of classifier system are more common than others. Bisang poses the simple question of why nominal classification is cross-linguistically more common than verbal classification. His argument goes as follows. The structural condition for nominal and verbal classification is the same, namely the availability of (syntactic or morphological) compounding followed by a process where the modifier develops into a classifier. So there is no a priori reason why verbal and nominal classification system should not be equally common. Assuming that both would have developed equally often, the question changes from why verbal classification systems arise less often to why they would seem to have disappeared more often. Bisang’s answer is that verbal classifiers are more prone to grammaticalize further into other functions in systems that are not compatible with the original classificatory function. For instance, if a classification system develops into an aspectual system with just two members – such as perfective and imperfective – this would spell the end of a classifier system. In contrast, a nominal classification system can develop new functions, such as a definite/indefinite distinction, and still continue to work as a classification system. This scenario is largely hypothetical since Bisang does not provide direct evidence to show that the grammaticalization path of verbal classification systems proposed is, indeed, common.

3. Developments

As Seifart observes, semantically transparent classification can become opaque, but opaque systems can be replaced over time by transparent ones, as shown in Frenda’s chapter on Celtic languages, for instance (cf. below). With reference to languages such as Chinese and Japanese, Seifart argues that classifiers tend to originate in open-class lexical nouns, so the normal situation would be for classifiers to start out being transparent.

Data presented by Kießling support this, since the incipient Niger-Congo numeral classifier systems that he describes are indeed semantically transparent. Opacity can arise if the physical properties of the classified nouns change or if classifiers coalesce in homonymy. While broad studies of numeral classifier systems such as Gil (2005) and Aikhenvald (2000) assume that numeral classifier systems are largely absent in Africa, Kießling shows that they are actually not infrequent.

Another area where there has been a tendency to overlook numeral classifier systems is the area that Stilo designates the Araxes Sprachbund (northern and central Iran, the Republic of Azerbaijan, Armenia, Georgia, eastern Turkey, and northern Iraq). Given that numeral classifiers are also incipient here, Stilo's data can shed further light on the rise of numeral classifier systems. As in the African data, one already notes an areal spread in spite of the apparent recency of the phenomenon. Thus, numeral classifiers are found in neighboring languages of different genealogical origins, including Azerbaijani (Turkic), colloquial Armenian (Indo-European/Armenian), Vafsi (Indo-European/NW Iranian), two Neo-Aramaic dialects (Afroasiatic/Semitic), Colloquial Georgian (Kartvelian), and some others. Characteristic is also the incompleteness of the Araxes-type system: there are just two classifiers, a generic one, originally meaning 'seed, grain', and a specific one optionally applied to persons. Thus, this type of system only borders on being a true classifier system since it does not classify at least a substantial fraction of nominals. It may, however, well be the case that more extensive classifier systems, approaching full taxonomies, have their origin in such partially-classifying systems.

Lichtenberk's chapter also offers observations on the development of classifier systems, dealing with the system of possessive classifiers in Oceanic (Austronesian) languages, which is used for alienable possession. The system descends from Proto-Oceanic, which is argued to have had three such classifiers. In the construction in question, known as 'indirect possession' in the literature, a possessive pronoun is affixed to a classifier and this expression then modifies a noun. Since some languages have extended the number of possessive classifiers while other have reduced it, leaving languages with anywhere between zero and a score or more classifiers, the Oceanic indirect possessive construction is ideal for an examination of the diachronic dynamics of classification systems. The original system, as reconstructed by Lynch et al. (2002), consisted in at least three classifiers: one for foodstuffs, one for drinks, and a generic one (there may have been more). Examples of expansions of this system include the addition of reflexes of *bula for valuable possessions in some languages of Vanuatu, the addition of a classifier *sa-* for domestic plants and animals in Southeast Ambrym, a system in Mussau with 14 classifiers (trees, buildings, food, juicy foods, cutting instruments, water craft, abstract personal items, personal relationships, domestic animals, things giving light, things to drink, pieces

of land, gardens, general things, things to suck juice from). Other languages, such as Iaaí or Chuukese, have even more extensive systems.

It stands to reason that classifiers, when they become more general in semantic content, will also become less transparent. This would hold true, for instance, for generic numeral classifiers, which are frequently found in numeral classifier systems. Hull presents an intermediate case, that of the Ch'orti' (Mayan) numeral classifier *-kojt*, which is derived from the verb root *kot* "to kneel down" and originally referred to quadrupeds. Through generalization, partly driven by conscious efforts of language activists, this has been broadened, apparently in the latter half of the 20th century, to also classify other animals and humans, and even to objects such as tortillas.

Agreement systems are usually quite opaque since they tend to distinguish few categories. But such systems may become more transparent under conditions of language attrition. Frenda describes remarkably similar developments in both the last stages of extinct Celtic languages and Celtic languages which are still spoken, but which have become severely endangered. Thus, the last speakers of now-extinct Manx and Cornish ceased to distinguish the inanimate gender, and precisely this gender is also collapsing to various degrees among present-day speakers of Welsh and Irish. The tendency is to use masculine pronouns to refer anaphorically to inanimates.

A classification system can expand or get lost, but it can also mutate into an inflectional class system whose status as functional becomes questionable. Dammel and Kürschner examine this situation in several Germanic languages: Swedish, Danish, Dutch, and German.

4. Diffusion

The Proto-Oceanic system discussed by Lichtenberk is clearly involved in diffusion, although the details remain unclear. The system consists of a contrast between direct possession, restricted to inalienable possession, where the possessor is suffixed to the possessum, and indirect possession, indicating alienable possession, where the possessum is followed by one of three different classifiers to which the possessor is suffixed. The three different classifiers comprise a generic one, one referring to food, and one to drink. The sister group of Oceanic within Eastern Malayo-Polynesian, i.e., South Halmahera-West New Guinea, has a similar system of direct vs. indirect possession, the only difference being that there is only one non-generic classifier, which is one for food. This has the shape **na-*, the same as the generic classifier in Proto-Oceanic. Therefore, van den Berg (2009) suggests the existence of a Proto-Eastern Malayo-Polynesian system of

direct vs. indirect possession. It seems that *na- originated in a definite article, so Lichtenberk sees it as more likely that *na- was a generic classifier than a classifier for food. Other than that, it is unclear how to reconstruct the original system. More interestingly, perhaps, Klamer et al. (2008) and Donohue & Schapper (2008) both attribute the existence of the contrast between direct and indirect possession to diffusion from Papuan (non-Austronesian) languages. While Lichtenberk points out that this scenario needs to be fleshed out better in terms of identifying the relevant Papuan family of languages responsible for the transfer of the pattern, he argues that it is plausible and could have happened in a situation where a considerable number of speakers of some Papuan language shifted to Proto-Eastern Malayo-Polynesian in the latter's hypothetical homeland in or around the Bird's Head of New Guinea some 3400 years ago. The reason why he argues for a language shift is that Proto-Oceanic does not evince many Papuan loanwords, so a transfer via language contact is not likely – such a type of transfer would be expected to be accompanied by loanwords (Thomason & Kaufman 1988).

Van der Voort looks at languages of the Guaporé and Mamoré river basins in Southwestern Amazonia. Just as research has shown Northwestern Amazonian languages to share features of classifier systems, there are also shared similarities in Southwestern Amazonia. Classifiers in Amazonia stand out as being 'multi-functional' (Krasnoukhova 2012), occurring in many different grammatical environments, including numeral, possessive, attributive, demonstrative, nominal, and verbal environments. Van der Voort investigates classifiers in the isolates Kwaza, Kanoë, Aikanã, Movima, Itonama, and the following representatives of larger families: Lakondê (Nambikwaran), Cavineña (Tacanan), Arikapu (Jabutian), Baure (Arawakan), Karo (Tupian). The following are some of the generally shared characteristics revealed by his survey. In all the languages classifiers may combine with nominal roots to yield new nouns. In all but Karo they can also combine with verbal roots as incorporated elements or modifiers. In Kwaza, Movima, and Lakondê, classifiers can further function as nominalizers, as in Example (2), from Kwaza (van der Voort's example (31)).

- (2) *axy haje-a-mũ*
 house smear-1PL-CLF:liquid
 "paint" (lit. "the liquid we smear onto the house")

A particularly conspicuous shared feature of the Southwestern Amazon classifier systems is that a classifier or bound noun root may combine with a semantically empty (dummy) root to form a free noun (*e-* in Kwaza; *i-* in Kanoë; *ka-*, *i-*, or *e-* in Aikanã; *-i* in Movima; *ĩ-* in Lakondê; *i-* or *a-* in Sabanê; *e-* in Cavineña). Related to this phenomenon is that of similarly shaped prefixes used to express dummy possessors allowing for obligatorily possessed nouns to be used in a non-possessed

sense. In Arikapu the first person singular (*i-*) or plural (*txi-*) markers may be used for this purpose (e.g., *i-txawa* “my flower” or simply “flower”), and in Baure an impersonal possessor morpheme *e-* is used when the possessor is not specified (e.g., *e-ser* “someone’s tooth”).

The various shared general characteristics combined with the fact that specific shapes of some classifiers are widely shared is a very strong argument in favour of diffusion. Van der Voort offers speculations on specific directions of diffusion in some cases, but the overall picture is not at all transparent. Thus, while the circumstances under which these systems have diffused remain unclear, the evidence suggests that they have diffused fairly easily, something which may be true of classifier systems in general as suggested by other contributions in the volume, including McGregor’s.

McGregor offers diachronic data on an entirely different kind of classification system, that of verb classification. Although the focus is on Nyulnyulan languages, McGregor mentions that the type of system in question is widespread in northern Australia and is likely to have been diffused. In the typical system, an uninflecting verb (or preverb) is followed by an inflecting verb, forming a compound verb construction. The uninflecting verbs constitute an open class, whereas the class of inflecting ones is comprised by anywhere from some ten to some two hundred members. McGregor views the inflecting verbs as elements that classify the uninflecting preverbs that they collocate with in compound verb constructions. Whereas verb classification is seen as having diffused across northern Australia, the similarities among Nyulnyulan languages are argued to be inherited. McGregor proposes that at least eight inflecting verbs as well as a number of specific collocations of preverbs and inflecting verbs can be reconstructed for Proto-Nyulnyulan. Thus, the timing of diffusion would have been before the break-up of the Nyulnyulan family.

The various studies in this volume discussing diffusion contribute to the general impression that classification systems are prone to diffusion. Based on Seifart’s contribution one should perhaps add the hedge that semantically transparent systems are more prone to diffusion than semantically opaque systems. Seifart points out, with reference to Nichols (2003) and Wichmann & Holman (2009), that classifiers are more prone to diffusion than genders and noun classes, arguing that this is because classifiers are more transparent, something which he sees as a prerequisite for diffusion.

5. Final remarks

It is our hope that the papers included in this volume may stimulate further research on the diachrony of classification systems. The information that would allow us to flesh out the typical life-cycle of a classification system, from birth or adoption (borrowing) over growth (differentiation of categories) to decay, is still too scanty. Too little is known about the reconstructibility of systems within language families across the world, although this volume has improved the situation a little with information on two language families. We have a growing number of documented cases of diffusion, but it is still not clear whether it is exceptional or normal that some morphemes are diffused along with the idea of classification as has happened in Southeastern Amazonia, and we know next to nothing about the social processes involving language use which condition diffusion. Thus, the present volume is a small step towards a situation where we will be able to make broader generalizations about the diachrony of classification systems, and we hope that this step is worthwhile.

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Craig, Colette G. 1986. "Jaltec Noun Classifiers: A study in grammaticalization". *Lingua* 70: 4. 241–284.
- Donohue, Mark & Antoinette Schapper. 2008. "Whence the Austronesian Indirect Possession Construction?" *Oceanic Linguistics* 47.316–327.
- Gil, David. 2005. "Numeral Classifiers". *The World Atlas of Language Structures* ed. by Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, 226–229. Oxford: Oxford University Press.
- Grinevald, Colette & Frank Seifart. 2004. "Noun Classes in African and Amazonian Languages: Towards a comparison". *Linguistic Typology* 8.243–285.
- Janda, Laura A., Anna Endresen, Julia Kuznetsova, Olga Lyashevskaya, Anastasia Makarova, Tore Nessel & Svetlana Sokolova. 2013. *Why Russian Aspectual Prefixes Aren't Empty: Prefixes as verb classifiers*. Bloomington, Ind.: Slavica Publishers.
- Kilariski, Marcin. 2013. *Nominal Classification: A history of its study from the classical period to the present*. Amsterdam: John Benjamins.
- Kilariski, Marcin. 2014. "The Place of Classifiers in the History of Linguistics". *Historiographia Linguistica* 41.33–78.
- Klamer, Marian, Ger Reesink & Miriam van Staden. 2008. "East Nusantara as a Linguistic Area". *From Linguistic Areas to Areal Linguistics* ed. by Pieter Muysken, 95–149. Amsterdam: John Benjamins.
- Krasnoukhova, Olga. 2012. *The Noun Phrase in the Languages of South America*. Utrecht: LOT Publications.

- Lynch, John, Malcolm Ross & Terry Crowley. 2002. *The Oceanic Languages*. Richmond, Surrey: Curzon Press.
- McGregor, William B. 2002. *Verb Classification in Australian Languages*. Berlin: Mouton de Gruyter.
- Nichols, Johanna. 2003. "Diversity and Stability in Language". *The Handbook of Historical Linguistics* ed. by Brain D. Joseph & Richard D. Janda, 283–310. Malden, MA: Blackwell.
- Schultze-Berndt, Eva. 2000. *Simple and complex verbs in Jaminjung: a study of event categorisation in an Australian language*. Ph.D. dissertation, Catholic University of Nijmegen.
- Seifart, Frank & Doris L. Payne. 2007. "Nominal Classification in the North West Amazon: Issues in areal diffusion and typological characterization". *International Journal of American Linguistics* 73: 4.381–387.
- Senft, Gunther, ed. 2000. *Systems of Nominal Classification*. Cambridge: Cambridge University Press.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- van den Berg, René. 2009. "Possession in South Halmahera–West New Guinea: Typology and reconstruction". *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust* ed. by Alexander Adelaar & Andrew Pawley, 327–357. Canberra: Pacific Linguistics.
- Wichmann, Søren & Eric W. Holman. 2009. *Temporal Stability of Linguistic Typological Features* (= LINCOS Studies in Theoretical Linguistics, 42). Munich: LINCOS EUROPA.

The semantic reduction of the noun universe and the diachrony of nominal classification

Frank Seifart

University of Amsterdam & University of Cologne

Classifiers and noun class markers are often semantically general and semantically opaque compared to open-class nouns, and in this sense they constitute a semantic reduction of the noun universe. These two semantic characteristics also play important roles in the diachronic development of nominal classification systems. First, the need for semantically general forms for anaphoric reference may be a possible motivation for developing nominal classification in the first place. Second, opaque classification, which may, for example, emerge through coalescence of classes with homophonous markers, may be replaced by transparent classification because of the incompatibility of opaque classification and certain syntactic constructions, such as contrastive focus. Finally, opaque classification, typical of grammatical gender systems, is less likely to diffuse through language contact than transparent classification, which is typical for other types of systems, including numeral classifier systems.

1. Introduction

Nominal classification systems are language-specific, conventional reductions of the potentially infinite semantic detail provided by open-class nouns into a finite, often rather small, number of classes. This paper discusses how this central reducing aspect of nominal classification may bear, in important ways, on the diachronic development of nominal classification systems. The discussion focuses on two aspects of this semantic reduction – semantic generality and semantic opacity – and the interaction of these with morphosyntactic properties of nominal classification. Such properties include the regular collocations of nouns with classifying expressions, which ultimately define noun classes, and different morphosyntactic types of classification systems (for extensive discussion, see Aikhenvald 2000; Grinevald 2000). This discussion builds to some extent on previous research on nominal classification in the Northwest Amazonian language Bora-Miraña (Seifart 2005), putting selected aspects of that research into a new perspective and drawing on

additional examples from other languages. Little research has been done so far on the interaction between semantic opacity and generality and the diachronic development of nominal classification. Therefore some of the ideas presented in this paper are speculative and exploratory.

The following section (§2) provides a definition of nominal classification. Section 3 discusses a motivation for classification systems to develop in the first place, which is based on the interpretation of semantically general classifiers as anaphora. Section 4 takes a closer look at the notion of semantic generality and introduces the terms semantic opacity and semantic transparency. These notions describe the extent to which a classifying morpheme carries appropriate ‘descriptive content’ with respect to the noun it classifies. I then discuss a number of ways semantically transparent classification can become opaque. A central aim here is to show how semantically opaque classification has a distributional restriction in a number of contexts, where it is often diachronically replaced by semantically transparent classification. Finally, §5 relates semantic opacity to the areal diffusability of nominal classification, suggesting that the general prediction that more strongly grammaticalized items are more resistant to borrowing than lexical items also holds for opaque classification vs. transparent classification. Section 6 concludes.

2. A definition of nominal classification

Before entering the main discussion, it is useful to provide a definition of nominal classification under the following four criteria (also applied in Seifart 2010), an adaptation of McGregor’s (2002: 16–22) definition of ‘grammatical superclassification’:

- a. nouns collocate in well-defined grammatical environments with classificatory elements (these may be free forms, clitics, affixes, etc., and these may also occur elsewhere);
- b. the number of classificatory elements is larger than one but significantly smaller than the number of nouns;
- c. classificatory elements show different patterns of collocation with nouns, i.e., they impose a classification (some overlap is allowed; typically, but not always, there is a relatively equal division of the nominal lexicon by classificatory elements);
- d. at least a substantial subpart of nouns are classified in this way.

This definition captures the generalizing effect of classificatory morphemes (my current focus) with the requirements that there must be significantly fewer of these than the number of nouns, and that these must show different patterns of collocation with nouns. Additionally, this definition covers different morphosyntactic

instantiations of nominal classification, including strongly grammaticalized gender and noun class systems, as well as lexico-syntactic classifier systems, such as numeral classifiers. Previous definitions (e.g., Allan 1977: 285; Aikhenvald 2000: 13) tend to apply to the latter rather than to the former. In the following, I use the term ‘classifying morpheme’ as a cover term for the elements of classifier systems as well as of noun class and gender systems.

3. The pragmatics of semantic generality

Reference-tracking, i.e., the establishment of an anaphoric link to a previously mentioned noun, is a common function of nominal classification systems, irrespective of morphosyntactic type (see Contini-Morava & Kilarski 2013 for a comprehensive overview). The following examples illustrate the anaphoric use of Miraña noun class markers (1), of Jacalteco noun classifiers (2), and of Caddo verbal classifiers (3) (antecedent noun phrases and classifying morphemes that establish anaphoric links are in boldface).

- (1) Miraña noun class markers (Boran, Peru; own data)
- a. *uhkuú-ʔi tuhkénuú tsa-né ahtʃúú-ró:ru*
 take-PRED begin.NMLZ one-CL.INAN shine-CL:CHUNK
 “... took first **one flashlight** ...”
- b. *arró-náa tsáʔ té-ro:ru pé:te-tuú-ne*
 but-after NEG PN-CL:CHUNK SUB.burn-NEG-CL.INAN
 “... but then it (chunk, i.e., flashlight) did not work.”
- (2) Jacalteco noun classifiers (Mayan, Guatemala; Craig 1986: 264)
- a. *swatx’ ix ix ixim b’itx*
 made CL girl CL tamale(corn_bread)
 “The girl made the tamales (corn bread).”
- b. *xtson’ ix ixim yin’ how-eb’ sentavo*
 sold CL:FEMALE_NON_KIN CL:CORN for five-PL cents
 “She sold them for five cents.”
- (3) Caddo verbal classifiers (Caddoan, Oklahoma; Mithun 1984: 865)
- a. *kas-sah-kú-n-dân-na-’na’ kišwah*
 should-2.A-1.BEN-DAT-CL:GRANULAR-PL-make parched.corn
 “You should make me some parched corn.”
- b. *nas-sah-kú-n-dân-na-’nih-áh*
 when.FUT-2.A-1.BEN-DAT-CL:GRANULAR-PL-make-PERF
sinátti’ ci:yáhdí’a’
 then I_will_go_on
 “When you have made it (the granular substance) for me, then I will go on.”

In the following, I characterize classifying expressions as semantically general with respect to the nouns they classify, in order to relate the anaphoric use of classifying morphemes to a general principle of interpretation of semantically general nominal expressions. This, I suggest, is an important factor in the initial development of nominal classification.

I assume for discussion here (but see §3) the simple case that a classifying morpheme is a hypernym of the noun it classifies. This means that the extension of the meaning of the classified noun is a subset of the extension of the meaning of the classifying morpheme. In this view nominal classification is a taxonomy of the nominal lexicon, as illustrated in Figure 1 for a subsection of the Miraña nominal classification system.

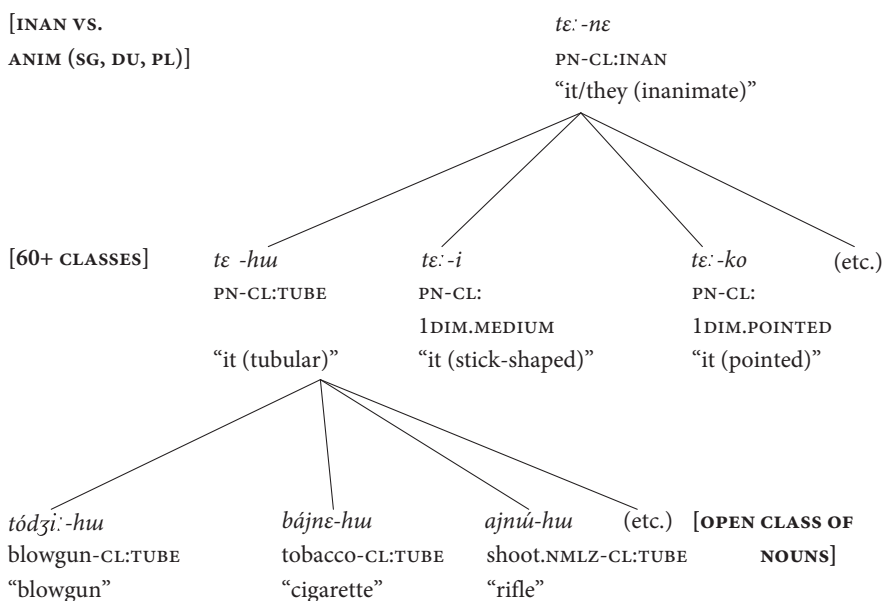


Figure 1. Hypernymic relations between Miraña nominal expressions

The lower part of Figure 1 illustrates that the open class of Miraña nouns is categorized by approximately 60 classifying elements. One of them is *-huu* “CL:TUBE”, which in Figure 1 appears suffixed to the semantically empty (or weak) root forming third person pronouns (*tɛ:-*). This expression is a hypernym with respect to the nouns which *-huu* “CL:TUBE” classifies, such as *tódzi:-huu* “blowgun”, *bájne-huu* “cigarette”, *ajnú-huu* “rifle”, etc. The upper part of Figure 1 illustrates that these approximately 60 classes are again categorized by a small set of ‘general class markers’ (which may occur in the same morphosyntactic slots as specific class markers,

e.g., with the pronominal root *tɛ:*). These categorize the nominal lexicon into a few general classes, including an ‘inanimate’ class (under which the classes in the lower part of Figure 1 fall), as well as animate singular, dual, and plural classes (not shown in Figure 1).

If one attempts a componential analysis of the semantic features in the intention of classifying expressions (Nida 1975), semantic generality can be represented as the elimination of semantic components, as in (4), to illustrate the semantic reduction of the noun universe through nominal classification.¹ The expression *tɛ:-huu* in (4b) is thus stripped of all semantic specifications of the noun *tódzi:-huu* ‘blowgun’ except for [inanimate], [singular], and [tube-shaped]. Miraña presents an additional, even more general classification device, so-called ‘general class markers’, which at the next higher level generalizes across all inanimate nominal expressions and all number values.

- (4) a. *tódzi:-huu* [inanimate] [singular] [tube-shaped] [hunting instrument] ...
 b. *tɛ:-huu* [inanimate] [singular] [tube-shaped]
 c. *tɛ:-nɛ* [inanimate]

The selection of semantic features that are preserved in semantically general classifying expressions is of course language specific, but the principle of semantic generality is common in nominal classification systems: otherwise they would not classify.

Given the above, the relation between a classifier and the classified noun can be viewed from the perspective of a general rule for the interpretation of semantically general nominal expressions, represented in Figure 2 (based on Givón 1983a: 18; Ariel 1988: 84; Gundel et al. 1993: 284; Levinson 2000: 267; Chafe 1994: 71–72). According to this principle, the use of a semantically general expression which provides only partial information about a nominal referent signals that the intended referent is already present in the discourse world and that the intended referent is not a new referent (see Seiler 1986 for an early discussion of the relationship between semantic generality and anaphora).

1. Note that I do not claim to present the only possible or the only correct componential analysis. I submit, however, that any reasonable componential analysis will show that classifying morphemes are semantically more general than the nouns they classify.

<i>semantic specificity:</i>	semantically specific		semantically general
<i>expression type:</i>	lexical noun	pronoun	zero
		free ----- bound	
<i>usual interpretation:</i>	new or inactive referent, non-coreferential reading		old or active referent, coreferential reading

Figure 2. Types of referential expressions and default interpretation

Example (5) illustrates this principle, showing that a coreferential (i.e., anaphoric) reading of the semantically relatively general noun *vessel* with the semantically relatively specific preceding noun *ferry* is naturally obtained in (5a). On the other hand, a semantically relatively specific noun such as *ferry* in (5b) cannot usually receive a coreferential interpretation with respect to the semantically relatively general preceding noun *vessel*, but it would normally be interpreted as introducing a new referent. A third person pronoun such as *it* in (5c), which is semantically general, is naturally interpreted as coreferential, according to the principle shown in Figure 2.

- (5) *Semantically general expressions and coreferential reading* (based on Levinson 2000: 269)
- The ferry_j hit the rock. The vessel_j capsized.*
 - The vessel_j hit the rock. The ferry_{j/i} capsized.*
 - The ferry_j hit the rock. The vessel_j capsized. It_j sank immediately.*

A componential analysis of the semantic features of the relevant expressions in (5) is given (6), illustrating that semantically relatively general expressions receive a coreferential (anaphoric) interpretation with respect to relatively specific expressions, but not the other way around.

- (6) a. *ferry* [inanimate] [singular] [floating] [artifact] [public transport] ...
 b. *vessel* [inanimate] [singular] [floating] [artifact] ...
 c. *it* [inanimate] [singular]

There is much empirical evidence for regular patterns of allocation of the amount of semantic information in discourse according to the basic schema sketched in Figure 2. Cross-linguistic evidence for this comes, for instance, from studies in the ‘topic continuity’ framework (see the contributions in Givón 1983b; see also Daley 1998; Fox 1987: 137–140; Payne 1988). It should be noted that this allocation of information (and accordingly the choice for an anaphoric expression) is subject to additional factors such as the topicality of the referent (Givón 1983a) and the paragraph structure, which may warrant an otherwise unexpectedly explicit mention of

a referent to signal the end of a paragraph (Fox 1987; on the function of Miraña class markers in this respect, see Seifart 2005: 245–306). As expected from a pragmatic principle (as opposed to a morphosyntactic rule), deviations are always possible, some of them under specifiable conditions such as signaling the end of a paragraph, as just mentioned.

The semantic generality itself is at least largely if not solely responsible for the pragmatic effect of coreferential readings, i.e., not (necessarily) the part of speech distinctions, such as noun vs. pronoun, free vs. bound pronoun. This is evident from (5a)–(b), where the pragmatic principle applies to two lexical nouns, i.e., members of the same part of speech. Another piece of evidence for the primacy of semantic generality over the part of speech distinctions in determining coreferential vs. non-coreferential readings comes from Miraña, which allows for two levels of semantic generality in its pronouns, general class markers, and specific class markers. Example (7), a continuation of (1), illustrates how the two types of class markers are used in reference tracking: On the one hand, a ‘specific class marker’ is used in (7a), which specifies the shape of the referent. On the other hand, a ‘general class marker’ is used in (7b), which only specifies inanimacy (see Figure 2, above) and is therefore semantically even more general. Example (7c) illustrates how the two levels of semantic specificity in Miraña reference-tracking pronouns are exploited to mark the end of a paragraph by an otherwise unexpectedly specific expression, in this case a ‘specific class marker’ used to mark a paragraph boundary.

- (7) a. *a:ró-náa tsáʔ tɛːʔo:u pɛːtɛ-tú-ne*
 but-after NEG PN-CL:CHUNK SUB.burn-NEG-CL:INAN
 ‘... but then it (chunk, i.e., flashlight) did not work ...’
- b. *a:-ne pɛːtɛːtú-né-dʒiːʔe*
 CONN-CL:INAN SUB.burn-NEG-CL:INAN-BEN
 ‘... and because it (inanimate) did not work ...’
- c. *píkoːbe ɪːne ɪːnú-hí-ʔadʒú-βú tɛːʔo:u*
 put-CL:masc.sg HES earth-CL:2DIM.ROUND-ON-ALL PN-CL:chunk
 ‘... he put, eh, it (chunk, i.e., flashlight) on the ground.’

The principle governing the coreferential readings of semantically general expressions is powerful and generally applicable. As such it helps ground the existence of nominal classification on general, communicative principles. Since it operates independently of parts-of-speech distinctions, it may predate the development of separate parts of speech that are semantically general and that serve dedicated reference-tracking functions, such as classifiers. Furthermore it may play a crucial role in shaping these. It is well known that classifiers develop from subsets of semantically general nouns. From the above discussion it follows that these may also be used for reference tracking, according to the above principle, in languages

without nominal classification. If they are used frequently in this function, they may eventually grammaticalize into closed sets of classifying morphemes (given a number of other factors, such as a dedicated classifier construction). The property of classifying morphemes as being semantically general may thus play an important role in the initial emergence of nominal classification as a reference tracking device.

The perspective on the emergence of nominal classification taken in this section additionally allows for some observations on the semantic contents of classifying morphemes as reference tracking devices. According to the principle mentioned above, dedicated anaphoric expressions (from which classifiers may develop) are semantically general and as such preserve only a subset of the semantic features present in the antecedent noun. But which features will these be? Which are the features of nouns that are appropriate as partial semantic specifications about a given referent to signal coreference? According to Givón (1976: 171), such features “represent only the top of the hierarchy of semantic features that underlie the noun universe” (see also Lehmann 1988: 61–62; Barlow 1992: 46–50). The semantics of nominal classification used for reference tracking may thus be used to explore which features are at the top of such hierarchies. Here, universal tendencies can be expected, such as the prominence of animacy and natural gender distinctions, as well as culture-specific classification preferences, such as social status (Aikhenvald 2000: 175–280).

In the domain of inanimates, it seems much less clear which features may be at the top of such a hierarchy. In the first three examples given above, the meaning components that are preserved in reference-tracking classifiers are quite diverse: shape (Miraña “chunk” for “flashlight”), material (Jacaltec “corn” for “tamale corn bread”), and consistency or configuration (Caddo “granular” for “parched corn”). Among the semantic features that are most often mentioned as being important in nominal classification of inanimates are shape, size, material, consistency, and function (Aikhenvald 2000: 275–280). However, generalizations about semantic domains have focused on differentiating morphosyntactic types of nominal classification systems, showing, for example, that the semantic domain ‘edibility’ is almost exclusively found in genitive classifiers and that the semantic domain ‘material’ is prevalent in noun classifiers (see Croft 1994; Aikhenvald 2000: 271–306; Grinevald 2000: 71–74). On the other hand, systems of various types (at least noun classes, numeral classifiers, and noun classifiers) may be used for reference tracking. Therefore it is still not clear which semantic features are most directly associated with the semantic generality of reference tracking and the emergence of nominal classification. The (reduced) semantic specification of a referent in reference-tracking classifying morphemes does not necessarily represent what is most important about this referent for the speakers, but it represents the appropriate semantic reduction for creating an anaphoric expression.

In sum, the fact that classifying morphemes are semantically general with respect to the nouns they classify can be related to a possible origin of nominal classification as a reference tracking device. On the one hand, this motivates the emergence of conventional and eventually grammaticalized morphosyntactic nominal classification systems. On the other hand, the role of semantic generality in reference tracking enables a new perspective on the selection of semantic domains encoded in nominal classification as those that are preserved in semantically general anaphoric expressions.

4. Semantic opacity and renewal

This section discusses further aspects of the semantic generality of classifying morphemes, in particular the distinction between semantically opaque and semantically transparent classification, and the role that this distinction plays in the diachronic process of the internal reorganization of nominal classification. I first discuss how semantically opaque noun class assignment can be distinguished from semantically transparent assignment (§4.1). Section 4.2 discusses briefly some aspects of the well-known fact that semantically transparent assignment can become opaque over time. Section 4.3 presents evidence for the opposite, less well-studied tendency, namely that opaque classifications can be replaced by semantically transparent classifications.

4.1 Descriptive content of classifiers, semantically transparent and opaque classification

In order to discuss the distinction between semantically transparent and opaque classification, it is useful to first introduce the notion of ‘descriptive content’ for a classifying morpheme (borrowing this term from Bosch 1988, who applies it to the meanings of pronouns; see below). Essentially, the descriptive content is the meaning that can be attributed to the classifying morpheme, and this is to be established for each classifying morpheme individually. As will become clear below, this notion is useful because there are differences in the extent to which classifying morphemes have descriptive content at all. A further question is how the descriptive content of a classifying morpheme relates to the meaning of the classified noun. This cannot be established once and for all for each classifying morpheme, but has to be established for each association of a classifying morpheme with a noun. Thus it distinguishes between semantically transparent vs. semantically opaque noun class assignment for each noun.

If a classifying morpheme is a free form, its descriptive content can be established in the same way as the meaning of any lexical item. If a classifying morpheme is bound, there may nevertheless be constructions in the language which allow us to isolate the descriptive content of the classifying morpheme. In *Miraña*, noun class markers can be combined with semantically weak pronominal stems, and these combinations can be used in a nominal or predicative function, as in Example (8) (descriptive content may also surface particularly well under contrastive focus; see below). For *Miraña*, this yields descriptions of the descriptive content of class markers, such as *-zo* “three-dimensional, oblong object” and *-ko-duu* “two-dimensional, pointed object”.

(8) *Predicating the meaning of a class marker over a referent in Miraña*

- a. *í-ne pá-zo-duú néɛ-ne*
 this-CL:INAN COP-CL:3DIM.OBLONG-COMP seem-CL:INAN
 “This is an oblong one.”
- b. *í-ne pá-ko-duú néɛ-ne*
 this-CL:INAN COP-CL:2DIM.POINTED-COMP seem-CL:INAN
 “This is a pointed one.”

Having established the descriptive content of a classifying morpheme (assuming, for now, that classifying morphemes have descriptive content), the next question is whether a noun that is grouped into a given noun class by the morphosyntactic nominal classification system of the language semantically relates to the descriptive content of the classifying morpheme. If it does, one can speak of semantically transparent noun class assignment. If not, one can speak of semantically opaque noun class assignment. In *Miraña*, the following test can be used to assess the semantic transparency or opacity of noun class assignment. This test is based on the predicative use of class markers, i.e., on the possibility of using class markers in a nominal expression that is used as a predicate nominal to attribute the properties denoted by the class marker to the referent of the subject of that clause (see also Example (8)). If the referent of a noun can be described with a predicate nominal that includes the same class marker that is also included in this noun, then the class marker assignment can be called semantically transparent, as in Example (9). In this case, the descriptive content of the class marker ‘fits’ the meaning of the noun.

(9) *Semantically transparent noun class assignment in Miraña*

- a. *úhi-zo pá-zo-duú néɛ-ne*
 banana-CL:3DIM.OBLONG COP-CL:3DIM.OBLONG-COMP seem-CL:INAN
 “A banana is like an oblong one.”
- b. *ka:túnuu-í:zo pá-í:zó-duú néɛ-ne*
 writing-CL:LITTLE_STICK COP-CL:LITTLE_STICK-COMP seem-CL:INAN
 “A pencil is like a little stick.”

rather than a categorical distinction. The scalar property reflects the gradual nature of the acceptability of constructions like those given for German and Miraña above. Note also that the notion of opaque assignment does not mean the absence of semantic categorization principles. On the contrary, there may well be semantic principles or semantic ‘motivation’ at work in opaque cases of assignment. For instance, Zubin & Köpke (1986) show that in German the vast majority of ‘superordinate terms’, such as *Obst* ‘fruit’, *Werkzeug* ‘tool’, and *Tier* ‘animal’, are neuter gender, while the associated ‘basic level terms’ tend to be either masculine or feminine, e.g., *Apfel* ‘apple(M)’, *Pflaume* ‘plum(F)’, *Traube* ‘grape(F)’. Another semantic principle is that nouns denoting highly imageable taxa, such as tree, bush, flower, or mushroom, tend to be masculine, while nouns denoting internally diverse taxa, such as *Kraut* ‘herb, cabbage, etc.’, tend to be neuter (see also Lakoff 1986, 1987 on Dyirbal). But this type of distant semantic motivation is of a different nature than semantically transparent assignment, which involves a straightforward ‘fit’ of the descriptive content of a classifying morpheme with the semantics of a classified noun. In the context of this paper, it is useful to maintain this distinction, even if it has fuzzy limits, because it plays a crucial role in some diachronic processes in nominal classification, as is shown in the following sections.

descriptive content <-----> fits classified noun	descriptive content does not fit classified noun
transparent assignment	opaque assignment

Figure 3. Semantically transparent and opaque noun class assignment

4.2 From semantically transparent to opaque assignment

Since classifiers overwhelmingly originate as open-class lexical nouns and have the semantic properties of lexical nouns, it is fair to assume that at early stages of development, nominal classification tends to be semantically transparent. How then does it become opaque? Due to lack of historical data for most languages, not many concrete processes of ‘opacitization’ have been described (see Erbaugh

1986; Wiebusch 2009 on Chinese; Downing 1996 on Japanese). Implicitly, at least, it seems to be often assumed that semantic extensions of noun classification categories that are found in synchronic data also correspond to diachronic development. Examples of such extensions described by Lakoff (1986) are, for instance, from fish to fishing implements and from women to birds, based on a belief that birds are the souls of deceased women (see Aikhenvald 2000: 404–408 for more examples). If the classifying morphemes retain descriptive content (e.g., ‘feminine’) at all, then the result of such extensions, especially if they are applied successively in chains (e.g., from ‘women’ to ‘sun’ to ‘sunburn’ to ‘hairy mary grub’, an insect that stings like a sunburn), is semantically opaque noun class assignment in the sense of §4.1. In Lakoff’s (1986) terms, such nouns would be less central members of a class.

In this section, I present some data from Miraña that illustrate perhaps less well-known processes, namely, different kinds of historical ‘accidents’ that may also contribute to a process by which originally transparent classification becomes opaque. First, the association of a noun with a classifying morpheme may become semantically less transparent if the noun changes semantically. This may be due to changes in the physical properties of the referents themselves. For instance, the spoons and axes the Miraña people used traditionally were differently shaped from those used today (13a)–(b), while the nouns denoting them, including their noun classes, remain the same. As a result, the noun class assignments of Miraña nouns for spoons and axes are more opaque now than they used to be.

(13) *Reduced semantic motivation through change of referents in Miraña*

a. *deihhu-gwa*

to_spoon_up.NMLZ-CL:2DIM.STRAIGHT

“spoon” (-*gwa* “flat, rigid, at least one **straight** edge”, traditionally pieces of wood with straight edges were used as spoons).

b. *ugwá:-hi*

metal-CL:2DIM.ROUND

“axe” (-*hi* “flat and **round**”, traditional stone axes are round).

A second historical process that may result in semantic opacity is when classifying morphemes become homophonous. In order to show this, I first illustrate how Miraña noun class markers most probably entered the system. This was by repeaters, i.e., the repetition of a noun in the class marker slot (14), with subsequent truncation of that noun in class marker slots (15). The form in Example (15b) is glossed as a class marker since it has effectively been extended to other nouns, although its origin as a partial repeater is still clearly recognizable.

- (14) Noun used as ‘repeater’ in agreement slot
- a. *ihka-báhuu* *tsá-bahuu* *báhuu*
 COP.SUB-RP.FOREST one-RP.FOREST forest
 “There is one (stretch of) forest.”
- b. *ihka-báhuu* *tsá-bahuu* *úhi-báhuu*
 COP.SUB-RP.FOREST one-RP.FOREST banana-forest
 “There is one banana plantation.”
- (15) Phonological reduction (partial repeater)
- a. *ihka-mo* *tša-mo* *mó:aj*
 COP.SUB-RP.RIVER one-RP.RIVER river
 “There is one river.”
- b. *ihka-mĩ* *tša-mĩ* *mĩ:nɛ*
 COP.SUB-CL:TRANSPORT one-CL:TRANSPORT canoe
 “There is one canoe.”

Examples (16a)–(c) illustrate the potential emergence of an opaque noun class through the coalescence of three homophonous partial repeaters, which could potentially become a single noun class in the further development of the system. If this marker retains descriptive content at all, based on the nouns from which it grammaticalizes, i.e., ‘signal drum’, ‘breast’, or ‘umari fruit’, this descriptive content would not fit many of the nouns that are associated with it, and the association of many nouns with this class would thus be semantically opaque.

- (16) Emergence of opaque class markers through repeaters in Miraña
- a. *tša-muu ku:muu*
 one-RP signal_drum
 “one signal drum”
- b. *tša-muu muhpajne*
 one-RP breast
 “one breast”
- c. *tša-muu ni:muu*
 one-RP umari_fruit
 “one umari (species of fruit)”

The process just described is similar to a particular way of integrating loanwords by which (initial or final, according to the systems) syllables of loanwords are reanalyzed as noun class markers. This can be observed in the Swahili Examples (17a)–(b), in which initial syllables of Arabic and Japanese loanwords have been reanalyzed as the Class 7 prefix *ki-*. According to the Swahili noun class system, these nouns are inflected for plural with class marker *vi-*.

- (17) *Reanalysis of initial syllable as class marker in Swahili*
- a. *ki-tabu* (PL. *vi-tabu*) < Arabic
“book”
 - b. *ki-mono* (PL. *vi-mono*) < Japanese
“kimono”

In sum, this section has shown that opacity in nominal classification may be a result of different processes in addition to semantically motivated extensions. Among these are semantic changes in the classified nouns, related to historical change in their typical referents, which results in greater distance between the descriptive content of the classifying morpheme and the semantics of the noun. Opacity may also arise when phonological processes intervene in the noun class assignment. This is the case when classifying morphemes become homophonous due to phonological reduction and when phonological segments of loanwords are reinterpreted as class-marking morphology. In both cases, the results are semantically less transparent subdivisions of the nominal lexicon, a different kind of semantic reduction.

4.3 Replacement of opaque classification by transparent classification

While the previous section discussed diachronic processes from transparent to opaque classification, we now turn to a process in the opposite direction, from opaque to transparent classification. This process has been reported for various languages. It will be argued that that opaque classification has an inherent tendency to be replaced by semantically transparent classification. The argument is based on the restricted distribution of opaque classification, when compared to transparent classification, in three constructions: (i) contrastive focus constructions, (ii) reference tracking after longer stretches of discourse, and (iii) in agreement on certain targets. The focus in this section is on the replacement of noun class or gender assignment, in particular in the context of agreement marking and reference tracking.

I begin by giving two examples of diachronic processes of the re-classification of nouns with opaque class assignment. First, in Miraña, animal names (faunal nouns) are assigned to shape-based ‘specific’ noun classes. These nouns are semantically opaque, i.e., they fail the test given above. The class assignment is evident from classifier forms that are recognizable as suffixes on the nouns, e.g., *kuú:muu-hi* (turtle-CL:2DIM.ROUND) “turtle”. However, these classes are used on noun class agreement with such nouns only in traditional songs (18), which instantiate archaic speech (also in other respects). In contrast, in spontaneously produced contemporary Miraña only animate ‘general’ class markers are used for agreement marking with such nouns on any target (19). This is a clear example of the replacement of opaque classification by transparent classification.

- (18) *Opaque class with animal name in Miraña traditional song*
mamáβe-hĩ=pe *kuú:muu-hĩ*
 learn-CL:2DIM.ROUND=PST turtle-CL:2DIM.ROUND
 “He learned, the turtle.”
- (19) *Re-classification with transparent class in contemporary Miraña*
- a. *aj:-di* *muúhuu-:be* *kuú:muu-hĩ*
 DIST-CL:M.SG be.big.SUB-CL:M.SG turtle-CL:2DIM.ROUND
 “that big turtle”
- b. *aj:-di* *muúhuu-:be* *ni:muú-ko*
 DIST-CL:M.SG be.big.SUB-CL:M.SG bird.sp-CL:1DIM.POINTE
 “that big bird (genus *Crax*)”

The second example comes from Swahili, which, like many other Bantu and Niger-Congo languages, has a complex noun class system. Most noun classes involve a considerable degree of opacity, although more or less distant semantic motivations can be discerned for most of them (Contini-Morava 1994, 1997). Class 1 (and its plural counterpart Class 2) are different in that they are semantically transparent, i.e., they are used almost exclusively for humans. Olstad (2011) has quantitatively substantiated the intuition that some Swahili classes (among them Class 5) are more opaque than others, especially Class 1/2. So-called ‘alliterative agreement’ (agreement marking on multiple targets by the same marker that is also present on the noun) is a pervasive and presumably relatively old pattern in Bantu (and some other Niger-Congo languages). However, in a number of modern Bantu languages, agreement with nouns that have semantically opaque assignment is now optionally or even preferably marked with semantically transparent noun classes. This is illustrated in Examples (20a)–(b) where the noun *zee* “old man”, which is lexically assigned to Class 5 (in a semantically opaque way), can now also be associated with the semantically transparent human Class 1/2 for agreement marking.

- (20) *Re-classification of animates with transparent class in Swahili (Bantu)*
 (Heine 1982: 195; see also Wurzel 1986: 84; Katamba 2003: 113)
 (Class 1 is human class)
- a. *zee* *yu-le*
 old_man(CLASS5) CLASS1-that
 “that old man”
- b. *zee* *li-le*
 old_man(CLASS5) CLASS5-that
 “that funny/extraordinary/extremely old man”

Heine (1982: 195) notes that the distribution of the opaque class with respect to (20) is restricted when compared to the semantically transparent agreement marking: “semantic agreement is unmarked as opposed to automatic agreement, which is

marked ... By means of the automatic agreement, the meaning of the derivative noun class is emphasized and/or modified.”

A similar process can be observed in the Bantu language Lingala (Example (21)). In this case, it is inanimate nouns which are re-classified as belonging to the (overwhelmingly inanimate) Class 7. To the extent that Class 7 is (or is becoming) a semantically transparent class for inanimates, this is another case of a re-classification from opaque to transparent. The re-classification is occurring, according to Heine (1982: 195), in the spoken language and in particular in the progressive urban varieties of Kinshasa.

- (21) *Alliterative noun class agreement in conservative Mankandza Lingala and Re-classification of inanimates with transparent inanimate Class 7 in progressive Kinshasa Lingala (Bantu)* (Bokamba 1977: 187–188; see also Aikhenvald 2000: 400)

- a. *mu-nkanda* *mu-ko-kweya*
 CLASS3-book/letter CLASS3-TAM-fall
 “A/the book will fall down.”
- b. *mu-nkanda* *e-ko-kweya*
 CLASS3-book/letter CLASS7-TAM-fall
 “A/the book will fall down.”

Why is there a recurrent tendency to replace opaque assignment with transparent assignment – particularly in certain constructions? I would like to suggest here that opaque assignment has distributional restrictions in certain linguistic contexts (when compared with semantically transparent assignment) and that these restrictions may lead to the gradual replacement of opaque assignment with transparent assignment.

The first context in which the use of opaque assignment is restricted is in pronouns under contrastive focus. This is discussed below with examples, first from German, which has semantically transparent gender assignment of animates and semantically opaque assignment of inanimates (masculine, feminine, neuter). Although semantic principles can be detected within gender assignment of inanimates (see §4.1), this is clearly of a different nature than the semantically highly transparent assignment of animate, especially human nouns (masculine vs. feminine, based on natural gender). Bosch (1988: 224–225) has observed that gender-marked pronouns cannot be used with contrastive focus to anaphorically refer to inanimate participants in German (22a), all of which have opaque assignment. However this is perfectly acceptable for gender-marked pronouns with animate antecedents, all of which are transparently assigned to a gender (22b). As Bosch (1988: 225) observed, contrastive focus requires expressions with a descriptive content. I may add here that it requires descriptive content that matches the semantics of the antecedent noun.

- (22) *Opaque assignment and contrastive focus in German* (Bosch 1988: 224–225)
- a. *Wenn du die Mutter(F) von dem Bolzen(M) lösen willst, dann musst du ??IHN(M) festhalten und ??SIE(F) nach rechts drehen.*
 “If you want to loosen the nut from the bolt, you must hold ??IT (pronoun marked for masculine, intended reference to bolt) and turn ??IT (pronoun marked for feminine, intended reference to nut) to the right.”
- b. *Der Mann(M) hatte Streit mit seiner Freundin(F), weil SIE(F) noch in eine andere Kneipe gehen wollte, aber ER(M) keine Lust mehr hatte.*
 “The man had a row with his girl friend, because SHE (pronoun marked for feminine, intended reference to girl friend) wanted to go on to another pub but HE (pronoun marked for masculine, intended reference to man) didn’t feel like it any more.”

The same restriction can be observed in Miraña, where there is semantically transparent assignment of some inanimate nouns but opaque assignment of other inanimate nouns and to nouns denoting animals (see above). As in German, the use of a pronoun with contrastive focus is not acceptable (or less acceptable) if the assignment is opaque (23a), but it is perfectly acceptable if the assignment is transparent (23b)

- (23) *Opaque assignment and contrastive focus in Miraña*
- a. *tsáihu-u nirmú-ko-o tohpá-u*
 once-REM currassow-CL:1DIM.POINTEd-and pigeon-CL:3DIM.ROUND
úménéba ʔadzú-ri á:-muútsi-dí-ʔú-u
 log top-LOC CONN-CL:M.DU-ANIM-ABL-REM
 ??té-*u-re* á:kité-ʔi
 PN-CL:3DIM.ROUND-FOC fall-PRED
 “A Currassow (family *Cracidae*) and a pigeon (species *Crypturellus cinereus*) were sitting on a log. Of these two, ??IT (pronoun marked for CL:3DIM.ROUND-class, intended reference to pigeon) fell down.”
- b. *tsáihu-u úhi-ʔo-o kúni-u*
 once-REM banana-CL:3DIM.OBLONG-and potato-CL:3DIM.ROUND
mérsá ʔadzú-ri á:-né-ku-tú-u
 table top-LOC CONN-CL:INAN-DU-ABL-REM
 té-*u-re* á:kité-ʔi
 PN-CL:3DIM.ROUND-FOC fall-PRED
 “A banana and a potato were sitting on a table. Of these two, IT (pronoun marked for CL:3DIM.ROUND-class, intended reference to potato) fell down.”

A second context where semantically opaque assignment is disfavored is anaphoric reference if the anaphoric expression is separated by longer stretches of discourse

from its antecedent. It has been observed – also for German – that third person pronouns can be used only with difficulty after longer stretches of discourse if the assignment of the intended antecedent is semantically opaque. As Comrie (1994: 4) puts it,

[I]n German, for instance, use of a masculine singular pronoun to refer across a long stretch of text to an inanimate antecedent of masculine gender is likely to cause bewilderment rather than retrieval of the appropriate referent, even if there are no intervening masculine singular referents.

It is the opaque assignment of inanimates, and not the animacy distinction itself, that is responsible for this restriction. This can be shown by comparison with Miraña, which has semantically transparent assignment of some inanimate nouns, as just mentioned. In Miraña texts, it is perfectly normal and common to use class-marked third person pronouns anaphorically to refer back to inanimate antecedents (Seifart 2005: 300–304). For example, *tɛː-huu* (PN-CL:TUBE) ‘it (tubular)’ is used in one text to refer back to *ajnú-huu* (shoot.NMLZ-CL:TUBE) ‘rifle’ minutes after its last mention with a full noun phrase and after a number of other intervening inanimate participants. This is possible because of the semantically transparent assignment, i.e., the matching of the descriptive content of the noun class morphology with the semantics of the classified noun.

Contrastive focus constructions and reference tracking over longer stretches of discourse may be among the first contexts where semantically transparent assignment is used in addition to opaque assignment, introducing alternative class membership for a given noun. Which other contexts may be affected by this gradual replacement can be captured with the Agreement Hierarchy (Corbett 1991: 225–230), which describes a restricted distribution of opaque classification if there is a choice between an opaque class and a transparent one for agreement marking. According to the Agreement Hierarchy, opaque class agreement (‘syntactic agreement’ in Corbett’s 1991 terms) is increasingly unlikely towards the top of the Agreement Hierarchy (Figure 4). Note that this hierarchy is probably also roughly congruent with the linear distance of the classifying element (or agreement target) from the classified noun (or head noun) in most cases.

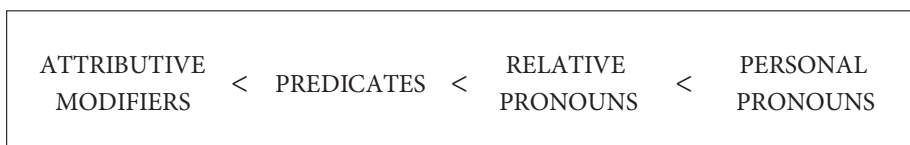


Figure 4. The Agreement Hierarchy (Corbett 1991: 225–230)

The diachronic hypothesis derived from the observations above in combination with the Agreement Hierarchy is that opaque assignment is gradually replaced by transparent assignment along the positions of the Agreement Hierarchy. Such replacement would thus first take place in personal pronouns – and in this position, perhaps first when these are used in contrastive focus and reference tracking over longer distances. Next, relative pronouns, predicates, and then attributive modifiers would be affected. An illustration of this process in Irish is given in Frenda (this volume). The endpoint of such a process may be a complete obsolescence of opaque classification, the only remnants of which may be fossilized class morphology on nouns, as in Miraña faunal nouns (see Wurzel 1986 on the repeated decay and renewal of nominal classification).

This section has discussed a number of pieces of evidence for a synchronically observable restriction in the use of opaque classification in certain constructions where transparent classification can freely occur. These include contrastive focus, anaphoric reference after longer stretches of discourse, and agreement on different targets. These restrictions are interpreted as a motivation for a diachronic process of re-classification by which opaque assignment is replaced by transparent assignment to a different class.

5. Differential diffusability of nominal classification

The previous sections discussed the role of semantic generality and semantic opacity in the internal development of nominal classification. This final section briefly considers the role of semantic opacity in contact-induced diachronic changes of nominal classification, in particular the differential diffusability of nominal classification systems of different types (for a morphosyntactic typology of nominal classification, see Aikhenvald 2000; Grinevald 2000). By ‘diffusion’ I mean the convergence of structures under the influence of contact, usually without the transfer of forms (i.e., loan words or loan morphemes).

Opaque assignment is a characteristic typically associated with particular types of nominal classification systems, namely with strongly grammaticalized systems such as small, obligatory noun class and gender systems. Opaque assignment is much less prevalent in other types of classification systems, i.e., in large and often semi-open classifier systems (in the narrow sense of ‘classifier’). For instance, Aikhenvald (2000: 229) claims that “the choice of a classifier in a multiple classifier language is always semantically based,” and this distinguishes multiple classifiers, as one kind of system of the type ‘classifiers’, from noun classes and genders, as another type of classification system.

In recent years, growing evidence has become available that some types of classification systems are diffused more easily than others. In particular, there is now quantitative evidence that classifiers are more prone to diffusion than genders and noun classes (Nichols 2003: 299–303; Wichmann & Holman 2009: 54–55; see also Nichols 1992: 132). These types are defined as clusters of a number of characteristics, but I suggest here that the difference in semantic opacity between these two types may be responsible for this difference in diffusability. This is consistent with ‘borrowability’ scales (see Wilkins 1996 for a summary) used in language contact research; these predict that elements with a higher degree of grammaticalization are less likely to be borrowed (although these scales usually refer to forms, not structures). Accordingly, classification systems that retain lexical semantic characteristics, i.e., descriptive content and semantically transparent assignment, are more likely to be diffused than systems with semantically opaque assignment, which is more closely associated with grammatical systems. The recruitment of classifier forms from native nouns, a requirement of structural convergence without borrowing of form, may also be facilitated by lexical-like meaning rather than semantically opaque forms.

6. Summary and conclusion

This paper has discussed various diachronic implications of semantic generality and semantic opacity, two key features of nominal classification. First, the pragmatics of semantically general forms in reference tracking provides a possible motivation for developing classification in the first place. Second, the possible emergence of opaque classification was discussed, and an account was suggested of how and why opaque assignment could be replaced by transparent assignment. Finally, the role of opacity (as closely associated with strong grammaticalization) in the diffusability of nominal classification was discussed. These issues show that semantic generality and semantic opacity may play an important role in the diachronic development of nominal classification systems.

Acknowledgments

I am grateful to Søren Wichmann and Bill McGregor as well as two anonymous reviewers for helpful comments on earlier versions of this paper. I am also grateful for funding from the Max Planck Institute for Evolutionary Anthropology’s Department of Linguistics headed by Bernard Comrie.

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith. 1977. "Classifiers". *Language* 53.283–310. doi:10.1353/lan.1977.0043
- Ariel, Mira. 1988. "Referring and Accessibility". *Journal of Linguistics* 24.65–87. doi:10.1017/S0022226700011567
- Barlow, Michael. 1992. *A Situated Theory of Agreement*. New York: Garland.
- Bokamba, Eyamba G. 1977. "The Impact of Multilingualism on Language Structures: The case of Central Africa". *Anthropological Linguistics* 19.181–202.
- Bosch, Peter. 1988. "Representing and Accessing Focussed Referents". *Language and Cognitive Processes* 3.207–231. doi:10.1080/01690968808402088
- Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Comrie, Bernard. 1994. "Towards a Typology of Reference-Tracking Devices". Handout for paper presented at the International Symposium on Language Typology, University of Tsukuba, January 1994.
- Contini-Morava, Ellen. 1994. *Noun Classification in Swahili (= Publications of the Institute for Advanced Technology in the Humanities, Research Reports, 2nd Series)*. Charlottesville, VA: University of Virginia. <http://www2.iath.virginia.edu/swahili/swahili.html>. (Accessed: 1 March 2014.)
- Contini-Morava, Ellen. 1997. "Noun Classification in Swahili: A cognitive-semantic analysis using a computer database". *African Linguistics at the Crossroads*. Papers from Kwaluseni, 1st World Congress of African Linguistics, Swaziland, 18–22. VII. 1994 ed. by Robert K. Herbert, 599–628. Köln: Rüdiger Köppe.
- Contini-Morava, Ellen & Marcin Kilarski. 2013. "Functions of Nominal Classification". *Language Sciences* 40.263–299. doi:10.1016/j.langsci.2013.03.002
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166119
- Craig, Colette G. 1986. "Jaltec Noun Classifiers: A study in grammaticalization". *Lingua* 70: 4.241–284. doi:10.1016/0024-3841(86)90046-X
- Croft, William. 1994. "Semantic Universals in Classifier Systems". *Word* 45.145–171. doi:10.1080/00437956.1994.11435922
- Daley, Karen Ann C. 1998. *The Use of Classifiers in Vietnamese Narrative Texts*. Arlington, Texas: Summer Institute of Linguistics and The University of Texas at Arlington.
- Downing, Pamela. 1996. *Numeral Classifier Systems: The case of Japanese*. Amsterdam: John Benjamins. doi:10.1075/sidag.4
- Erbaugh, Mary S. 1986. "Taking Stock: The development of Chinese noun classifiers historically and in young children". *Noun Classes and Categorization* ed. by Colette G. Craig, 399–436. Amsterdam: John Benjamins. doi:10.1075/tsl.7.25erb
- Fox, Barbara. 1987. *Discourse Structure and Anaphora: Written and conversational English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627767
- Givón, Talmy. 1976. "Topic, Pronoun, and Grammatical Gender." *Subject and Topic* ed. by Charles N. Li, 149–188. New York: Academic Press
- Givón, Talmy. 1983a. "Topic Continuity in Discourse: An introduction." Givón, ed. 1983b.1–41

- Givón, Talmy, ed. 1983b. *Topic Continuity in Discourse: A quantitative cross-language study*. Amsterdam: John Benjamins. doi:10.1075/tsl.3
- Grinevald, Colette. 2000. "A Morphosyntactic Typology of Classifiers." *Systems of Nominal Classification* ed. by Gunter Senft, 50–92. Cambridge: Cambridge University Press
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69.274–307. doi:10.2307/416535
- Heine, Bernd. 1982. "African Noun Class Systems." *Apprehension: Das sprachliche Erfassen von Gegenständen, Vol. I: Bereich und Ordnung der Phänomene* ed. by Hansjakob Seiler & Christian Lehmann (= Language Universal Series), 189–216. Tübingen: Gunter Narr
- Katamba, Francis. 2003. "Bantu Nominal Morphology." *The Bantu Languages* ed. by Derek Nurse & Gérard Philippson, 103–120. London: Routledge.
- Lakoff, George. 1986. "Classifiers as a Reflection of Mind." *Noun Classes and Categorization* ed. by Colette Craig, 13–52. Amsterdam: John Benjamins doi:10.1075/tsl.7.04lak
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What categories reveal about the mind*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001
- Lehmann, Christian. 1988. "On the Function of Agreement." *Agreement in Natural Languages* ed. by Michael Barlow & Charles A. Ferguson, 55–65. Stanford: Center for the Study of Language and Information.
- Levinson, Stephen C. 2000. *Presumptive Meaning: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
- McGregor, William B. 2002. *Verb Classification in Australian Languages*. Berlin: Mouton de Gruyter. doi:10.1515/9783110870879
- Mithun, Marianne. 1984. "The Evolution of Noun Incorporation." *Language* 60.847–894. doi:10.1353/lan.1984.0038
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226580593.001.0001
- Nichols, Johanna. 2003. "Diversity and Stability in Language." *The Handbook of Historical Linguistics* ed. by Brain D. Joseph & Richard D. Janda, 283–310. Malden, MA: Blackwell. doi:10.1002/9780470756393.ch5
- Nida, Eugene. 1975. *Componential Analysis of Meaning: An introduction to semantic structures*. The Hague: Mouton.
- Olstad, John. 2011. "Noun Class Assignment in Swahili via Bayesian Probability." *On Words and Sounds: A selection of papers from the 40th PLM, 2009* ed. by Kamila Dębowska-Kozłowska & Katarzyna Dziubalska-Kołaczyk, 180–194. Newcastle: Cambridge Scholars..
- Payne, Thomas E. 1988. "Referential Distance and Discourse Structure in Yagua." *Studies in Language* 12.345–392. doi:10.1075/sl.12.2.04pay
- Seifart, Frank. 2005. *The Structure and Use of Shape-Based Noun Classes in Miraña (North West Amazon)* (= MPI Series in Psycholinguistics, 33). Nijmegen: Max Planck Institute for Psycholinguistics. <http://hdl.handle.net/2066/26990>. (Accessed: 1 March 2014.)
- Seifart, Frank. 2010. "Nominal Classification." *Language & Linguistics Compass* 4:8.719–736. doi:10.1111/j.1749-818X.2010.00194.x
- Seiler, Hansjakob. 1986. *Apprehension: Language, object, and order. Part III: The universal dimension of apprehension*. Tübingen: Gunter Narr.
- Wichmann, Søren & Eric W. Holman. 2009. *Temporal Stability of Linguistic Typological Features* (= LINCOM Studies in Theoretical Linguistics, 42). Munich: LINCOM EUROPA.

- Wiebusch, Thekla. 2009. "The Diachronic Development of Chinese Numeral Classifiers." Paper presented at The Diachrony of Classification Systems, Netherlands Institute for Advanced Studies in the Humanities and Social Sciences, March 2009.
- Wilkins, David P. 1996. "Morphology". *Contact Linguistics: An international handbook of contemporary research* ed. by Hans Goebel, Peter H. Nelde, Zdenek Starý & Wolfgang Wölck, Vol. I, 109–117. Berlin: Mouton de Gruyter.
- Wurzel, Wolfgang Ullrich. 1986. "Die wiederholte Klassifikation von Substantiven: Zur Entstehung von Deklinationsklassen". *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39:1.76–96.
- Zubin, David & Klaus Michael Köpke. 1986. "Gender and Folk Taxonomy: The indexical relation between grammatical and lexical categorization". *Noun Classes and Categorization* ed. by Colette Craig, 139–180. Amsterdam: John Benjamins. doi:10.1075/tsl.7.12zub

Niger-Congo numeral classifiers in a diachronic perspective

Roland Kießling
University of Hamburg

Numeral classifier systems have only recently come to be recognized in various African languages where they either co-exist with fully fledged noun class systems or supersede residual ones. This chapter explores the semantic and morphosyntactic properties of Niger-Congo numeral classifier systems in a typological and a diachronic perspective. Due to an incipient stage of development, most of these systems are fairly transparent etymologically. With respect to lexical source concepts, the classifier items originate in nouns for concrete objects such as body parts or in basic level terms, most of which relate to the botanical domain. Syntactically, the emergent classifier constructions provide counter-evidence to current generalizations in that they separate the classifier from the numeral, allowing for morphophonological fusion of classifier and enumerated noun rather than classifier and numeral. These syntactic properties reflect constituency relations directly inherited from associative predecessor constructions.

1. Introduction

Numeral classifier systems have been believed to be absent in African languages for a long time. Seminal works such as Craig (1986) and general overviews such as Dixon (1986) and Craig (1994) remain silent on African numeral classifier systems. Even since Suano Ikoró's (1994) extensive description of a full-fledged classifier system in the Cross River language Kana, numeral classifier systems are still regarded as very rare and fairly exotic in Africa, as testified by their drastic underrepresentation in the *World Atlas of Language Structures* (Gil 2005) and Aikhenvald (2000: 98–124). According to a widely shared view, this underrepresentation is largely due to the predominance of functionally complementary grammaticalized noun class systems in large parts of Sub-Saharan Africa, e.g. in most branches of Niger-Congo, in Afroasiatic, in Eastern Nilotic and in Khoisan. However, upon

closer inspection of Niger-Congo languages in general and Benue-Congo languages of the Nigerian-Cameroonian Bantu borderland in particular, incipient numeral classifier systems of a restricted type could be found to prevail. This contribution explores the semantic and morphosyntactic properties of the Niger-Congo numeral classifier systems in a typological and a diachronic perspective.

Section 2 presents an overview of the numeral classifier systems which have been found so far in Niger-Congo languages. Due to an incipient stage of development, most of these classifier systems are fairly transparent etymologically, providing an excellent laboratory for deriving generalizations pertaining to the lexical sources of numeral classifiers (Section 3), the syntactic sources of classifier constructions (Section 4) and the processes of grammaticalization (Section 5). Puzzles concerning potential explanations of their distribution, their diversity and motivations of their emergence are discussed in Section 6. Section 7 summarizes the findings and points out their significance for comparative Niger-Congo.

2. The numeral classifier systems in Niger-Congo

The *World Atlas of Language Structures* (WALS) presents Africa roughly as a white spot on the map classifier-wise. The hotbed of numeral classifiers in South and East Asia contrasts sharply with just four dots representing attestations of numeral classifier systems across the whole continent of Africa (Gil 2005: 228): Kana (Benue-Congo), Toussian (Gur), Boko (Mande), So (Kuliak).

At the same time, typological comparisons such as Grinevald & Seifart (2004) tend to contrast 'African' vs. Amazonian systems of nominal classification in which features of the Bantu noun class system (Katamba 2003) are taken to represent the 'African' pole. While it is certainly true that fully grammaticalized noun class systems characterize the vast majority of Bantu languages in Africa, this is only true to a much lesser extent for Niger-Congo (Williamson & Blench 2000). The tendency to invoke prototypical Bantu as a reference point and model of Niger-Congo in general rather reflects the longstanding Bantu-focussed research tradition into African noun class systems.

Beside the noun class systems in Niger-Congo, there is increasing evidence of emergent numeral classifier systems in many more Niger-Congo¹ languages than

1. As will become clear, the frame 'Niger-Congo' is taken in a very broad sense, genetically speaking. Numeral classifiers discussed here are restricted to three distinct families in West and Western Central Africa, i.e. Benue-Congo, Gur and Mande. Taking Greenberg (1966) as a reference point, all three families are included in his model of Niger-Congo as a phylum.

represented in *WALS*, which will entail a more differentiated picture of nominal classification in Africa in general. Table 1 provides a contrastive overview of the numeral classifier² systems found in Niger-Congo languages so far, arranged according to major typological parameters such as lexical coverage, semantic categories of classification, morphosyntactic properties of the classifier constructions and the presence of a coexistent and competing noun class system. Except for Kana, the study of these numeral classifier systems is in its infancy. Many details of their morphosyntax and their semantics remain to be worked out, since most of the available descriptions barely reach beyond the level of preliminary sketches. Yet, they are definitely sufficient to provide a reliable impression of the basic properties and the outlines of the numeral classifier systems under discussion.

However, views on the genetic status of Niger-Congo are controversial. They range from his conception which includes six coordinate branches – Atlantic, Mande, Gur, Kwa, Benue-Congo, Adamawa – over reorganizations which provide much more fine-grained modifications of the internal historical relations of branches and subgroupings (Williamson & Blench 2000) to a reconception which rejects the genetic unity of Niger-Congo altogether, isolating individual families such as Mande as separate genetic units (Dimmendaal 2011) which rather form part of a larger areal convergence zone which has come to be called Macro-Sudan (Güldemann 2008). Without intending to subscribe to Greenberg’s model of the phylum, I take Niger-Congo as a handy cover term in order to include phenomena which occur in the three families mentioned – Benue-Congo, Gur, Mande – for purely practical reasons. Whether Mande is genetically linked to the rest of Niger-Congo or not is not crucial to the argument here, since all instances of numeral classifiers rather represent fairly recent innovations. Thus, irrespective of the genetic affiliation of Mande, the question of contact-induced innovation will have to be taken into account seriously even when it comes to determine the emergence of numeral classifiers in disproportionately closer related subgroups within Benue-Congo such as Nyang and Ekoid. It goes without saying that it is highly desirable to check various other families both within Niger-Congo and outside – especially those which have demonstrably been in massive contact with Benue-Congo such as Chadic – for numeral classifiers or functionally related phenomena. However, as nothing comparable has been reported so far for language families such as Chadic, nor other branches of Niger-Congo such as Adamawa, Kru and Kwa, nothing definitive could be said on the limits of the distribution of numeral classifiers, since much more descriptive groundwork needs to be done here in order to achieve a full picture. Consequently, statements about the possible motivations for the independent emergence of numeral classifiers, their diversity and the role of culture and mind in this context would be much too premature here and should await a fuller account of the details in more languages.

2. Other types of classifiers such as possessive classifiers in Ubangi (Pasch 1985, 1986) and classifiers relating to the distinction of inalienable vs. alienable possession, e.g. in Mande (Welmers 1973), are not dealt with here.

Table 1. Typological parameters of Niger-Congo numeral classifier systems

Language	Classification	Restriction type	Semantic parameters	Construction type	Competing noun class system	Source
Bafanji	Benue-Congo	lexical	shape, size, aggregation	NUM [CLF N]	reduced	Hamm (2010)
Bali-Mungaka	Benue-Congo	lexical	shape	[CLF N] NUM	reduced	Tischhauser & Stöckle (1993)
Busa	Mande	none	animacy, shape, number	N [CLF NUM]	no	Wedekind (1972), Jones (1998)
Bwamu	Gur	?	animacy	N [CLF NUM]	reduced	Manessy (1961)
Denya	Benue-Congo	lexical	shape, size, aggregation	[CLF N] NUM	yes	Fieldnotes by Abangma & Kießling
Ejagham	Benue-Congo	lexical	shape, size, aggregation	[CLF N] NUM	yes	Watters (1981)
Gban	Mande	?	animacy, shape, size	N [CLF NUM]	no	Zheldtov (p.c.)
Ibibio	Benue-Congo	lexical (probably)	shape	[CLF N] NUM, [CLF NUM] N	no	Kaufman (1972)
Isu	Benue-Congo	lexical + numerical	shape	[CLF N] NUM	yes	Fieldnotes by Kießling
Kana	Benue-Congo	none	shape, size, animacy, function, partition, aggregation	NUM [CLF N]	no	Ikoro (1994, 1996)
Ngiemboon	Benue-Congo	lexical (probably)	shape, size	N [CLF NUM]	yes	Vinogradov (2009)
Ngwe	Benue-Congo	lexical	shape, aggregation, stages of processing	N [CLF NUM]	yes	Lovegren (2010)
Obolo	Benue-Congo	lexical (probably)	animacy	N [CLF NUM]	no	Rowland-Oke (2003)
Pana	Gur	?	animacy	N [CLF NUM]	no	Prost (1969)
Toussian	Gur	none	animacy, shape, number	N [CLF NUM]	no	Zaugg-Coretti (2005), Mous (p.c.)
Yemba	Benue-Congo	lexical (probably)	shape	N [CLF NUM]	yes	Harro & Haynes (1991)

Categories of classification observed in the numeral classifier systems in Table 1 conform to what has been identified as universal semantic properties of classifiers (Allan 1977: 297; Craig 1994: 567; Aikhenvald 2000: 286–293). Enumerated objects are differentiated for their material or essence (human vs. animate vs. inanimate), their shape (saliently one-dimensional long shape vs. two-dimensional flat shape vs. three-dimensional round shape), their size (small vs. big) and their arrangement or aggregation (cluster vs. bunch vs. bundle vs. lump). Among the prime shapes, the most salient category which seems to emerge first as a distinct category is the three-dimensional round shape which originates in lexical meanings such as ‘eye’, ‘grain’ or ‘fruit’.

Typologically, the numeral classifier systems in Table 1 fall into three distinct categories which may reflect successive stages in their emergence. The first category includes numeral classifier systems with almost full lexical coverage, i.e. without lexical restriction. The second category presents systems in which the obligatory presence of a numeral classifier is restricted to certain semantic domains such as objects of round vs. oblong shape small enough to hold in a hand and thus displays restricted lexical coverage. The third category presents the least developed systems which are characterized by restricted lexical plus numerical coverage, i.e. the occurrence of a numeral classifier is restricted both to certain semantic domains and to low numerals. The second restriction is caused by the fact that the classifier conflates haptic notions with notions such as contra-expected deficiency in number.

While only three language families are represented, i.e. Gur, Mande and Benue-Congo, the numeral classifier systems with almost full lexical coverage are found exclusively in those branches which lack a coexistent noun class system, either because a prior noun class system has completely broken down, i.e. in Kana of the Kogoid languages of the Cross River branch of Benue-Congo and in Toussian and some other languages of the Gur family, or because there might never have been a prior noun class system at all, i.e. in Busa and some closely related Mande languages of the Eastern branch. It is also precisely these languages which include animacy as a central semantic feature and, to a lesser extent, number.

Both features are absent from the classifier systems with restricted lexical (and numerical) coverage, which have been found in various subgroups of Benue-Congo, i.e. Ekoid Bantu (Ejagham), Nyang (Denya) and Grassfields Bantu (Bali-Mungaka, Bafanji, Ngwe, Ngiemboon, Yemba and Isu). Remarkably, these systems share a semantic domain: products of plants, i.e. fruits, grains and tubers. Rather than being based on animacy, they are predominantly organized by criteria of shape, size and aggregation and basically differentiate globular vs. oblong objects small enough to hold in the hand. Systems such as these emerge in languages which retain, to varying degrees, a prior noun class system, allowing for a coexistence of two systems of nominal classification: the older fully grammaticalized noun class system inherited

from Proto-Benue-Congo or Proto-Niger-Congo and the newly emerging numeral classifier system. Animacy and number are absent as categories in these classifier systems, since they are taken care of by the competing noun class system.

The fact that the existence of these numeral classifier systems has largely gone unnoticed so far is due to three interrelated factors. First, the majority of Niger-Congo languages still lack adequate documentation and analysis. Second, due to the emergent nature of most of these numeral classifiers systems, at times in competition with coexistent noun class systems, their lexical coverage is often not complete. So they might easily escape the attention of researchers who aim at a standard first grammatical sketch. Finally, as argued in Grinevald & Seifart (2004), the early recognition of Bantu noun class systems has restricted the research paradigm to identification of variations of this system in the modern languages, resulting in the dominance of a biased descriptive practice which distracts attention from parallel systems of classification which might coexist with the well-established one. However, with a more fine-grained research agenda, many more of these emergent systems may show up in the future.

3. Lexical sources and cognitive models of classifiers

Due to the incipient stage of development, the etymology of most Niger-Congo classifiers is either transparent or could be retrieved easily by internal reconstruction, as detailed in Tables 2–3, 11, 17, 21, which list lexical sources and semantic ranges of numeral classifiers, specified for their attestation in Niger-Congo. The general overview in Table 2 reveals a predominant application of cognitive models based on body parts and generic or basic level concepts, most of which originate in the botanical domain. To a lesser extent concepts of aggregation and partition are exploited.

Table 2. Cognitive models of numeral classifiers in Niger-Congo

Lexical sources of classifiers	
Basic level terms:	'person', 'people', 'child', 'mother', 'thing'
Botanical terms:	'tree (trunk)', 'fruit', 'grain', 'stock', 'pod', 'leaf', 'stick', 'palm kernel'
Body part nouns:	'body', 'eye', 'hand', 'breast', 'finger', 'skin', 'head'
Terms of aggregation and partition:	'bunch', 'bundle', 'heap', 'lump (of clay)', 'bottom', 'part', 'piece', 'half'

Table 3. Body part nouns as sources of numeral classifiers

Lexical source of classifier	Semantic range of classifier and attestation
'body'	inanimates (Busa)
'breast'	bunches of objects which are recognizably subdivided into clusters (Isu)
'eye'	small compact globular objects (Denya, Bali-Mungaka, Busa) spot- or dot-like objects (Kana) objects with curved shapes (Busa, Bali-Mungaka, Bafanji) measurements (Toussian) contra-expectual deficiency in number (Isu, Men)
'finger'	fruits of oblong shape (Denya)
'hand'	long objects in a cluster (Isu) long parts of plants (Ngiemboon)
'head'	big globular fruits (Ngiemboon) contra-expectual deficiency in number with round flat objects (Isu)
'skin'	flat objects (Kana)

3.1 Body parts

The application of body part concepts is attested by numeral classifiers which are derived from body part nouns such as 'eye', 'hand', 'head', 'finger', 'breast, udder', 'skin' and the noun for 'body' itself (see Table 3).

Among these, the noun 'eye' stands out particularly as the most frequent across languages from various branches of Niger-Congo. It is attested in Grassfields Bantu languages such as Isu (1g), Men, Bali-Mungaka (1e) and Bafanji (1f), beyond these within Benue-Congo in Denya (1c)–(d) and Kana (1a), in Gur languages such as Toussian (1b) and in East Mande languages such as Busa.

(1) Source item 'eye' as classifier

- a. *Kana* (Ikoro 1996: 96): *děě* "eye"
zìì děě múù
 one CLF:eye injury
 "one injury"
- b. *Toussian* (Zaugg-Coretti 2005: 25): *yí* "eye" (PL. *íyò*)
mé tǎn mè wè báá yí nèn
 1SG shadow DEM:SG see limit:SG CLF:eye one
 "I have seen this film one time."

- c. *Denya*: classifier *ɲɲi* (5/2) ‘eye’ (PL. *àmé*)³
ɲi-ɲi né ɲjègésè né-mâ
 5-CLF:eye ASS.5 groundnuts 5-one
 ‘one groundnut’
- d. **ɲi-ɲi né gè-sòó né-mâ*
 5-CLF:eye ASS.5 7-sand 5-one
 ‘one grain of sand’
- e. *Bali-Mungaka* (Kähler-Meyer 1941/42: 142): *líɪ* ‘eye’ (PL. *míɪ*)
bó yá mǝʔ líɪ nsú ɲgʰ
 3PL see one CLF:eye fish big
 ‘Then they saw a single big fish.’
- f. *Bafanji* (Hamm 2010): *ndí* (9/6) ‘eye’
tii mǝ-ndi mbigi
 three 6-CLF:eye mosquito
 ‘three mosquitos’
- g. *Isu* (Kießling, fieldnotes): *ísí* (5/6) ‘eye’
wǝ fǝn ká ná ísí í^ltsíy yíy í mǝʔ
 2SG return only with 5.CLF:eye ASS.5:5.fish 5:OF ASS.5 one
 ‘You have returned with only a single fish (when you should have brought more).’
- h. *Cameroonian Pidgin English* (Kießling, fieldnotes):
no one eye mosquito
 NEG one CLF:eye mosquito
 ‘not even a single mosquito’

The saliency of this model is highlighted by the fact that it has radiated beyond Benue-Congo languages of Cameroon into the most widespread lingua franca, Cameroonian Pidgin English (1h). In most cases, the noun ‘eye’ is expanded to refer to small globular objects such as grains, seeds, kernels and nuts. In *Denya* it is restricted to objects of organic origin, as attested by the ungrammaticality of (1d). *Isu*, on the other hand, demonstrates a broader usage for any object, irrespective of haptic criteria, constrained only by the notion of contra-expectual deficiency in number (1g). While the cognitive model of ‘eye’ for classification of globular objects is quite widespread, it is definitely not universal. Thus, the Ngiemboon classifier *tɔn* for globular fruits and the Ejagham classifier *Ñ-màgè* (PL. *à-màgè*) for small globular objects are etymologically not transparent any longer, both most probably not derived from the noun meaning ‘eye’, while *Kana* classifies small globular objects

3. Numbers in interlinear glossing refer to person when combined with number indicating glosses such as SG and PL. Otherwise they refer to noun classes following the Bleek & Meinhof conventions for Bantu (Katamba 2003).

such as grains by the classifier *ásúú* ‘grain’ and bigger globular objects such as pears by the classifier *bēē* ‘fruit’.

The lexical source concept ‘hand’ is used in Isu for clusters of long objects such as bananas and plantains (2a). The usage in Ngiemboon seems to be similar in its restriction to oblong parts of plants such as blades of grass (2b). However, the significance of the cluster notion is not clear.

(2) *Source item ‘hand’ as classifier*

a. *Isu: káwó* (7/8) ‘hand’

má má kòꞤ áwó á káfè yiy à bè
 1SG IMM.PST.FOC see 6.hand ASS.6 7.plantain 6:OF ASS.6 two
 “I have seen only two clusters of plantains.”

b. *Ngiemboon: mb'o* ‘hand’

taꞤ mb'o legí
 one CLF:hand grass
 “one blade of grass”

The body part concept ‘breast, udder’ classifies bunches of objects which are recognizably subdivided into clusters, e.g. plantains and bananas in Isu (3). On the other hand, ‘finger’ is used for classifying fruits of oblong shape which may occur in such clusters, e.g. plantains and bananas, or which may occur separately, e.g. maize cobs and bean pods, as in Denya (4). However, it is not the case that oblong fruits are generally classified by a noun meaning ‘finger’, e.g. Ngiemboon uses a classifier *ndɔŋ* for oblong fruits which is etymologically not transparent any longer.

(3) *Source item ‘breast, udder’ as classifier*

Isu: íyé (5/6) ‘breast, udder’

íyé í káfè í mòꞤ í má fwàì
 5.udder ASS.5 7.plantain ASS.5 one SBJ.5 IMM.PST.FOC rot
 “One single bunch of plantains was rotten.”

(4) *Source item ‘finger’ as classifier*

Denya: gèlò (7/8) ‘finger, pod’ (PL. *ùlò*)

ù-lò ú-^hkwá ú-péá
 8-CLF:finger ASS8:8-plantain 8-two
 “two plantains.”

The body part concept ‘head’ is used for varied classifier functions: while Ngiemboon usage seems to be restricted to big globular fruits (5b), Isu usage includes the notion of contra-expected deficiency in number combined with a haptic notion which allows for application to round objects which are flat, e.g. leaves (5a). In Kana, flat objects such as these are instead classified by the body part concept ‘skin’ (6).

- (5) *Source item ‘head’ as classifier*
- a. *Isu: fātúw* (19/6a) “little head”
 túw fífəp fə mɔ̃?
 19.little.head ASS.19:5.leaf ASS.19 one
 “one single leaf only”
- b. *Ngiemboon: tʰò* “head”
 nanasi tʰò lepfuzó
 pineapple CLF:head nine
 “nine pineapples”
- (6) *Source item ‘skin’ as classifier*
Kana: ákpá “skin”
 bàè ákpá àkà
 two CLF:skin biscuit
 “two biscuits”

The source concept ‘body’ is used to refer to inanimate objects generally, e.g. in Busa (7).

- (7) *Source item ‘body’ as classifier* (Wedekind 1972: 167)
Busa: mè “body”
 lí mè ààgɔ̃
 tree CLF:body three
 “three trees”

3.2 Basic level terms

Generic or basic level concepts such as ‘person’ (8), ‘people’ (9) and ‘thing’ (10) are used to classify human singulars vs. human plurals (Busa, Toussian, Gban, Pana, Bwamu) and inanimates (Toussian, Bwamu), respectively, as listed in Table 4.

Table 4. Nouns for basic level concepts as sources of numeral classifiers

Lexical source of classifier	Semantic range of classifier and attestation
‘person’	human singulars (Busa, Gban, Toussian, Win, Pana, Bwamu, Kana (marginal))
‘people’	human plurals (Busa, Toussian, Win)
‘child’	young animates (Kana)
‘mother’	default: entities, animals, things (Kana)
‘thing’	inanimates (Toussian, Bwamu)

- (8) *Source item ‘person’ as classifier*
- a. *Busa: gbē* “person” (Wedekind 1972: 167)
tōfē gbē dō
 hunter CLF:person one
 “one hunter”
- b. *Toussian: náy* “person” (PL. *nǒw*) (Zaugg-Coretti 2005: 22)
lèè náy nèn
 uncle CLF:person one
 “one uncle”
- c. *Pana: la* “person, man” (Prost 1969: 894)
a na balo la-nyi
 1SG see men CLF:human-two
 “I have seen two men.”
- d. *Kana: nēē* “person” (Ikoro 1996: 93)
zìì nēē sāānēē
 one CLF:person guest
 “one guest”
- (9) *Source item ‘people’ as classifier*
- a. *Busa: gǒrǒn* “people” (Wedekind 1972: 167)
ànō gǒrǒn ààgō-mò-nō
 3PL CLF:people three-DEF-PL
 “they, three people”
- b. *Toussian: nǒw* “people” (Zaugg-Coretti 2005: 22–25, 45)
ìkpà nǒw tó
 child.PL CLF:people.PL three
 “three children”
- (10) *Source item ‘thing’ as classifier*
Toussian: kpà “thing” (PL. *yì*) (Zaugg-Coretti 2005: 22–25, 45)
tén-nó yì nìn
 mortar-PL CLF:thing.PL two
 “two mortars”

The source items ‘child’ and ‘mother’⁴ are only attested in Kana. While ‘child’ is used for classifying young animates (11), ‘mother’ has undergone semantic bleaching insofar as it is used as default classifier for entities, animals and things in general without any obvious criterion (12).

4. Another item, ‘wife’, could be added here, provided that the Pana classifier *ka* for non-humans and human inferiors (Prost 1969: 894) could be shown to derive from the noun *kā* ‘wife’.

- (11) *Source item ‘child’ as classifier*
 Kana: *ɲwíí* “child” (PL. *míónj́*) (Ikoro 1996: 93)
zǐ́ ɲwíí pée
 one CLF:child goat
 “one kid”
- (12) *Source item ‘mother’ as classifier*
 Kana: *kà* “mother” (Ikoro 1996: 90–91)
lòb kà kúǐ
 ten CLF:mother basket
 “ten baskets”

It is striking that these basic level source items, along with the animacy category they establish in classifier systems, are restricted to languages which display almost complete lexical coverage of classifiers and have no coexistent noun class system, i.e. in Mande, in Kana and in some Gur languages. The reason for their absence in classifier systems of restricted lexical coverage, i.e. in most of Benue-Congo, e.g. Ejagham, Denya and Grassfields, is that animacy is usually taken care of by the coexistent noun class systems in those languages.

3.3 Botanical terms

Botanical source concepts are manifest in numeral classifiers which are derived from generic botanical terms or nouns for parts of plants (Table 5) such as ‘grain’, ‘fruit’, ‘tree (trunk)’, ‘pod’, ‘leaf’, ‘stick’ and ‘palm kernel’.

Table 5. Nouns for botanical concepts as sources of numeral classifiers

Lexical source of classifier	Semantic range of classifier and attestation
‘tree’	wooden objects > vehicles of transportation (Kana) objects of oblong shape with a vertical orientation (Gban)
‘tree trunk’	trees (Ejagham)
‘fruit’	fruits (Denya) globular objects (Gban) globular objects > drops of liquid (Bafanji) globular objects > offspring (Kana)
‘grain’	individual units extracted from an organic aggregation (Ngiemboon, Kana) small objects of oblong or roundish shape, both animate (but non-human) or inanimate, e.g. small animals, roots, tubers, leaves (Toussian)
‘pod’	shells with several smaller fruits inside (Bafanji)
‘leaf’	flat objects (Gban)
‘stick’	objects of elongated shape (Gban)
‘palm kernel’	bitter kola kernels (Ngwe)

The concept ‘grain’ is employed for classifying individual units extracted from a stock or a bunch (Kana, 13a–b) or some organic aggregation (Ngiemboon, 13c), while in Toussian (13d) it has become generalized to classify all kinds of small objects of oblong or roundish shape including small animals on the one hand and inanimate objects on the other hand, attesting the most advanced stage of grammaticalization. Thus, classifiers derived from the noun ‘grain’ overlap in their range of application with classifiers derived from the body part concept ‘eye’. In languages such as Kana where both coexist one might be inclined to think that the source noun ‘grain’ tends to remain restricted to the botanical domain in its classifier function, maintaining its relation to an organic aggregation, while the source noun ‘eye’ (1a) is extracted more easily from its basic frame of a human or animal body to refer to a generalized haptic concept. However, this does not seem to be the case, as demonstrated by (13a). It rather seems to be the notion of an aggregation which differentiates the classifier noun ‘grain’ from the classifier noun ‘eye’ here.

(13) *Source item ‘grain’ as classifier*

- a. *Kana: ásúú* “grain” (Ikoro 1996: 95)
bàè ásúú m̀̀ǹ̀ǹ̀
 two CLF:grain fist
 “two fists”
- b. *l̀̀òb̀̀ ásúú kp̀̀àkp̀̀à*
 ten CLF:grain maize
 “ten maize grains”
- c. *Ngiemboon: mv̀̀m̀̀* “grain” (Vinogradov 2009)
mekú mv̀̀m̀̀ tá
 beans CLF:grain three
 “three beans (of haricot)”
- d. *Toussian: p̀̀ùe* “grain” (PL. *pl̀̀ú*) (Zaugg-Coretti 2005: 23)
ǰ̀̀ẁ̀ p̀̀ùè ǹ̀ǹ̀
 antelope.SG CLF:small.SG one
 “one single antelope”

The lexical source concept ‘fruit’ either remains, as a classifier, restricted to the domain of fruits (Denya, 14a) or undergoes a haptic generalization in its application to objects of roundish shape (Bafanji, 14b) or is additionally expanded by a functional generalization of the offspring notion (Kana, 14c).

(14) *Source item ‘fruit’ as classifier*

- a. *Denya: g̀̀èl̀̀óó* (7/8) “fruit”
ń̀̀ ǰ̀̀s̀̀ g̀̀è-l̀̀óó gé m̀̀è-f̀̀i gé-m̀̀á
 1SG pick 7-CLF:fruit ASS.7 3-mango 7-one
 “I have picked one mango.”

- b. *Bafanji*: *tēē* (7/6) “fruit” (Hamm 2010)
tēē ləʔɥ
 7.CLF:fruit sweet.yam
 “one piece of sweet yam”
- c. *Kana*: *bēē* “fruit” (Ikoro 1996: 91)
bàè bēè ɲwíí
 two CLF:fruit child
 “two children”

The lexical source concept ‘tree’ undergoes double metonymy in its use as a classifier in Kana where it is first expanded to classify wooden objects in general (15a). In a second semantic expansion it is used for vehicles of transportation (15b). In Gban, it rather undergoes a haptically motivated generalization in its application to objects of oblong shape with a vertical orientation. In Ejagham, the lexical source ‘tree trunk’ is expanded by synecdoche to classify trees in general (15c).

(15) *Source item ‘tree (trunk)’ as classifier*

- a. *Kana*: *té* “tree” (Ikoro 1996: 91)
tāà té náā
 three CLF:tree gun
 “three guns”
- b. *Kana*: *té* “tree” (Ikoro 1996: 92)
tāà té itòbtòb
 three CLF:tree motorcycle
 “three motorcycles”
- c. *Ejagham*: *èsán* “tree trunk” (Watters 1981: 312)
è-sán í-cókùd já-d
 5-CLF:trunk ASS.5 19-orange.tree 5-one
 “one orange tree”

Two further lexical source items in Gban, ‘stick’ and ‘leaf’, undergo haptically motivated expansion in their classifier application to objects of elongated shape and objects of flat shape, respectively. Note that for deriving the second classifier category the botanical model as applied in Gban is in contrast with Kana where a similar function is rather derived from the body part concept ‘skin’.

Marginally, two further botanical source models of classifiers occur: the item ‘pod’ is used to classify shells with several smaller fruits inside (Bafanji), and the item ‘palm kernel’ which is applied to classify bitter kola kernels (Ngwe).

3.4 Terms of aggregation and partition

Beside body part nouns and generic or basic level terms, classifiers are also found to be derived from terms for aggregational concepts such as ‘bunch’ (Kana, Denya, 16), ‘bundle’ (Bafanji), ‘stock’ (Kana, 17), ‘heap’ (Kana, 18) and ‘lump (of clay)’ (Bali-Mungaka, 20) on the one side and partitional concepts such as ‘bottom’ (Ejagham, 19), ‘part’ (Kana), ‘piece of’ (Kana) and ‘half’ (Kana) (see Table 6).

Table 6. Terms for aggregational and partitional concepts as sources of numeral classifiers

Lexical source of classifier	Semantic range of classifier and attestation
‘bunch’	objects of bunch- or package-form (Kana) fruits which come in distinctive clusters (Denya)
‘stock’	objects of a stock or a bunch (Kana)
‘bundle’	2 types: wrapped bundles vs. tied bundles (Bafanji)
‘heap’	objects with irregular shape which come in heaps or chunks (Kana)
‘lump (of clay)’	lumpy, bulging or clotted objects (Bali-Mungaka)
‘bottom’	trees, shrubs, vegetables (Ejagham)
‘part’	potentially separable parts of objects (Kana)
‘piece’	pieces of objects, splashes of liquids (Kana)
‘half’	objects segmented into parts (Kana)

(16) *Source item ‘bunch’ as classifier*

- a. *Denya: ndòó* (9/10) “bundle, cluster” (SG. = PL.)

ndòó ú- ↓kwá é-péá

10.CLF:bunch ASS.10:8-plantain 10-two

“two bunches of plantains”

- b. *Kana: kō* “bunch”

(Ikoro 1996: 98)

zǐ̃ kṑ zó̀

one CLF:bunch palmtree

“one bunch of palmtrees”

(17) *Source item ‘stock’ as classifier*

Kana: kūm “stock”

(Ikoro 1996: 98)

zǐ̃ kū̀m zó̀

one CLF:stock palmtree

“one stock of palmtrees”

(18) *Source item ‘heap’ as classifier*

Kana: kpò “heap”

(Ikoro 1996: 95)

tū̀̀b à̀kpò dēm

twenty CLF:heap brick

“twenty bricks”

- (19) *Source item ‘bottom’ as classifier*
Ejagham: Ñ-dǎb (PL. *à-rǎb*) “bottom” (Watters 1981: 312)
à-rǎb ‘ *ǝ-rǎbé* *á-bá[↓]é*
 6-CLF:? ASS.6 14-bean 6-two
 “two bean plants”
- (20) *Source item ‘lump’ as classifier*
Bali-Mungaka: túʔ “lump (of clay)” (Tischhauser & Stöckle 1993: 363–364)
(n)túʔ nǎm íbá
 lump meat two
 “two pieces of meat”

3.5 Obscure etymology

The above generalizations are very preliminary for two reasons. First, in most, if not all cases, the attested numeral classifiers need more accurate description in terms of their range of application. Second, more sophisticated etymological analyses combined with progress in low-level comparative work might unearth etymologies of classifiers which have not been included in Tables 2–6 either because the lexical source suggested has not been confirmed yet, e.g. with Ejagham *Ñ-mǎgè* (PL. *à-mǎgè*) “small globular object” (21a), or because no lexical source has been recognized at all. This holds for the Ejagham classifiers *Ñ-súm* (PL. *à-súm*) for longish fruits and tubers (21b) and *è-rám* (PL. *Ñ-dám*) for roundish fruits or tubers which come in bunches (21c) (Watters 1981: 310–313), for the Kana classifiers *ábá* for objects with tube-like shape (21d), *àpā* for conically-shaped objects (21e), *ákpé* for objects with vertically high extension (21f), and for the majority of Bafanji, Ngwe and Ngiemboon classifiers (see Appendix).

- (21) *Classifiers without established lexical source*
- a. *Ejagham: Ñ-mǎgè* (PL. *à-mǎgè*) “small globular object(s)”
à-mǎgè ‘ *ǝ-rǎbé* *á-bá[↓]é*
 2-CLF:small.round.object? ASS.2 14-bean 2-two
 “two beans”
- b. *Ejagham: Ñ-súm* (PL. *à-súm*) (longish fruits, tubers)
à-súm ‘ *ǝ-rǎbé* *á-bá[↓]é*
 6-CLF:? ASS.6 14-bean 6-two
 “two bean pods”
- c. *Ejagham: è-rám* (PL. *Ñ-dám*) (fruits, tubers in bunches)
Ñ-dám *ì* *Ñ-bǎŋ* *é-bá[↓]é*
 9-CLF:? ASS.9 9-coconut 9-two
 “two coconuts”

- d. *Kana: ábá* (objects with tube-like shape) (Ikoro 1996: 95)
 z̩̀í ábá z̩̀íá
 one CLF:? yam
 “one yam tuber”
- e. *Kana: àpā* (conically-shaped objects) (Ikoro 1996: 97)
 bàē àpāà má
 two CLF:? breast
 “two breasts”
- f. *Kana: ákpé* (objects with vertically high extension) (Ikoro 1996: 97)
 z̩̀í ákpé dáé
 one CLF:? wall
 “one wall (of house)”

Most of these cases might reflect an advanced stage of grammaticalization insofar as the items have lost their syntactic autonomy, while they retain most of their original nominal properties, e.g. assignment to a gender and pluralizability.

4. Syntactic sources of classifier constructions

In this section, the Niger-Congo numeral classifier constructions are analysed for three crucial morphosyntactic parameters: word order, immediate constituency and cliticization, i.e. morphophonological fusion.

With regard to word order, numeral classifier constructions in Niger-Congo roughly fall into two major types (22). In type I the classifier follows the enumerated noun, whereas type II is characterized by the reverse order, i.e. the classifier precedes the enumerated noun. Depending on the position of the numeral, type II presents two subtypes: in type IIa the numeral follows both classifier and enumerated noun, while in type IIb the numeral precedes the entire classifier construction. No cases have been found where the numeral intervenes between classifier and enumerated noun (*Ic, *IIc). Furthermore, for type I there is no corresponding subtype which places the numeral in the initial position.

(22) Word order types of numeral classifier constructions

- I enumerated noun + classifier:
- a. enumerated noun + classifier + numeral
 - *b. numeral + enumerated noun + classifier
 - *c. enumerated noun + numeral + classifier
- II classifier + enumerated noun:
- a. classifier + enumerated noun + numeral
 - b. numeral + classifier + enumerated noun
 - *c. classifier + numeral + enumerated noun

The basic division of numeral classifier construction type I vs. type II roughly correlates with the word order in possessive associative constructions, the classifier position corresponding to the position of the modified head noun which encodes the possessed item, while the position of the enumerated noun corresponds to the position of the dependent modifier which encodes the possessor. This correlation stands out most clearly in a comparison of numeral classifier constructions across typologically different languages which deviate with respect to word order in possessive constructions, i.e. Mande and Gur on the one side vs. Benue-Congo on the other side.

All numeral classifier constructions of Mande and Gur languages attested so far belong to type I, i.e. the enumerated noun precedes the classifier, while the numeral follows in terminal position (23). This order perfectly matches the general word order found in the noun phrase of type B languages such as these, i.e. nominal (possessive) modifiers in associative constructions precede their head, while all other modifiers such as adjectives, demonstratives and numerals follow (Heine 1976: 41–42).

- (23) Type I: correlation with NP-internal order (Mande and Gur)
 classifier construction:
 enumerated noun + classifier + numeral
 general order in NP:
 possessor + possessed head + modifiers

This correlation clearly emerges from a comparison of word order in classifier constructions vs. the respective associative constructions in Busa (24), Toussian (25) and Pana (26).

- (24) Busa
- a. *Numeral classifier construction* (Wedekind 1972: 167)
mátò mè pláā
 car CLF:body two
 “two cars”
- b. *Associative construction* (Jones 1998: 70)
bòó búa dafu daaleā
 Boko farm new beginning
 “the beginning of a new Boko farm”
- (25) Toussian
- a. *Numeral classifier construction* (Zaugg-Coretti 2005: 22)
íkpa nŏw tó
 child.PL CLF:person.PL three
 “three children”

- b. *Associative construction* (Zaugg-Coretti 2005: 59)
nòw-té mlí fà
 head-owner nose hole
 “hole of the nose of the chief”
- (26) Pana
- a. *Numeral classifier construction* (Prost 1969: 894)
tyibi la-nyi wora a ka
 brothers CLF:human-two be.located 1SG with
 “Two younger brothers are with me.”
- b. *Associative construction* (Beyer 2006: 111)
ká nìmi jénjé-lè
 woman calabash small-DEF
 “the small calabash of a woman”

In most Benue-Congo classifier constructions, on the other hand, the classifier precedes the enumerated noun, while the numeral follows in terminal position (27). This order matches the general word order in the noun phrase of type A languages such as Ejagham (28), Denya (29), Isu (30), Bali-Mungaka (31), i.e. all modifiers – including possessors – follow their head noun (Heine 1976: 39–40).

- (27) Type IIa: correlation with NP-internal order (Benue-Congo)
 classifier construction:
 classifier + enumerated noun + numeral
 general order in NP:
 possessed head + possessor + modifiers

- (28) Ejagham
- a. *Numeral classifier construction* (Watters 1981: 310–313)
à-màgè ' ó-ròbé á-bá¹é
 2-CLF:small.round.object ASS.2 14-bean 2-two
 “two beans”
- b. *Associative construction* (Watters 1981: 354)
Ñ-tég i Ñ-sé òmé
 9-village ASS 1-father 1:my
 “my father’s village”

- (29) Denya
- a. *Numeral classifier construction*
à-mé jìci á-léé
 2-CLF:eyes ASS.2:9.maize 2-three
 “three grains of maize”
- b. *Associative construction* (Mbuagbaw 1996: 64)
mè-kpó ' m̀p̀
 5-head ASS:5 9.cow
 “head of a cow”

- (30) Isu
- a. *Numeral classifier construction*
túw fífəp fə mɔ?
 19.little.head Ass.19:5.leaf Ass.19 one
 “one single leaf only”
 - b. *Associative construction*
tə-yáw tə-fə-ɲwòní
 13-wings 13–19-bird
 “wings of a bird”
- (31) Bali-Mungaka
- a. *Numeral classifier construction* (Tischhauser & Stöckle 1993: 164)
bón lí? nsú íbá
 children CLF:eye fish two
 two little fish”
 - b. *Associative construction* (Hombert 1980: 161)
bəm mfɔn
 5.belly:Ass.5 1.chief
 “the chief’s belly”

Type IIb differs from type IIa in only one respect: the numeral precedes the classifier construction (32), as observed in some Benue-Congo languages such as Kegboid and Bafanji. For Kegboid (33) this order clearly conforms to the general NP-internal order, i.e. the nominal possessor follows its head noun while most other modifiers precede, e.g. pronominal possessors, attributive adjectives and most specifiers (Ikoró 1996).⁵ For Bafanji, however, this order is largely at odds with the NP-internal order, since it is only quantifiers and numerals which precede the head noun, while all other modifiers follow (Moritz 1994: 95, 96).

- (32) Type IIb: correlation with NP-internal order (Kegboid)
 classifier construction:
 numeral + classifier + enumerated noun
 general order in NP:
 modifiers + possessed head + possessor

5. The situation is more complex: pre-head position holds for attributive adjectives, numerals, definite markers and pronominal possessors in alienable possession, whereas post-head position holds for nominal possessors in associative constructions, inalienable body-part possessive pronouns and locative adverbs which fulfill the function of demonstratives.

- (33) Kana
- a. *Numeral classifier construction* (Ikoro 1996: 99)
bàè kááná àpāà bári
 two nice CLF:? fish
 “two nice fish”
- b. *Associative construction* (Ikoro 1996: 82)
bèè méné
 home.ASS chief
 “a chief’s home”

These structural correlations of numeral classifier and associative constructions on the one side and the transparent etymology of most of the classifiers themselves, as laid out in Section 3, confirm the conclusion that in most of the observed cases numeral classifier constructions actually originate in associative constructions in which the head noun tends to undergo nominal decategorization and incipient grammaticalization as a classifier, as argued in Section 5.

Remarkably however, these harmonic word order correlations between classifier construction and associative construction do not seem to hold for some of the Bamileke languages. Thus, in Ngwe, Ngiemboon and Yemba numeral classifier constructions follow type I, i.e. the enumerated noun precedes the classifier, whereas associative constructions generally follow type II, i.e. the possessed noun precedes the possessor, as outlined in (34) and exemplified in (35)–(36).

- (34) Ngiemboon and Yemba: non-harmonic mixed type I/IIa
- a. classifier construction:
 enumerated noun + classifier + numeral
- b. unmarked general order in NP:
 possessed head + possessor + modifiers
- c. marked order in NP:
 possessor + possessed head + modifiers
- (35) Ngiemboon
- a. *Numeral classifier construction* (Vinogradov 2009)
fu mbáj lekúá
 lice CLF:? four
 “four lice”
- b. *Associative construction* (Anderson 1983: 32)
lètyó lékwòŋ
 5.point 5:5.spear
 “point of a spear”

- (36) Yemba
- a. *Numeral classifier construction* (Harro & Haynes 1991: 30–31)
metíta atá píá
 6.potato 7.CLF:round.oblong.fruit 8.two
 “two potatoes”
- b. *Associative construction* (Harro & Haynes 1991: 38)
nzā e efo
 9.axe 9 1.chief
 “chief’s axe”

Some of these Eastern Grassfields languages allow for possessors to precede their heads under special pragmatic circumstances though, e.g. Yemba (Harro & Haynes 1991: 27–28). Thus, the exceptional word order in classifier constructions might simply be due to their emergence from pragmatically marked (34c) rather than unmarked associative constructions (34b). Otherwise, the lack of structural isomorphism might be regarded as an outcome of advanced grammaticalization maximizing the distinction of the emerging classifier construction from its predecessor construction. In order to improve the understanding of the diachronic morphosyntax of these classifier systems, however, much more effort will have to be invested into the study of their synchronic details.

Moreover, in Ngiemboon and Yemba the numeral ‘one’ demands an exceptional non-harmonic word order which represents a mirror-image of the general word order in numeral classifier constructions: the numeral ‘one’ preceding the classifier construction in which the classifier precedes the enumerated noun, as indicated in (37) and exemplified in (38).

- (37) Ngiemboon and Yemba: exceptional non-harmonic order with numeral ‘one’ classifier construction:
 enumerated noun + classifier + numeral
 classifier construction ‘one’:
 numeral ‘one’ + classifier + enumerated noun
- (38) *Exceptional order in classifier constructions with ‘one’*
- a. *Ngiemboon* (Vinogradov 2009)
taʔ mb^ho legí
 one CLF:hand grass
 “one blade of grass”
- b. *Yemba* (Harro & Haynes 1991: 30–31)
taʔ atá metíta
 one 7:CLF:round.oblong.fruit 6:potato
 “one potato”

The reason for this deviation of the numeral ‘one’ is unclear. It might be the case that the shared cognate *taʔ* ‘one’ is not a genuine numeral in these languages, but rather a special noun. This is supported by evidence from Bali-Mungaka. While in general Bali-Mungaka numerals follow both classifier and enumerated noun (31), the numeral *mɔʔ* ‘one’ precedes them (1e). In addition to this numeral, Bali-Mungaka possesses a set of classifying numerals which conflate the meaning ‘one’ with a haptic classifying concept. Thus, *ntàm* is restricted to single objects with oblong shape (39a), while *ndàʔ* applies to single artefacts such as pieces of clothing (39b) and *táʔ*, which is cognate to the Yemba and Ngiemboon items in (38), is used for single collections of entities (39c).

(39) *Bali-Mungaka: classifying numerals ‘one’*

- a. *ntàm* ‘single oblong object’ (Tischhauser & Stöckle 1993: 254)
ntàm *nǎŋ-tú*
 CLF:single.oblong hair-head
 ‘a single hair’
- b. *ndàʔ* ‘single artefact?’ (Tischhauser & Stöckle 1993: 229)
yó yá ndàʔ ndzí
 this 1SG.POSS CLF:single dress
 ‘This is my only dress.’
- c. *táʔ* ‘single unit composed of distinct individuals’
 (Tischhauser & Stöckle 1993: 351)
táʔ wàŋ nín
 CLF:single group one
 ‘one single group’

All of these single unit classifying numerals precede the enumerated noun, occupying the head position and betraying their character as nouns. Although Bali-Mungaka has been documented rather extensively, as compared to most other languages of the Grassfields, unfortunately no more information concerning these classifying numerals could be retrieved from the sources. So this issue needs further dedicated research.

Following Greenberg (1972), Allan (1977: 288) states ‘a universal principle that a CLASSIFIER CONCATENATES WITH A QUANTIFIER, LOCATIVE, DEMONSTRATIVE, OR PREDICATE TO FORM A NEXUS THAT CANNOT BE INTERRUPTED BY THE NOUN IT CLASSIFIES.’ This entails that numeral classifier constructions allow for only four possible constituent orders: (a) [NUM CLF] N, (b) N [NUM CLF], (c) [CLF NUM] N, (d) N [CLF NUM], as restated in Aikhenvald (2000: 104–105): ‘These orders exhaust all possibilities – numerals and classifiers are always adjacent.’ A generalization such as this claims that sequences such as CLF N NUM and NUM N CLF are not permissible in numeral classifier constructions. Apart from types

- b. *Isu*
má má kòʔ àwó á káfè yiy à bè
 1SG IMM.PST.FOC see 6.hand ASS.6 7.plantain 6:OF ASS.6 two
 “I have seen only two clusters of plantains.”
- c. *Denya*
gè-lò gé ʔjici gé-mâ
 7-CLF:finger ASS.7 maize 7-one
 “one cob of maize”

In numeral classifier constructions, two of the three minimally involved items tend to form a tighter morphosyntactic unit which acts as a domain of morphophonological rules to the exclusion of the third item, as indicated by bracketing in (42).

(42) Immediate constituency in numeral classifier constructions

Benue-Congo:

Kana, Bafanji:	NUM [CLF N]
Bali-Mungaka:	[CLF N] NUM
Ejagham, Denya, Isu:	[CLF N] NUM
Ngiemboon, Yemba, Ngwe:	N [CLF NUM]

Gur:

Toussian, Pana:	N [CLF NUM]
-----------------	-------------

Mande

Busa, Gban:	N [CLF NUM]
-------------	-------------

In most numeral classifier languages worldwide, it is the classifier and the numeral which form such an immediate constituent to the exclusion of the enumerated noun (Dixon 1986; Aikhenvald 2000: 105). This generalization also seems to be valid for the Niger-Congo numeral classifier systems of construction type I, e.g. Pana and Toussian. Thus, Prost (1969) actually describes the Pana classifiers as prefixes to the numeral (8c). However, for Niger-Congo languages which present numeral classifier constructions of the other types IIa and IIb, the classifier rather forms a constituent with the enumerated noun to the exclusion of the numeral.⁶

6. In a wider diachronic – or rather prognostic – perspective, the difference in immediate constituency and resulting tendencies towards morphophonological fusion indicated in (42) lays the foundation for different paths of grammaticalization in a prospective development of noun classes from the numeral classifiers attested so far: while a fusion of the classifier to the numeral in all Mande and Gur languages and in the Bamileke languages Ngiemboon, Yemba and Ngwe would account for the emergence of concordial markers first in numeral agreement targets as part of an emergent covert noun class system, the fusion of the classifier with the enumerated noun in all Benue-Congo languages (except the Bamileke cases cited above) would account for a direct development of nominal prefixes as part of an emergent overt noun class system, rather than concordial markers.

These constituency relations are directly inherited from the associative predecessor construction, as indicated in (40). In associative constructions of type IIa and IIb languages, it is always the possessed head noun and the modifying possessor which form the inseparable core of a more complex noun phrase which must not be disrupted by any of the other modifiers such as determiners, demonstratives, attributive adjectives and numerals. Since the classifier originates in a possessed head noun, while the enumerated noun derives from the modifying possessor, their morphosyntactic and morphophonological unity simply continues the prior unity of possessed head and modifying possessor in the associative predecessor construction.

For Kana, Ejagham and Isu the tight unit of classifier and enumerated noun is manifest in morphophonological, morphotonological and morphosyntactic effects which link both items to the exclusion of the numeral. Thus, in Kana classifier and enumerated noun are linked by the grounding of the floating low tone associative marker which docks to the classifier rather than to the numeral (Ikoro 1994: 20), changing the classifier's tone from mid, *bēē*, to falling from mid to low, *bēè* (43a). Moreover, syntactic expansions by the diminutive marker *í-* (43b) and by attributive adjectives such as *kááná* 'nice' (43c) might separate the numeral from the classifier but leave the union of classifier and enumerated noun unaffected (Ikoro 1994: 21–22, 26; Ikoro 1996: 99–100).

(43) *Kana: immediate constituency of classifier and noun*

- a. *Associative marker links classifier and enumerated noun* (Ikoro 1994: 20)
tùùb bēē ɲwíí
tùùb bēē- ɲwíí
 twenty CLF:fruit-ASS child
 "twenty children"
- b. *Diminutive marker separates numeral and classifier* (Ikoro 1996: 99)
zī í-bēē ɲwíí
 one DIM-CLF:fruit:ASS child
 "one little child"
- c. *Attributive adjective separates numeral and classifier* (Ikoro 1996: 99)
bàè kááná àpàà bári
 two nice CLF:?:ASS fish
 "two nice fish"

Contrary to Aikhenvald's statement (2000: 110–111), Kana and the Kegboid languages are not exceptional in this respect. Along with it go Denya, Ejagham, Bali-Mungaka and Isu, possibly others as well. Thus, in Ejagham the tighter unity of classifier and enumerated noun is reflected in morphotonological effects such as the grounding of the floating high tone associative class marker in the class prefix

of the immediately following enumerated noun (44a), turning the underlying form *ɔ̀rɔ̀bɛ́* “beans” into the surface form *ɔ̀rɔ̀bɛ́* (Watters 1981: 100–102, 354). Isu presents two pieces of evidence for the immediate constituency of the classifier and the enumerated noun, as illustrated in (44b). First, the associative marker *i* of the classifying head noun and the associative low tone undergo fusion with the enumerated noun, while at the same time the terminal numeral could be separated from the enumerated noun by intervening determiners such as the out-of-focus marker *yíy*.

(44) *Evidence of immediate constituency of classifier and noun*

a. *Ejagham*

à-súm *ɔ̀-rɔ̀bɛ́* *á-bá[↓]é*

à-súm *´-ɔ̀-rɔ̀bɛ́* *á-bá[↓]é*

6-CLF:long ASS.6–14-bean 6-two

“two bean pods”

b. *Isu*

ìsǐ́ *í[↓]tsíy* *yíy* *í* *mɔ̀ɔ̀*

ì-sǐ́ *í-´-í-tsíy* *y-íy* *í* *mɔ̀ɔ̀*

5-eye ASS.5-AT-5-fish 5-OF ASS.5 one

“only a single fish”

While the origin of Niger-Congo numeral classifier constructions in associative possessive constructions is very common cross-linguistically (e.g., Bisang 1999), it has been shown that it leads to a word order pattern in some Niger-Congo languages which does not seem to be attested elsewhere: [CLF N] NUM. Rather, cross-linguistically the juxtaposition of classifier and numeral is preferred due to their tighter cognitive relation. The maintenance of the cognitively discontinuous order [CLF N] NUM after its emergence from plain associative constructions in languages such as Ejagham, Denya and Isu is possible because of the presence of the coexistent noun class system which caters for overtly marking the coherence of classifier and numeral by class agreement. This is obvious in (41) and (44) where the terminal numerals bear the class markers of the respective classifier nouns. Once the word order [CLF N] NUM is firmly established in a language, it may be strong enough to be preserved even if the noun class system breaks down and agreement can no longer be used to mark constituency, e.g. in Bali-Mungaka. In languages such as Yemba, Ngiemboon and Ngwe, though, it seems as if the strong cognitive bond between classifier and numeral has caused the syntactic shift juxtaposing both constituents in emancipation from the prior associative construction (34)–(36).

5. Grammaticalization: From noun to classifier

The emergence of distinct numeral classifier systems is accompanied by both morphological and syntactic indicators of grammaticalization. On the syntactic level, reanalysis of prior associative constructions in terms of newly emerging numeral classifier constructions is reflected in word order changes. On the morphological level, a progressive loss of nominal properties in the classifier nouns points to their gradual emancipation from their lexical sources.

Syntactically, synchronic mismatches regarding word order in classifier constructions vis-à-vis associative constructions, e.g. in Ngwe, Ngiemboon and Yemba (34)–(36), mark the advancement of syntactic reanalysis which works towards an emancipation of the numeral classifier construction from its associative predecessor construction.

Other indications might be subtler. Thus, in Toussian the position of attributive adjectives in numeral classifier constructions seems to be different from their position elsewhere. While in associative constructions they follow the possessed head noun as do all non-possessor modifiers (45b), attributive adjectives such as *pièw* “white” are placed directly after the enumerated noun, *sây* “feather”, in classifier constructions (45c), testifying to a distinct order in classifier constructions (45a).

(45) *Toussian: distinct order in numeral classifier constructions*

- a. classifier construction:
enumerated noun (+ adjective) + classifier + numeral (+ determiner)
- b. general order in NP:
possessor + possessed head + modifiers
- c. *Position of the attributive adjective in classifier constructions:*

(Zaugg-Coretti 2005: 45)

sây-pièw *púé* *nēn kè-pè-n* *ɲòŋ mè* *wò*
feather-white CLF:grain.SG one NEG-COP-IPFV fowl DEM:SG LOC
“This guinea fowl does not have a single white feather.”

This distinction might be motivated by the intention to prevent the classifier and the numeral from being separated by the adjective, since in Toussian classifier and numeral form a tight unit, a “syntagme numéral” (Zaugg-Coretti 2005: 46) which must not be split by any insertions, conforming to Dixon’s morphosyntactic generalization (1986: 106) and contrasting with most Benue-Congo numeral classifier systems where it is rather the enumerated noun and the classifier which form a unit.

On the morphological level, incipient grammaticalization is reflected in a gradual loss of nominal properties in classifier nouns, indicating a functional split of lexical source items vs. a newly emergent word class of numeral classifiers. Nominal

properties affected by loss pertain to (a) syntactic autonomy, (b) number distinction and (c) concordial agreement features. Thus, individual classifiers could be located on a cline ranging from a full-fledged noun over various stages of defective nominal status to a full-fledged classifier which does not retain any nominal properties. Full-fledged classifiers at the extreme end of this cline seem to be rare for two reasons: either, as in most Benue-Congo languages which retain coexistent noun class systems, the development of classifiers seems so recent that they could not have had the time to fully dissociate from their nominal sources, or classifiers are attested in languages such as Kana and Eastern Mande, which widely lack a distinct nominal morphology, as enshrined in a noun class system, so that diagnostic criteria (b) and (c) fail and only criterion (a) remains.

The loss of syntactic autonomy of classifying nouns is attested for Kana classifiers *ábá* for “objects with tube-like shapes”, *ápā* for “conically-shaped objects” and *ákpé* for “vertical-high objects” (Ikoro 1996: 90, 95–97), exemplified in (21d)–(f), for Ejagham classifiers *Ñ-súm* (PL. *à-súm*) for longish fruits and tubers and *è-rám* (PL. *Ñ-dám*) for roundish fruits or tubers which come in bunches (Watters 1981: 310–313), exemplified in (21a)–(c), and for the Toussian diminutive classifier *púe* (PL. *plú*) which is restricted to compounds such as *sár-púe* “oil palm grain” (PL. *sēn-plú*), but does not occur in isolation. For most of the Bafanji, Ngwe and Ngiemboon classifiers it might be concluded preliminarily that they lack syntactic autonomy. In all instances, these items seem to be restricted to numeral classifier constructions and cannot act as autonomous nouns any longer. While for Ejagham, Ngwe, Bafanji, Ngiemboon and Toussian these classifiers visibly retain nominal properties, this is more difficult to decide for Kana since criteria such as noun class assignment and pluralizability are lacking.

A loss of the number distinction in the transition from a fully fledged noun to a classifier is attested in Bali-Mungaka and Yemba. Thus, Bali-Mungaka *líʔ* “eye” has a distinct plural form *míʔ* “eyes” (46a). In classifier function, however, *líʔ* does not vary for number: even in plural contexts such as (46b) it is the singular form which occurs.

(46) *Bali-Mungaka: number neutralization in classifier*

(Tischhauser & Stöckle 1993: 164)

- a. *míʔ m-a*
 5.eyes 5-1SG.POSS
 “my eyes”
- b. *bón líʔ nsú íbá*
 children CLF:eye fish two
 “two little fish”

In Yemba, the number distinction in classifier nouns seem to be neutralized in favour of the singular, while gender assignment and concordial properties are retained (Harro & Haynes 1991: 30–31). A classifier such as *atá* (for round, oblong objects such as mango, avocado, taro) belongs to gender 7/8 and forms a distinct plural *etá*. Thus, in classifier constructions such as (47) it assigns agreement according to gender 7/8 in the numeral ‘two’ which is spelt out in terms of the plural class 8, whereas the classifier itself remains invariant in its singular form of class 7.

- (47) *Yemba: number neutralization in classifier*
metíta atá píá
 6:potato 7:CLF:round.oblong.fruit 8:two
 “two potatoes”

In Isu, the functional split of classifier usage and original lexical meaning of *ísí* ‘eye’ is reflected in its tonal behaviour. Thus, the lexical use of *ísí* ‘eye’ (48a) demands the associative form of the possessor noun which is marked by an initial floating low tone (labelled AT) which prevents the high tone in the associative marker *í* from spreading to the possessor noun *wù* ‘person’. The classifier use of the noun *ísí* ‘eye’ (48b), however, dispenses with this floating low tone and therefore allows the high tone of the associative linker *í* to spread to the enumerated second noun creating a falling tone *wù* ‘person’.

- (48) *Isu: loss of concordial agreement features in classifier*
 a. *ísí í ` wù ì mǝ?*
 5.eye 5 AT 1.person 5 one
 “one eye of a person”
 b. *ísí í wù ì mǝ?*
 5.eye 5 1.person 5 one
 “one single person only”

The loss of the syntactic autonomy, nominal number distinction (46)–(47) and concordial agreement features (48) in classifying nouns is diagnostic of their incipient decategorization and visibly marks the functional split of genuine noun and classifier. Instances such as Bali-Mungaka *lí?* (46), Yemba *atá* (47) and Isu *ísí* (48) represent defective nouns halfway to being fully grammaticalized: while they retain some nominal properties such as the capacity to trigger gender agreement and to act as heads of a NP, they have lost some other crucial nominal properties, e.g. syntactic autonomy and pluralizability.

6. Puzzles

Apart from the generalizations on lexical sources of classifiers, syntactic sources of classifier constructions and formal symptoms of grammaticalization compiled in Sections 3–5, many open questions remain. To make progress in understanding the emergence and development of African numeral classifier systems, some puzzles pertaining to the ultimate motivations of the emergence of numeral classifiers and their widely varying degrees of elaboration need to be solved.

There is little etymological conformity across the classifier items recognized so far. Despite the close genetic relationship of languages such as the Eastern Grassfields Bantu languages Bafanji, Bali-Mungaka, Ngiemboon, Ngwe and Yemba, only very few classifiers share a common etymology which would allow a reconstruction at an earlier level, e.g. Proto-Bamileke or Proto-Eastern Grassfields. Thus, Ngwe *àtáá* (classifier for round objects) and Yemba *atá* (classifier for roundish oblong objects) are cognate, and Ngwe *ntfʒʔ* (classifier for long objects which hang in bunches), Yemba *ntsʔ* (classifier for long objects) and Bafanji *tʒʔú* (classifier for a objects in a group) derive from a common source. In order to decide about the cognacy of Ngwe *mbāŋ* ‘palm kernel’ and Ngiemboon *mbáŋ* (classifier for small insects), both classifiers have to be determined more accurately for their function. Apart from these cognates, most of the other numeral classifiers of the Eastern Grassfields seem to be recent, independent innovations.⁷ In order to derive reliable conclusions concerning the historical level of innovation of classifiers, much more groundwork of language description and reconstruction still has to be done on the level of individual subgroups. In general, however, the lack of etymological conformity across families as well as within families points to the fact that classifiers present innovations which have arisen independently in Eastern Mande, several Gur languages and various branches of Benue-Congo.

With regard to potential motivations of the emergence of numeral classifier systems in Niger-Congo, the Gur and Kegboid examples suggest that it is the decline and the loss of a prior noun class system which triggers off a process of innovation of numeral classifiers, since it is precisely these languages in which numeral

7. Beyond the Eastern Grassfields level, some of the classifiers which originate in body part nouns ‘eye’, ‘head’ and ‘hand’ are cognate. Thus, cognate forms of the noun ‘eye’ are used for small roundish objects in Bafanji (*ndí*), Bali-Mungaka (*líʔ*) and the Ring subgroup of Grassfields Bantu (e.g., Proto-West Ring **isi*). All of them go back to Proto-Benue-Congo **lito* (de Wolf 1971: 54). While this might suggest that the noun ‘eye’ could have been employed as a numeral classifier at this level already, it does not prove it. In the absence of evidence of classifier usage of cognates in other branches of East Benue-Congo, e.g. Plateau, Dakoid, Mambiloid, Tivoid and Beboïd, it rather seems unlikely.

classifier systems with almost full lexical coverage are found. However, such a view of noun class compensation turns out to be too simplified for three reasons. First, languages can do very well with neither noun classes nor numeral classifiers. While some Eastern Mande languages such as Busa, Boko, Bokabaro and Gban have innovated numeral classifiers, other branches of Mande are neither found to innovate classifiers nor do they present any evidence of a prior noun class system at all. Second, decay and loss of prior noun class systems does not automatically entail the emergence of classifiers in compensation, as in various other branches and sub-branches of Niger-Congo such as most of Kwa and Adamawa, Dogon, Ijo, Kru and various sections of Benue-Congo (Williamson & Blench 2000). Third, numeral classifiers also arise in languages which retain a prior noun class system such as Ejagham, Denya, Ngiemboon, Yemba and Isu, allowing for a coexistence of two alternative systems of nominal classification. The remarkable point, however, is that lexical coverage in these latter cases is very imbalanced, i.e. the restricted lexical coverage of these classifier systems betrays their incipient stage. Thus, while the presence of a prior noun class system does not rule out the possibility of innovation of a competing system of nominal classification such as numeral classifiers, the full proliferation of a numeral classifier system seems to require the decline and loss – or absence – of a prior noun class system after all. So the basic motivation underlying the development of numeral classifier systems in these cases has to do with notions such as countability and individuation. Since the prior noun class system in most Niger-Congo languages includes a conflation of noun class and nominal number up to the point that a nominal number distinction outside the class distinction does not exist (Schadeberg 2001), the collapse of this noun class system often entails the loss of the nominal number distinction as well which could be seen most clearly in languages such as Bali-Mungaka. As is characteristic of the Nun subgroup of Eastern Grassfields, Bali-Mungaka has reduced the inherited Bantoid noun class system radically to six distinct classes grouped in six genders with the pronominal possessive as the only remaining agreement target (Hombert 1980: 153ff.; Tischhauser & Stöckle 1993: 3–6, 14–16) as indicated in Table 7. In a noun class system on the verge of elimination such as this, the nouns themselves could be seen to lose their formal distinction of singular vs. plural, e.g. nouns with a homorganic nasal prefix *N-* and nouns with a zero prefix have become vague with respect to number. Only a dwindling minority of nouns, those with prefix *mV-* (class 1 SG.) and those with prefix *bV-* (class 2 PL.), remain specific with respect to number.

Against this backdrop, the innovation of numeral classifiers seems to compensate, to some extent, for the loss of number specification caused by the breakdown of the prior noun class system and restore explicit marking of individuation in counting contexts. Yet, the question remains why there should be a strong necessity

Table 7. Agreement classes and their targets in Bali-Mungaka

reference number	singular-plural-pairing (gender)	agreement in pronominal possessives	noun prefix
1 SG.	1/2	-ʼ	<i>N-</i> , <i>mV-</i> , \emptyset
2 PL.	1/2, 3/2, 7/2	<i>b-</i> ʼ	<i>bV-</i> , \emptyset
3 SG.	3/2, 3/6	-ʼ	<i>N-</i>
5 SG.	5/6	<i>t-</i> ʼ	\emptyset
6 PL.	3/6, 5/6, 7/6	<i>m-</i> ʼ	<i>N-</i>
7 SG.	7/2, 7/6	-ʼ	\emptyset

in some Niger-Congo languages for explicitly marking individuation which needs compensation once the noun class system with its singular/plural distinction has vanished, while there is obviously no such tendency in others, e.g. many Kwa and Western Benue-Congo languages.

Another puzzle pertains to the diversity of numeral classifier systems. In systems with almost complete lexical coverage, classifier contrasts range between a maximum of 19 (Kana) and a minimum of four (Busa). Systems with restricted lexical coverage do not indicate a quantum leap: their contrasts range between a maximum of 15 (Bafanji) and a minimum of five (Ejagham, Denya). While this seems to suggest that the degree of elaboration of classifier systems does not crucially depend on their lexical coverage, it does not provide any clue as to the reasons of their diversity.

In order to fully understand and explain the ultimate motivations for the independent emergence of numeral classifiers and their diversity, certainly a broader approach in the sense of coevolutionary linguistics is needed, where cognition and culture are considered in their interrelation. Along these lines, fine-grained studies of the cultural and economical background of classifier usage will be necessary in order to reveal practices and preoccupations which favour the emergence of numeral classifiers in some semantic domains rather than in others. In the absence of such studies, however, the classifier systems presented so far provide some strands of evidence which point to potential motivating factors: language contact and additional semantic notions conflated in classifiers.

In the case of Bali-Mungaka and Bafanji, it might be argued that contact to languages which retain a full-fledged noun class system might stimulate the emergence of classifiers in compensation of a fading singular/plural distinction in nouns. This does not seem to hold, however, for cases such as Kana, Busa and Toussian.

The initial motivation for the emergence of classifiers might reside in the explicit marking of additional semantic notions such as definiteness, quantitative

limitation and contrastive emphasis on individuation or singularity. Thus, Isu classifiers *ísí* ‘eye’ (1g) and *fátúw* ‘little head’ (5a) conflate haptic notions with the notion of a contra-expected deficiency in number. In a similar vein, in Toussian (Zaugg-Coretti 2005: 22–25) the application of the diminutive classifier *plú* ‘grains’ (49b) rather than the expected inanimate classifier *yí* ‘things’ (49a) indicates an emphasis on limitedness, separateness or contra-expected low number.

(49) *Toussian: semantic notions of numeral classifiers*

- a. *tén-nó yí nín*
 mortar-PL CLF:thing.PL two
 ‘two mortars’
- b. *tén-Ø plú nín-nò*
 mortar-SG CLF:grain.PL two-PL
 ‘two mortars (only)’

With the Toussian item *íkpa* ‘child’, the selection of the human classifier indicates that the set is complete, the possibility of the mother having more children is excluded (50a), whereas the use of the inanimate classifier suggests that she could continue having more children (50b) (Zaugg-Coretti 2005: 22).⁸

(50) *Toussian: semantic notions of classifiers*

- a. *à íkpà nów tó wè téré*
 3SG child.PL CLF:human.PL three find 3SG.EMPH
 ‘She has born him three children (and it stops there).’
- b. *à íkpà yí tóó wè téré*
 3SG child.PL CLF:inanimate.PL three find 3SG.EMPH
 ‘She has born him three children (and she could have more).’

The importance of semantic conflation of haptic notions with notions such as deficiency in the context of numerals is also manifest in the classifying numerals for ‘one’ in Bali-Mungaka (45), a phenomenon which might be more widespread and definitely needs more attention and research.

Denya evidence suggests that classifiers might have emerged as one particular strategy to mark definiteness or specificity in nouns of certain semantic domains such as plants and products of plants such as fruits, tubers, nuts and grains. While the classifiers in counting contexts such as (51) are used to refer to definite or specific plantains or mangos, their absence in (52) renders the reference indefinite.

8. It is not clear if the diminutive classifier could be used in this context to produce the notion of a contra-expected low number or whether this notion is included in (50a).

(51) *Denya*: definiteness indicated by classifiers *gèlò* (7/8) “finger” and *gèlòó* (7/8) “fruit”

- a. *gè-lò* *gé-↓kwá* *gé-mâ*
7-CLF:finger ASS.7:7-plantain 7-one
“one plantain”
- b. *ù-lò* *ù-↓kwá* *ù-péá*
8-CLF:finger ASS.8:8-plantain 8-two
“two plantains”
- c. *ń* *↓só* *gè-lòó* *gé* *mè-fi* *gé-mâ*
1SG pick 7-CLF:fruit ASS.7 3-mango 7-one
“I have picked one mango.”
- d. *ń* *↓só* *ù-lòó* *mè-fi* *ù-péá*
1SG pick 8-CLF:fruit ASS.8:3-mango 8-two
“I have picked two mangos.”

(52) *Denya*: absence of classifiers indicates indefinite use of nouns

- a. *mëndèè wà* *à* *tyèé* *gé-↓kwá*
wife 1SG.PS SBJ.1 cook 7-plantain
“My wife has cooked plantains.”
- b. *ù-↓kwá* *ù ↓kwé*
8-plantain 8 fall
“Bunches of plantains have fallen (here and there).”
- c. *ń* *↓só* *mè-fi*
1SG pick 3-mango
“I have picked mangos (of one sort).”
- d. *ń* *↓só* *mà-fi*
1SG pick 6-mango
“I have picked several sorts of mangos.”

In this respect it is worthwhile to investigate to what extent the emergence of classifiers for functions such as definiteness and/or specificity is related to the loss of the augment in Benue-Congo which covers comparable functions (de Blois 1970; Hyman 2005; Williamson 1993, 2000).

While it is not clear which ultimate motivations account for the genesis of numeral classifier systems in Niger-Congo languages, the most promising hints point to a scenario of their innovation as carriers of notions such as definiteness/specificity and limitation or deficiency. Once classifier nouns have thus become established, the original notions of definiteness/specificity and limitation or deficiency might fade out in the course of semantic generalization, leaving a genuine classifier system as a side effect.

7. Conclusion

The numeral classifier systems presented above are no exceptional isolates, but rather form the tip of an iceberg which is yet to be explored. Their broader attestation, either in coexistence with a fully fledged noun class system or superseding a residual noun class system, calls for a major revision of Africa's linguistic landscape with respect to nominal classification.

Since almost all numeral classifier systems represent rather recent innovations, both the etymology of many classifiers and their source constructions remain transparent or could be retrieved fairly easily. With respect to lexical source concepts, classifiers are found to originate in nouns for concrete objects such as body parts or in basic level terms, most of which relate to the botanical domain. Syntactically, the classifier generally occupies the same position with respect to the enumerated noun as the possessed noun with respect to the possessor in an associative possessive construction, betraying the origin of the classifier in a noun which acts as head of an associative construction. While the classified noun takes the position of a modifier, the numeral comes in peripheral position, agreeing with the noun class of the classifier head noun in languages with a coexistent noun class system. In these syntactic and semantic properties, classifier constructions are very similar to associative constructions which display a dependency reversal, frequently found in Bantoid and Bantu languages of Cameroon. In those constructions, there is a mismatch of syntactic and semantic dependency, qualitative nouns such as Isu *fə̀nàŋə̀* (19/6a) “small thing” (53a) and *kándáŋ* (7/8) “tall thing” (53b) take the position of syntactic heads in associative constructions, but rather act as qualifiers or modifiers on the semantic level in that they encode a quality or the manifestation or embodiment of a quality, whereas the syntactically modifying (possessor) nouns, e.g. *↓ndáw* “house” and *ísòŋ* “tooth”, rather encode more specific concepts and must therefore be considered the semantic heads (van der Velde 2009).

(53) *Isu: Associative constructions with dependency reversal*

$[N_1]_{\text{QUALITY}} \text{ PP}_{X_1} [N_2]$

- a. *fə̀nàŋə̀* (19/6a) “small thing”

nàŋə̀ *fə̀* *↓ndáw*

19.small.thing 19 9.house

“tiny house”

- b. *kándáŋ* (7/8) “tall thing”

ndáŋ *k-ísòŋ*

7.tall.thing 7–5.tooth

“long tooth”

In constructions like these the qualitative noun in head position plays a role which is similar to that of a classifier both syntactically and semantically.

Since numeral classifier constructions inherit properties of the associative predecessor constructions, type A languages with an NP-internal word order of possessed noun + possessor noun + numeral (i.e. most Benue-Congo languages, except Ngwe, Ngiemboon and Yemba) regularly produce the word order classifier (noun) + enumerated noun + numeral in emergent classifier constructions, violating the generalization of non-separability of classifier and numeral (Allan 1977; Greenberg 1972; Dixon 1986). Moreover, it is these Benue-Congo languages which provide morphophonological, morphotonological and morphosyntactic evidence showing that it is the classifier and the enumerated noun which form a morphological unit rather than the classifier and the numeral – which had been taken as the only permissible possibility of fusion (Dixon 1986). As could be seen here, the general explanation which accounts for both – the allegedly non-permissible separation of classifier and numeral by the enumerated noun and the allegedly non-permissible fusion of classifier and enumerated noun – simply resides in two fundamental differences vis-à-vis the classifier constructions cited in Greenberg (1972), Allan (1977), and Aikhenvald (2000): a difference in constituency relations prevailing in most Benue-Congo classifier constructions and, perhaps most of all, in the presence of coexistent noun class systems which provide for explicit marking of discontinuous constituents (such as numerals) by agreement.

Numeral classifier systems in Niger-Congo languages are fairly underresearched and have escaped the attention of linguists to a large extent so far. This is due to three interrelated factors. First, the majority of Niger-Congo languages still lack adequate documentation and analysis. Second, due to the functionally competing noun class system, most emergent numeral classifiers in Niger-Congo languages tend to be restricted to certain semantic domains, e.g. plants. Last but not least, the inadequate state of documentation is partly due to the long-standing dominance of the paradigm the inherited noun class system in Niger-Congo has imposed and still imposes on the practice of language documentation. Sketch grammars and standard first descriptions of African languages do not normally include sections on how to count small roundish or longish objects. Therefore it will be one of the future tasks to investigate Niger-Congo languages for the way they handle counting objects of specific semantic domains in order to check them for their incipient numeral classifier systems.

Since the numeral classifier systems in Niger-Congo present fairly transparent embryonic stages of nominal classification, the cognitive models employed in their genesis might contribute, in the long run, to the understanding of the emergence of the classical Niger-Congo noun class system by providing clues both to the ultimate lexical sources and the syntactic structures from which the classical Niger-Congo noun class system might have been developed (Claudi 1997).

Acknowledgments

The present paper has profited greatly from discussions with Tom Güldemann, Maarten Mous and Thilo Schadeberg as well as from feedback from the editors, Bill McGregor and Søren Wichmann, and two anonymous reviewers. In particular, I want to express my gratitude to the DFG (German Research Foundation) whose generous grants enabled the research on which this contribution is based, to Philip Nguessimo Mutaka and Pius Tamanji for their effective administrative help and their hospitality in Yaoundé and Cameron and Valerie Hamm, Jesse Lovegren, Viktor Vinogradov and Alexander Zheltov for sharing their data on Bafanji, Ngwe, Ngiemboon and Gban, respectively. Moreover, I am deeply grateful to Samson Negbo Abangma who has provided the data on Denya and to Bong Marcellus Wung of Isu who has untiringly helped to clarify the grammatical and semantic intricacies of the use of the noun *ísí* “eye” in his mother tongue. Last, but not least, I am very grateful to Britta Neumann who assisted in eliciting additional Isu data on my behalf. All mistakes remain my own.

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith. 1977. “Classifiers”. *Language* 53.285–311. doi:10.1353/lan.1977.0043
- Anderson, Stephen. 1983. *Tone and Morpheme Rules in Bamileke-Ngyemboon*. Ph.D. dissertation, University of Southern California.
- Beyer, Klaus. 2006. *La langue pana*. Köln: Rüdiger Köppe.
- Bisang, Walter. 1999. “Classifiers in East and Southeast Asian Languages. Counting and beyond”. *Numeral Types and Changes Worldwide* ed. by Jadranka Grozdanovic, 113–185. Berlin: Mouton de Gruyter. doi:10.1515/9783110811193.113
- Claudi, Ulrike. 1997. “Some Thoughts on the Origin of Gender Marking”. *African Linguistics at the Crossroads: Papers from Kwaluseni, 1st World Congress of African Linguistics, Swaziland, 18–22. VII. 1994*. ed. by Robert K. Herbert, 63–74. Köln: Rüdiger Köppe.
- Craig, Colette. ed. 1986. *Noun Classes and Categorization*. Amsterdam: John Benjamins. doi:10.1075/tsl.7
- Craig, Colette. 1994. “Classifier Languages”. *The Encyclopedia of Language and Linguistics* ed. by Ronald E. Asher & James M. Y. Simpson, Vol. II, 565–569. Oxford: Pergamon Press.
- de Blois, Kornelis Frans. 1970. “The Augment in the Bantu Languages”. *Africana Linguistica* 4.87–165.
- de Wolf, Paul. 1971. *The Noun Class System of Proto-Benue-Congo*. The Hague: Mouton.
- Dimmendaal, Gerrit. 2011. *Historical Linguistics and the Comparative Study of African Languages*. Amsterdam: John Benjamins. doi:10.1075/z.161
- Dixon, R. M. W. 1986. “Noun Classes and Noun Classification in Typological Perspective”. *Noun Classes and Categorization* ed. by Colette Craig, 105–112. Amsterdam: John Benjamins. doi:10.1075/tsl.7.09dix
- Gil, David. 2005. “Numeral Classifiers”. *The World Atlas of Language Structures*. Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie, eds. 2005, 226–229. Oxford: Oxford University Press.
- Greenberg, Joseph H. 1966. *The Languages of Africa* (2nd edition). Bloomington & The Hague: Mouton & Co.

- Greenberg, Joseph H. 1972. "Numeral Classifiers and Substantival Number: Problems in the genesis of a linguistic type". *Working Papers in Language Universals* 9 .1–39. [Reprinted in Greenberg (1990: 166–193)].
- Greenberg, Joseph H. 1990. *On Language: Selected writings of Joseph H. Greenberg*. Stanford: Stanford University Press.
- Grinevald, Colette & Frank Seifart. 2004. "African and Amazonian Noun Classes: Towards a comparison". *Linguistic Typology* 8.243–285. doi:10.1515/lity.2004.007
- Güldemann, Tom. 2008. "The Macro-Sudan Belt: Towards identifying a linguistic area in northern sub-Saharan Africa". *A Linguistic Geography of Africa* ed. by Bernd Heine & Derek Nurse, 151–185. Cambridge: Cambridge University Press.
- Hamm, Cameron. 2010. "Numeral Classifiers in Bafanji (Nun, Eastern Grassfields)". Ms., SIL Cameroon.
- Harro, Gretchen & Nancy Haynes. 1991. *Grammar Sketch of Yemba*. Yaoundé: SIL.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie, eds. 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Heine, Bernd. 1976. *A Typology of African Languages*. Berlin: Dietrich Reimer.
- Hombert, Jean-Marie. 1980. "Le groupe noun". *L'Expansion bantou I* ed. by Luc Bouquiaux, 143–163. Paris: SELAF.
- Hyman, Larry. 2005. "Initial Vowel and Prefix Tone in Kom: Related to the Bantu augment?" *Studies in African Comparative Linguistics with Special Focus on Bantu and Mandé* ed. by Koen Bostoen & Jacky Maniacky, 313–341. Tervuren: Royal Museum for Central Africa.
- Ikoro, Suanu M. 1994. "Numeral Classifiers in Kana". *Journal of African Languages and Linguistics* 15.7–28. doi:10.1515/jall.1994.15.1.7
- Ikoro, Suanu M. 1996. *The Kana Language*. Leiden: CNWS.
- Jones, Ross M. 1998. *The Boko/Busa Language Cluster*. München: Lincom.
- Jones, Ross M. 2004. *Busa dictionary*. München: Lincom.
- Kähler-Meyer, Emmi. 1941/42. "Märchen in der Balisprache aus dem Grasland von Kamerun". *Zeitschrift für Eingeborenen Sprachen* 32.135–160, 224–236.
- Katamba, Francis. 2003. "Bantu Nominal Morphology". *The Bantu Languages* ed. by Derek Nurse & Gérard Philippson, 103–120. London: Routledge.
- Kaufman, Elaine Marlow. 1972. *Ibibio Dictionary*. Washington, D.C.: Office of Education.
- Lovegren, Jesse. 2010. "Notes on Ngwe Numeral Classifiers". Ms., University at Buffalo.
- Manessy, Gabriel. 1961. "Le Bwamu et ses dialectes". *Bulletin de l'Institut Fondamental de l'Afrique Noire* 23.119–178.
- Mbuagbaw, Tanyi Eyong. 1996. *Denya phonology*. Yaoundé: Cameroon Bible Translation Association.
- Moritz, Luc. 1994. "The Facts of Nchufie DP's: A report". *Aspects of Nchufie Grammar* ed. by Hilda Koopman & Murat Kural, 91–113. Los Angeles: University of California, Los Angeles.
- Mous, Maarten. 1999. "Numeral Classifiers in Toussian: First preliminary version". Ms., Leiden University.
- Pasch, Helma. 1985. "Possession and Possessive Classifiers in 'Dongo-ko'. *Afrika und Übersee* 68.69–85.
- Pasch, Helma. 1986. *Die Mba-Sprachen: Die Nominalklassensysteme und die genetische Gliederung einer Gruppe von Ubangi-Sprachen*. Hamburg: Helmut Buske.
- Prost, André. 1969. "Le pana". *Bulletin de l'Institut Français de l'Afrique Noire B* 31:3886–911.
- Rowland-Oke, Mary. 2003. *Description systématique de la langue Obolo-Andoni – langue du groupe Cross River*. Paris: L'Harmattan.
- Schadeberg, Thilo C. 2001. "Number in Swahili Grammar". *Afrikanistische Arbeitspapiere* 68 (= Swahili Forum, VIII).7–16.

- Tischhauser, Georg & Johannes Stöckle. 1993. *Mungaka (Bali) Dictionary*. Köln: Rüdiger Köppe.
- van der Velde, Mark. 2009. "Two Non-Prototypical Uses of the Bantu Connective Construction". Paper presented at the WOCAL 6, Cologne 17–21.08.2009.
- Vinogradov, Viktor. 2009. "Strange Kinds of Nominal Classification in some West African Languages". Paper presented at the WOCAL 6, Cologne 17–21.08.2009.
- Watters, John R. 1981. *A Phonology and Morphology of Ejagham, with Notes on Dialect Variation*. Ph.D. dissertation, University of California, Los Angeles.
- Wedekind, Klaus. 1972. *An Outline of the Grammar of Busa (Nigeria)*. Ph.D. dissertation, University of Kiel.
- Welmers, William E. 1973. *African Language Structures*. Berkeley: University of California Press.
- Williamson, Kay. 1993. "The Noun Prefixes of New Benue-Congo". *Journal of African Languages and Linguistics* 14.29–45. doi:10.1515/jall.1993.14.1.29
- Williamson, Kay. 2000. "Towards Reconstructing Proto-Niger-Congo". *Proceedings of the 2nd World Congress of African Linguistics, Leipzig 1997* ed. by H. Ekkehart Wolff & Orin Gensler, 49–70. Köln: Rüdiger Köppe.
- Williamson, Kay & Roger Blench. 2000. "Niger-Congo". *African Languages: An introduction* ed. By Bernd Heine & Derek Nurse, 11–42. Cambridge: Cambridge University Press.
- Zaugg-Coretti, Siliva. 2005. *Le syntagme nominal en toussian du nord (langue voltaïque du Burkina Faso)*. M.A. thesis (Lizenziatsarbeit), University of Zürich.
- ZheltoV, Alexander. N.d. "Numeral Classifiers in Gban (South Eastern Mande, Niger-Congo) – preliminary remarks". Ms., St. Petersburg State University .

Appendix

This appendix presents a paradigmatic overview of the numeral classifier systems discussed in the text.

Bafanji (Grassfields) (Hamm 2010)

form	lexical source	range of application
<i>ndí</i> (9/6)	'eye'	small, non-round objects, seeds, insects
<i>téé</i> (7/6)	'fruit'	whole fruit, round objects, drop of liquid
<i>mbóruú</i> (9/6)	'bundle'	bundle (wrapped, not tied)
<i>póruú</i> (7/6)	'bundle'	bundle (that is tied with string)
<i>túró</i> (?/?)	?	edible, made to be round
<i>tjórú</i> (?/?)	?	group
<i>ndú</i> (?/?)	'pod'	shell with several smaller fruits inside
<i>tjúrú</i> (?/?)	?	stem/stalk with seeds
<i>ŋkúrú</i> (7/6)	?	bare, rigid plant part
<i>ŋkèé</i> (9/6)	?	all fruit including stem from which it grows
<i>pjé</i> (7/6)	?	group of fruits that form together
<i>tjágí</i> (7/6)	?	blade
<i>ŋgàà</i> (?/?)	?	twig, small stem
<i>kú</i> (7/6)	?	heap, pile
<i>kàí</i> (?/?)	?	thick clump

Bali-Mungaka (Grassfields) (Tischhauser & Stöckle 1993)

form	lexical source	range of application
<i>líʔ</i> (PL. <i>míʔ</i>)	‘eye’	small roundish object
<i>ntàm</i>	?	single oblong object
<i>túʔ</i>	‘unformed lump of clay’	lumpy, bulging or clotted objects

Busa (East Mande) (Wedekind 1972: 167; Jones 2004)

form	lexical source	use as classifier
<i>gbē</i>	‘person’	human (singular)
<i>gʒrʒn</i>	‘people’	human (plural)
<i>mè</i>	‘body’	object (non-human)
<i>wén</i>	‘eye’	small roundish objects (also: curved or curly shape): grains, points of arrows, hair

Denya (Nyang) (Abangma & Kießling)

form	lexical source	range of application
<i>ɲɲi / àmé</i> (5/2)	‘eye’	grains, seeds, kernels, nuts (small globular organic materials)
<i>gèlò / ùlò</i> (7/8)	‘finger, pod’	fruits of oblong shape
<i>ndòó</i> (9/10)	‘bunch, cluster’	fruits which come in distinctive clusters
<i>gèlòó / ùlòó</i> (7/8)	‘fruit’	any other fruits

Ejagham (Ekoid-Bantu) (Watters 1981: 310–313)

form	lexical source	range of application
<i>è-sán / Ñ-sán</i>	‘tree trunk’	trees
<i>Ñ-dáb / à-ráb</i>	‘bottom’	trees, shrubs, vegetables
<i>Ñ-màgè / à-màgè</i>	‘small globular object’?	grains, seeds, kernels, nuts
<i>Ñ-sùm / à-sùm</i>	?	fruit or root which is long
<i>è-rám / Ñ-dám</i>	?	fruit or root which is round or in a cluster

Gban (South Eastern Mande) (Zheltev n.d.)

form	lexical source	use as classifier
<i>ví</i>	‘person’	human; default
<i>tí</i>	‘tree’?	objects of oblong shape with vertical orientation
<i>wlí</i>	‘fruit’	objects of roundish shape
<i>lã</i>	‘leaf’	objects of flat shape
<i>kwě</i>	‘stick’	objects of elongated shape, diminutive
<i>míã</i>	‘board, plank’	objects of flat and elongated shape

Isu (Grassfields) (Kießling)

form	lexical source	range of application
<i>ísí</i> (5/6)	‘eye’	single globular units
<i>kówó</i> (7/8)	‘hand’	units with oblong extremities
<i>íyé</i> (5/6)	‘breast, udder’	bunches of objects which are recognizably subdivided into clusters
<i>fātúw</i> (19/6a)	‘little head’	items with round flat shape

Kana (Cross River) (Ikoro 1996: 89ff.)

form	lexical source	range of application
<i>bēē</i>	‘fruit’	bulb-like objects, seed-like objects, offshoot, offspring, abstract
<i>té</i>	‘tree’	wooden objects > vehicles of transportation
<i>kūm</i>	‘stock’	objects of a stock or a bunch
<i>kéré</i>	‘half’	objects segmented into parts
<i>ηwíí</i> (PL. <i>míóηó</i>)	‘child’	young animates (without reference to birth positions)
<i>kō</i>	‘bunch’	objects of bunch- or package-form
<i>kà</i>	‘mother’	default classifier for “entities, animals and things”
<i>ákpó</i>	‘length’	inanimate objects which possess a trunk
<i>ápéé</i>	‘piece of’	pieces of objects, splashes of liquids
<i>ásúú</i>	‘grain’	individual units extracted from a bunch
<i>àkpò</i>	‘heap’	objects with irregular shape which come in heaps or chunks
<i>ábá</i>	?	objects with tube-like shape
<i>àbà</i>	‘part’	potentially separable parts of objects
<i>sś</i>	‘time’	periodically recurrent event
<i>děē</i>	‘eye’	spot- or dot-like objects
<i>ákpá</i>	‘skin’	objects of flat shape
<i>nēē</i>	‘person’	marginal, only used for ‘guest’
<i>àpā</i>	?	conically-shaped objects
<i>ákpé</i>	?	objects with vertically high extension

Ngwe (Grassfields) (Lovegren 2010)

form	lexical source	range of application
<i>àtàá</i> (7?/8?, PL. <i>ètáá</i>)	?	round objects
<i>nífóɔ</i> (9?/?)	?	long objects, objects which hang in bunches
<i>ēndrò</i> (?/?)	?	some unpeeled fruits
<i>ēngà</i> (?/?)	?	whole bunches
<i>mbāη</i> (?/?)	‘palm kernel’	bitter kola kernels
<i>ályz</i> (7?/6)	?	bunches, bundles, chunks, lumps
<i>mbàm</i> (?/?)	?	grains, small insects

Ngiemboon (Grassfields) (Vinogradov 2009)

form	lexical source	range of application
<i>mbáŋ</i>	?	insects?
<i>mvôm</i>	'grain'	vegetables
<i>ntaʔ</i>	?	oblong tools?
<i>tɔn</i>	?	globular fruits
<i>ndɔŋ</i>	?	oblong fruits
<i>mb'o</i>	'hand'	leaves, oblong parts of plants?
<i>tʰò</i>	'head'	big globular fruits

Pana (Gur) (Prost 1969: 894–895)

form	lexical source	use as classifier
<i>la</i>	'man'	humans
<i>ka</i>	'wife?'	non-humans, inferior humans

Toussian (Gur) (Zaugg-Coretti 2005: 21–26)

singular	plural	lexical source	use as classifier
<i>náy</i>	<i>nõw</i>	'person'	human
<i>kpâ</i>	<i>yí</i>	'thing'	non-human, inanimate
<i>pûe</i>	<i>plú</i>	'grain'	non-human, inanimate, small, objects of roundish shape, small animals, roots, tubers, leaves, small oblong objects
<i>yî</i>	<i>îyô</i>	'eye'	measurements (time, basket)

Win / South Toussian (Gur) (Mous 1999)

singular	plural	lexical source	use as classifier
<i>nɛ</i>	<i>nɛpî</i>	'person'	humans, anthropomorphized animals
<i>pɔ́</i>	<i>plá</i>	?	discrete mobile entities

Semantic generalization in Ch'orti' Mayan numeral classifiers

Kerry Hull

Brigham Young University

This study examines the process of linguistic change in the classification system of Ch'orti' Mayan, a language spoken by roughly 12,000 people in southern Guatemala. I trace the origin of the classifier *-kojt* from its original meaning of “quadruped”, based on the positional root *kojt*- “kneel down”, to all animals in general; then (somewhat unexpectedly) to include certain foods; and most recently to humans. I also describe the impact of ideologically driven decisions by those involved in the Pan-Maya Movement in the Ch'orti' area relating to the use of *-kojt*, namely in accelerating its natural semantic broadening to now include all human referents to compensate for the lack of such a classifier in the language. I conclude that this new use of *-kojt* represents both a case of standard semantic generalization as well as a case of hypercorrection due to revised notions of language ‘correctness’.

1. Introduction

The Ch'orti' Mayan language spoken in southern Guatemala has a rich system of numeral classifiers (NC), both sortal and mensural. In the last century, however, several sortal classifiers have undergone a semantic shift due to both language internal and external influences. In this study I trace this movement towards semantic generalization through an analysis of all extant written Ch'orti' documents, in addition to data gathered during nearly three years of my own fieldwork among the Ch'orti'. Based on this analysis, I identify the developmental steps in the process of change of two of the most common NCs, *-te'* and *-kojt*, and identify many of the factors motivating their semantic shifts. Finally, I discuss the role of modern language revitalization groups in facilitating these semantic extensions for reasons relating to idealized notions of ‘language purity’.

Ch'orti' is part of the Eastern Branch of the Ch'olan family of Mayan languages and is one of its three remaining members since the extinction of Ch'olti' around the seventeenth century. Ch'orti' is spoken by roughly 12,000 speakers

today – a considerable decrease in number within the last hundred years. The influence of Spanish on Ch'orti' since the Conquest has penetrated into most facets of the language, particularly in the commonly found SVO word order in Ch'orti' today – a departure from other Mayan languages that tend to be VSO or VOS. Lexical borrowing has been consistent and is particularly salient in highly frequent adverbial phrases such as *entonces*, *pues*, *porque*, and others. These borrowings have also had an effect on Ch'orti' syntax as speakers often attempt to conform to Spanish sentence structures. Since the 1990s, however, efforts have been underway to 'purge' the Ch'orti' language of many of these borrowed terms.

In the last two decades, the Ch'orti' language has experienced a considerable surge in published material, grammatical descriptions, and oral tradition documentation by the Ch'orti' themselves. The attention being paid to documenting and describing the Ch'orti' language is part of a larger movement within Guatemala to redefine the identity of individual Maya groups. The Pan-Maya Movement, begun in the late 1980s, has resulted in the consolidation of certain universal ideals in disparate Maya communities that represent the first steps in an effort to create a unified 'Maya voice' in Guatemala. One of the key aspects of the Pan-Maya Movement is the preservation and promotion of Mayan languages. Thus, professionally trained native speaker linguists are actively producing linguistic sources as a means of solidifying some of these faltering languages and increasing the language resources available to speakers.

In addition, the proliferation of pedagogical materials in Mayan languages as well as school instruction in these languages is a major component of these revitalization efforts. One common approach to accomplishing their linguistic goals is the replacing of borrowed Spanish terms and the creation of Mayan neologisms. Some Ch'orti' involved in this process have described this same need to 'purify' their language, even of Ch'orti' forms that are perceived to be 'incorrect'. The natural consequence of this ideology is that spoken forms are replaced with idealized 'correct' forms in what amounts to a language-wide house cleaning of unwanted terms. NCs in Ch'orti' have been affected by these redactive processes in several important ways.

Since NCs are not fundamentally important to meaning in languages, Ch'orti' speakers of the last generation have steadily reduced the number of active NCs in their repertoire. Indeed, young Ch'orti' today often know no more than 3–5 classifiers (sortal and mensural), and the trend is certainly towards simplification, generalization, or elision.¹ As I have observed it, the use of a wide range of NCs is a

1. Nick Hopkins (personal communication, July 2010) has noted a similar phenomenon in Ch'ol, where there seems to be a move towards using only the most general classifiers today by many speakers, i.e., *tyikil* "people", *kojt* "animals", *tyejk* "plants", and *p'ejl* "things (everything else)". In fact, for some people, there are only two classifiers: *tyikil* and *p'ejl*.

marker of a competent native speaker in most cases, whereas those who have lesser abilities tend to use far fewer, relying either on the general classifier *-te'* to a larger extent or using no classifier at all.

The degradation of the NC system is in part a result of near total use of Spanish numbers when speaking Ch'orti'.² Only the numbers 1–4 (*jun*, *cha'*, *ux*, and *chan*) are still in use in spoken Ch'orti' today. A full set of indigenous terms was once part of the language, but the use of Spanish numbers has rendered the original Ch'orti' counting system nearly obsolete, with only numbers 1–3 regularly used, and 4 much less commonly so. Since Spanish numbers do not fit the morphemic structural patterns of Ch'orti' numerals in the formation of NCs, they are not usually used with NCs. For example, in Ch'orti', “one bunch of string” would be expressed as

- (1) *inpoj* *sukchij*
 in-pojk sukchij
 one-NC(BUNCH) string
 “a bunch of string”.

However, since the numerals in Ch'orti' are proclitics and Spanish numerals are not, one could not use the Spanish *un* “one”, to produce **unpoj* *sukchij*. This restriction discourages the use of native Ch'orti' classifiers with all numbers over 4, thereby contributing significantly to their decline in daily speech.

2. Numeral classifiers in Ch'orti'

Most Mayan languages have a system of NCs, but their use and stages of development vary. Kanjob'alan languages, such as Jakaltek, also have noun classifiers in addition to numeral classifiers. Craig has convincingly argued that the noun classifier system of Jakaltek is in fact quite a recent development since it shows none of the morphologization found in older systems (1986: 281; cf. Grinevald 2004: 1027). The late development of many aspects of both the elaborate sortal numeral classifications systems and mensural classifiers found in numerous Mayan languages today is also evidenced epigraphically. Hieroglyphic writing of the Late Classic period shows the existence of a limited set of numeral classifiers in use. As Lacadena (2010) has pointed out, the Spanish conquest brought the Maya into contact with many

2. Shklovsky (2005: 40) has noted a similar phenomenon occurring with the NCs in Tzeltal: “Much of the system is falling into disuse due to the fact that many communicative acts requiring numbers (such as buying and selling) are transacted in Spanish, and in many cases the generic classifier is replacing more specific ones.”

new objects that needed classification; others that conflicted with their traditional classification forced changes to that system.

For the most part, NCs are monosyllabic in Ch'orti', the only exceptions being *-kab'an* "times" and *-b'akar* "ears of corn".³ The constituent order for numeral classifiers in Ch'orti' is [NUM-CL]-NOM. The only morphemes allowed to come between the NUM and the CL are adjectives and possessive pronouns. Most Ch'orti' numeral classifiers and mensurals are formed through the intransitivization of the root with the infix *-j-*, e.g., *-gojr* "roundish thing", which derives from the transitive root *gor-* "to roll". Whereas in some Mayan languages like Akatek where all sortal classifiers derive from positional roots (Zavala 2000: 129), Ch'orti' classifiers in general come from three different sources. Of the roughly 40 sortal and mensural classifiers I have identified in Ch'orti' (see Hull in preparation), the majority (about 80%) are derived from transitive roots, while the remainder come from positional or nominal roots.⁴

Sortal and mensural are the two primary types of classifiers in Ch'orti'. Following Craig (1992: 280), sortal classifiers are defined as "morphemes that specify units (not quantity) in terms of which the referent of the head noun can be counted, although they may be used in contexts other than quantification ... They often appear to be semantically redundant, expressing one of the inherent semantic characteristics of the head noun." Mensural classifiers, on the other hand, qualify mass nouns and therefore are part of all languages of the world. Mensurals cluster in distinctive features such as dimensionality, shape, and function, essentially expressing "the type of arrangement of the figure in units or containers that are necessary for mathematical operations" (Zavala 2000: 130). While dimensionality is especially relevant to differentiating classifiers, function and material make up also figure prominently into this division, especially in their interaction with the physical properties of the objects (Aikhenvald 2003: 292). Ch'orti' has more mensurals than sortal classifiers, though many mensurals have also fallen into disuse in the last 50 years.

Functionally, NCs are commonly used as anaphoric pronouns in Ch'orti'.

3. Wichmann (1999: 125, note 2) has cautioned that *-b'akar* may not be a NC but may rather be more comparable to the expression in English of "a glass of (water, etc.)." However, I have heard it used as a classifier in numerous cases. Note this sentence: *Ch'ama uxb'akar nar, jixi twa' e sa'* "Take three ears of corn (and) remove the grains for the *chilate* (drink)". Commonly *-b'akar* functions as a NC for the head nouns *a'n* and *nar* (both meaning "ear of corn"). When used as a noun by itself, however, *b'akar* refers to a corncob without any kernels.

4. There are another 20 possible classifiers mentioned in other sources that I have not been able to confirm in the field. With these added to the total, the percent of forms derived from transitive roots rises slightly to 85%.

- (2) *Ub'ajyu ayi uwira inkojt.*
 u-b'ajy-u ayi u-ir-a in-kojt
 3A-feel-THEME QUOT 3A-see-THEME one-NC(ANIM)
 "One of them touched it softly and looked at it."

While Ch'orti' has a fair distribution of both sortal and mensural classifiers, for the purposes of this study I focus on the development and use of only two sortal classifiers, *-kojt* and *-te'*.

3. The origin of *-kojt* and semantic generalization

The NC *-kojt* in Ch'orti' and its cognates in other Ch'olan languages derive from the Proto-Ch'olan **kot* "bent over, crouching; on all fours, kneeling" (cf. Kaufman & Norman 1984: 123).⁵ The form found in Ch'orti', *-kojt*, is derived from the positional root *kot-*, meaning "to kneel down". The infix *-j-* (= IPA [x]) is a derivational morpheme for creating intransitive roots. It is this type of CVjC construction that forms the base for most Ch'orti' classifiers, both sortal and mensural.⁶

A comparative search of relevant Mayan languages shows that the essence of the verbal root *kot-* is "to be down on all fours", originating with quadruped animals. Thus, in Ch'ol, according to Coon & Preminger (2011: 215), a cognate classifier, *-kojty*, also "appears to be derived from the positional root *koty*, [and is] used to describe the position of being on four limbs":

- (3) *Tyi i-tsāñ-s-ä ux-kojty wakax.*
 PRFV 3A-die-CAUS-TV three-NC(ANIM) cow
 "He killed three cows."

Elsewhere, Coon (2005: 7) notes that the NC *koty* in Ch'ol comes from the positional root *kot* "standing on four legs".

For Proto-Tzeltal-Tzotzil Kaufman (1972: 106) reconstructs **koht* meaning "animal". In Tzeltal, Shklovsky (2005: 40) also notes the use of the numeral classifier *koht* not only for "all animals" as in (4), but also for "cars".

5. In the scant resources left to us of Ch'olti', the now-extinct language that was closely related to Ch'orti', the NC *-kojt* does not appear. The most common NC is *-te'*, just as in modern Ch'orti' (cf. Robertson et al. 2010).

6. Only CVC roots ending in sibilants do not show the medial "j"; e.g., *-b'is*, "packet of", not **-b'ijs* (Hull in preparation). Coon (2005: 5) has noticed a parallel phenomenon in Ch'ol: "CVC roots in which the final C is a fricative [s], [ʃ], or [h] are unable to form intransitives with a lengthened aspirated vowel."

- (4) *juju-koht mut*
 every-NC chicken
 “every chicken”.

The further extension of meaning from “four-legged animal” to “four-wheeled cars” was a natural association based on shape. The use of *koht* with a chicken is an extension of meaning from animals on all fours to a two-legged animal.

A further semantic generalization can also be seen in the same root *kot* in Tzotzil, which Laughlin (1975: 178) describes as follows: “In reference to: all animals except man, machines on legs or wheels, harp, fiddle, guitar, table, chair, stool, metate, chili.”

Here we can see the gradual broadening of scope of this classifier from “four-legged animals” to the present day “all animals”, and then to tables and chairs with four legs by association, and finally to machines on legs or wheels. All of these meanings ultimately come from the verbs *kotih*, meaning to “stand on all fours”, and *kotkon*, “walking on all fours (child, devil, animal)” (Laughlin 1975: 178). The idea of “bent over” found in Proto-Ch’olan is also found in the entries: *koti* “rainbow”, *kotkot* “arched, bend (back, tree), curved (bill), hooked (nose)”, and *kotleh* “standing on all fours, bent over (person with age)”.

Hurley & Ruíz Sánchez (1978: 27) also note this root in Tzotzil as the intransitive verb *cotijel* “to crawl”, or as *cot* as a numeral classifier for “*animales o cosas de cuatro pies* (animals or things that have four legs).” García de León (1971: 53) provides a similar description for the meaning of the root *kot-* in Tzotzil: “*clasificador numeral de animales u objetos tetrápodos; vago. Ox-kot ts’i tres perros, jun-kot mexa una mesa; kot-ol en cuatro patas* (numeral classifier of animals or four-legged objects; *Ox-kot ts’i* “three dogs”, *jun-kot mexa* “one table”; *kot-ol* “on four legs”).”

Haviland likewise states that the NC *kot* can modify an “*animal, mesa, silla, carro, cosa a gatas* (animal, table, chair, car, something on all fours)” (1981: 372). Once again, we see the process of semantic generalization to all things that have four legs, be they animate or not.

In Ch’ol Schumann (1973: 31) notes that *-kojt* “*se usa para contar animales* (is used to count animals).” Attinasi’s Ch’ol dictionary supports this interpretation, although he records the form as *ko:t* as the numeral classifier for “animal” (1973: 281). Warkentin & Scott (1980: 112) state that in Ch’ol *-kojt* is used for “*un chile o un animal, un ave, una cosa con cuatro patas (mesa, carro, etc.)* (a chile or an animal, a bird, something with four legs [table, car, etc.]).” The meaning in Ch’ol has similarly extended to all animals (even two-legged birds) and four-legged objects such as tables, and also to four-wheeled cars (cf. Coon 2005: 7).

Based on the linguistic data in the Mayan languages presented above, we can tentatively reconstruct the gradual linear semantic extension of the positional root *kot* “bend down on fours” to a NC *-kojt* meaning (1) four-legged animals >

(2) animals in general (3) > non-animal objects with four 'legs' (4) > various other more general objects (harps, violins, guitars, chiles, etc.). However, the order of extension suggested here is not absolute since it is also possible (though I would think less likely) that Step 3 preceded Step 2.

4. Sources of linguistic data on Ch'orti'

Having established a progressive semantic generalization in related Mayan languages with a cognate to the Ch'orti' NC *-kojt*, we can now turn our attention to one of the main themes of this study – the semantic generalization that has occurred in the last century with *-kojt* in Ch'orti'. This evolution of meaning can be traced over time through an analysis of all early linguistic data compared to the use of *-kojt* in present-day spoken Ch'orti'. What will become clear here is that *-kojt* has undergone a considerable semantic broadening – a process still continuing today.

In order to properly contextualize this discussion, we must first review the primary sources for Ch'orti' linguistic data and what they tell us as to the use of *-te'* and *-kojt*.

Our earliest sources of Ch'orti' data come from a short wordlist compiled by Juan Galindo in 1834. In 1892, Alberto Ruano Suárez collected a modest list of Ch'orti' terms in Chiquimula, Guatemala. Karl Sapper also worked on lexical and grammatical documentation at this time and published part of his data in 1897. We also have a sizable list of about 566 words, some verbal paradigms, and a handful of sample sentences assembled by Alberto Membreño and published in 1897, though a portion of his work was lifted straight from Suárez's recently published wordlist at the time (Membreño 1897). Marcos E. Becerra also published a short list of Ch'orti' words in 1910. Later, one of the most important scholars in Ch'orti' studies, Rafael Girard, published an extensive Ch'orti' vocabulary in his monumental work *Los chortís ante el problema maya* (Vol. 1) in 1949 (Girard 1949). Girard, like Membreño, however, borrowed heavily from Suárez's work, but there is still very valuable new information not found in Suárez's published wordlist.

Charles Wisdom's prolific studies among the Ch'orti' from 1931 to 1933 culminated in his publication of *The Chorti Indians of Guatemala* (Wisdom 1940). His unpublished dictionary of Ch'orti' Maya is one of the largest sources of data from the early mid-twentieth century (Wisdom 1950). Then in the 1950s, Helen Oakley, a missionary with the Friends Mission, worked in the Ch'orti' area, helped to establish an alphabet to write Ch'orti', and made the first translation of parts of the New Testament into Ch'orti' in 1969. In 1966, her publication, *Chorti*, contained ethnographic data, grammatical descriptions, some short sample expressions, and a transcribed and translated text (the earliest-known published text transcription

in Ch'orti') (Oakley 1966). Also in the 1960s John Fought conducted considerable fieldwork among the Ch'orti', resulting in his Ph.D. dissertation *Chorti (Mayan) Phonology, Morphophonemics, and Morphology* (Fought 1967) and his all-important book *Chorti (Mayan) Texts* (Fought 1972).

In more recent times there have been numerous articles relating to Ch'orti' grammar (see Wichmann 1999 for further details on other Ch'orti' published sources). In 1989, John Lubeck and Diane Cowie published the first pedagogical grammar of the Ch'orti' language. Around that same time, Metz et al. (1992) produced a dictionary of Ch'orti' with over 1,000 terms that is still unpublished.

In the 1990s, however, with the burgeoning Maya Movement sweeping through parts of Guatemala, several groups were formed in the Ch'orti' area dedicated to the revitalization of the Ch'orti' language, promoting education in Ch'orti', furthering human rights, and improving the desperate economic situation of the Ch'orti'. With the establishment of such groups came the first linguistic sources about Ch'orti' created by the Ch'orti' themselves. The significance of this milestone in Ch'orti' studies can hardly be overstated.

The largest single source in the Ch'orti' language is the translation of the New Testament done by John Lubeck in conjunction with his Ch'orti' consultants, which appeared in 1996. John Lubeck has spent over 30 years proselytizing among the Ch'orti' and is known throughout the region for his fluency in the language. The Ch'orti' New Testament translation is an untapped resource of valuable data on the Ch'orti' language (see below).

The Proyecto Lingüístico Francisco Marroquín (PLFM) was first to publish a descriptive grammar of Ch'orti' in 1994 by Vitalino Pérez Martínez. This was followed soon after by a dictionary in 1996 (Pérez Martínez et al. 1996). Since this time, there has been a steady stream of published materials, counting numerous volumes containing dozens of oral traditions, by the Academia de Lenguas Mayas, Guatemala (ALMG), the PLFM, CONALFA, COIMCH, COMACH, the Comunidad Lingüística Ch'orti', and various other groups dedicated to promoting the Ch'orti' language. These sources produced by native-speaker Ch'orti' Maya have been an absolute boon for Ch'orti' studies and, more importantly, have begun to have a noticeable impact the endangered status of the Ch'orti' language.

In 1999, Søren Wichmann produced the most concise grammar of the Ch'orti' language to date that drew upon many of the sources mentioned above and his own fieldwork in the area. More recently, the ALMG compiled a short dictionary of Ch'orti' in 2001 and in 2004 published a descriptive grammar. In 2005, I produced a 1,500-word trilingual dictionary of Ch'orti' Maya-Spanish-English (Hull 2005). Schumann (2007) has also recently published his *Introducción a la morfología verbal del chortí*, which has valuable data on the differences in the speech of younger and

older Ch'orti'. Most recently, in 2016, I published *A Dictionary of Ch'orti' Mayan – Spanish – English* with roughly 9,000 Ch'orti' entries (Hull 2016).

The involvement of the Ch'orti' in the study of their own language has resulted in a sharp increase in published material in Ch'orti' and is helping to fully document dialectal forms in the different Ch'orti' communities. Yet an unexpected consequence of Ch'orti' language revitalization groups' contributions is the internal push to modify the language to fit their idealized visions of it. In the remainder of the study, I outline some of the efforts underway by certain Ch'orti' language revitalization groups to 'purify' their language by 'purging' it of some forms while altering others to accommodate either their understanding of how the language once was in the past or what it SHOULD be (regardless of whether it actually was that way previously).

5. The use of *-kojt* in Ch'orti'

In order to demonstrate this process, it is necessary to first establish the earliest uses and interpretations of certain morphemes, namely the NCs *-kojt* and *-te'*. These two classifiers have undergone semantic shifts caused both by natural processes of generalization and through a recent reappraisal of their function by modern Ch'orti' language revitalization groups. The most evident of these semantic movements is found with *-kojt*.

I first want to discuss one of the more intriguing seeming incongruities in Ch'orti' NCs – the use of *-kojt* with certain tortillas, tamales, and possibly bananas,⁷ which gives us a glimpse into its early semantic range. In one of my early fieldwork trips among the Ch'orti' I was a bit taken aback when I heard a husband yell across the room to his wife, *B'ajnan inkojt yar nipa'*, "Warm me up a little tortilla!" Since the root *-kojt* refers to "being on all fours", its use as a NC for tortillas (and food made from tortillas) was somewhat surprising. Note another example:

- (5) *Tzepe* *cha'kojt* *u't* *e* *pa'*
 tzep-e cha'-kojt u-ut e pa'
 stack.up-IMP two-NC(QUADRUPEd) 3A-face DEF tortilla
 "Stack up two tortillas!"

7. Several consultants have reported to me that it is possible to use *-kojt* with bananas, but I have never heard it used this way in daily conversation. Other consultants disagreed, so this identification remains tenuous for the moment.

Table 1 contains the attested tortilla types (or foods made with tortillas) that can take *-kojt*:

Table 1. Tortilla types attested with NC *-kojt*

Ch'orti'	Spanish	English
1. <i>pa'</i>	<i>tortilla</i>	tortilla
2. <i>pixtonpa'</i>	<i>memela, torilla grueso</i>	thick tortilla
3. <i>tz'ak'watz'</i>	<i>tortilla de elote tierno</i>	tortilla made from young corn
4. <i>to'mpa'</i>	<i>empanada, tamal de frijol</i>	stuffed bread pastry, bean tamale
5. <i>sispa'</i>	<i>tortilla fría</i>	cold tortilla

All of these terms are forms that are commonly used in quotidian speech, though those over 40 years of age tend to use *-kojt* with much more consistency than younger speakers, who often use *-te'* instead or no classifier at all.

Oakley's 1966 data contains the first published use of *-kojt* with tortillas, *incojt nipa* "one of my corn cakes [i.e., tortillas]" (1966: 248). Also in John Fought's unpublished Ch'orti' fieldnotes from the 1960s, *-kojt* appears in the expression *cha'ko-jtpa'*, or "two tortillas" (Fought n.d.a: 368). The dictionary *Ojronerob' ch'orti'* of the ALMG, one of the largest language revitalization groups operating in Guatemala, also notes that *-kojt* is the correct classifier for *pixtomp'a'* "thick tortillas", *pa'* "tortillas", and *sispa'* "cold tortillas" (2001a: 215). One Ch'orti' speaker (who has worked for the ALMG) explained to me that they use *-kojt* with certain foods that are "essential to life" and that they are "so closely associated with 'life' that use *-kojt*."⁸ This may be a folk etymology, but the question remains as to why these objects that could otherwise be easily described with other classifiers based on their shape take a classifier with its origins in the bending down of four-legged animals. What can be said is that most of the oldest speakers I have worked with regularly (but not exclusively) use *-kojt* with certain types of tortillas and foods made using them, and there is no indication that it is a recent development in the language.

8. If one further entertains this etymology, one might also assume the since tortillas are made of corn – the most important, indeed sacred, food of the Maya, – this might explain the use of a classifier that has come to be associated with "living things". However, no other words relating to corn take the classifier *-kojt*. For example, *jechò'* "elote (young corn)" takes *-te'* instead as do many other words relating to different varieties of corn, so it is not simply the fact they are made of corn that necessitates the use of *-kojt*. Furthermore, there are considerable chronological problems with this interpretation in that the shift from using *-kojt* with quadrupeds to "living things" almost certainly post-dates the use of *-kojt* with tortillas.

There is no direct⁹ data on the use of *-kojt* prior to 1950 in any context, though it was surely part of the language. It does not appear in the work of Suárez, Membreño, or Sapper.¹⁰ It is not until Wisdom's 1950 dictionary that we find the first explicit evidence of the use of *-kojt*, where he defines it as a numeral classifier referring to a "quadruped, on all fours". It is crucial to note the singularity of his definition as referring only to "quadrupeds" or being "on all fours", with no mention of any further semantic generalization at this stage in Ch'orti'. Somewhat unexpectedly, in all of Wisdom's entries there are only two examples of use of *-kojt*: *ux.koht ci'* [*uxkojt tz'i'*] "three dogs" and *in koht wakax* [*inkojt wakax*] "one cow". Thus, the earliest recorded use of *-kojt* in Ch'orti' is strictly limited to quadruped animals. There is no indication that it could be used with a human referent whatsoever. Indeed, a careful analysis of all Ch'orti' linguistic data provided in the footnotes of Wisdom's 1940 ethnography shows in ALL that cases humans take the *-te'* numeral classifier.

The next published source pertaining to *-kojt* was in 1966 by Helen Oakley, where she glosses *-kojt* as "animal-like" (1966: 247). In the Ch'orti' text she provided, *-kojt* only appears one time with the word "*sapo* (toad)" (1966: 249). Once again, *-kojt* is never used in conjunction with any human referents. In fact, humans consistently take *-te'* (e.g., *inte' winic* "a man" and *inte' ixic* "a woman" in Oakley's data [1966: 249]).

Around the same time in the 1960s, Fought's work on Ch'orti' was producing much more linguistically detailed descriptions of the language. Fought (1972: 52) states that *-kojt* is used for things that are "bent over, crouching". Thus, we find quadrupeds such as mules, cows, and frogs taking *-kojt* in his texts (cf. Fought 1972: 178, 404). In Fought's large corpus of unpublished audio and written data held in the AILLA repository (www.ailla.utexas.org/), four-legged animals (cows, horses, etc.) regularly take *-kojt*. In his published data Fought also notes that *-kojt* is found with "children, animals and poultry" (1972: 53). What is significant here is the extension of meaning to include "animals", not just "quadrupeds", as Oakley also recorded it. What is more, in both his published and unpublished data chickens and turkeys also can take *-kojt* even though they are bipeds.¹¹ Even non-legged

9. There is, however, substantial indirect evidence in Wisdom's 1940 ethnography, where his transcriptions contain uses of *-te'* in places where today one might hear *-kojt*, thereby indicating not so much when *-kojt* was used, but when it was *not* used.

10. A caveat to this statement is that a decent portion of Sapper's handwritten notes on Ch'orti' is nigh illegible. A more careful and expert analysis of his orthography, however, could reveal the presence of the *-kojt* NC since he does have numerous example sentences in his work.

11. Fought's published texts have one example of *-kojt* being used for a chicken (1972: 776). Data from López de Rosa (2001: 35), however, seems to indicate that the use of *-kojt* for "chicken" is now declining among young speakers of Ch'orti' today.

animals such as worms (*b'i*) or snakes (*chan*) take the classifier *-kojt* (though more often *-te*) (Fought n.d.b: 251, 252, 255). The expansion in function to include other animals around this time finds curious confirmation in Fought fieldnotes from the 1960s where *-kojt* is labeled as “animate” (Fought n.d.b.: 251).¹² What is more, on the same page of fieldnotes Fought gives the first clear evidence of the use *-kojt* with a human reference: *cha'kojt*, which he translates anaphorically as “*dos personas* (two people)” (Fought n.d.b.: 251, orthography altered).

In addition to being used for “non-human animals”, Brent Metz (personal communication, July 2010), has also noted the use of *-kojt* among the Ch'orti' with “children”. The grouping of animals and infants into one class is not uncommon in languages with NCs. However, the reason for viewing them in a similar light can vary considerably. In Tamil, for example, animals and infants are classified together because they are both “irrational beings” and therefore do not necessitate a NC (Asher 1985: 135). In Ch'orti', however, the reason almost certainly lies in the semantics of the original term. The likely answer is that children spend much of their first year of life ‘on all fours’, crawling as their primary means of mobility. Since the origin of *-kojt* lies in the verbal root *kot* “be down on all fours”, the extension to children would have been an easy one.

Lubeck & Cowie's (1989: 357) explanation of the meaning of *-kojt* in their pedagogical grammar sheds little direct light on the status of the NC *-kojt*, saying only that it refers to “*algo completo* (something complete)” – a definition that is hard to square with its use today in Ch'orti', or indeed with their own data.¹³

Schumann (2007: 127) describes *-kojt* as being used “*para contar animales en general* (to count animals in general).” His work contains a number of cases of its use such as *cha'kojt* for “two (animals)”, referring to pigs (2007: 253), *jay-kojt* “¿Cuántos animales? (How many animals?)” (2007: 240), and *inkojt inkojt* for “*a cada animal* (each animal one by one)” (2007: 130).

To summarize so far, Fought's fieldnotes in the 1960s contain the earliest single reference to the NC *-kojt* being used with a human head noun. There is one case, however, in the texts of Fought (1972) which deserves our attention. In the story “Buzzard”, the Buzzard states: “Me, my – the meat I eat is not all the same. I eat cow, I eat horse, I eat dog, and chicken, turkey – and if – and if there is a person (*inkojt jente*) lying dead, I eat him too” (Fought 1972: 178, 180, orthography altered). Note that the NC for “person” (Spanish *gente*, which I write in Ch'orti' as *jente*) is preceded by the NC *-kojt*, though it is referring to a dead person. Based on these two

12. Somewhat confusingly, however, the word “not” (“animate – not”) is written next to this notation, making his original intent somewhat opaque.

13. I have heard a similar definition given by one of my Ch'orti' consultants, though he could not produce an example. Still, this unexpected meaning needs further investigation.

examples from Fought's data, we could hypothesize that the language had already begun to generalize the semantics of *-kojt* even further at this point to include some human head nouns. While no data exist confirming the use of *-kojt* with humans from this time until Lubeck and Cowie's 1989 pedagogical grammar (see below), it can be assumed that the form continued in use and likely slowly expanded its range of human referents.

6. A shot in the arm for semantic generalization

Language change is at times systematic, following certain predictable chronological patterns. Yet there are moments when forces not internal to the system bear sway on its features. It appears that the advent of Ch'orti' language revitalization groups has played a direct role in further facilitating the generalization from "animal" to "human", i.e., "living thing in general", for the NC *-kojt*. By looking at all the major works produced by these revitalization groups in Ch'orti' we can reconstruct to some extent the process as *-kojt* began to further shift semantic domains to include all human referents. This progression has been neither neat and tidy nor immediate, but the wheels are certainly in motion.

In fact, no amount of detective work is needed to conclude that many Ch'orti' language revitalization groups have promoted the use of *-kojt* to fill a void in the language – the lack of a true animacy marking NC. I have discussed the issue with many of the leaders of these groups (many of them trained linguists), and they have stated outright that *-kojt* was chosen for this role since it is "the closest thing we have to a general classifier for animate things" in that it already refers to LIVING ANIMALS. In other words, since it is already used for living animals, it is half way to a full animacy marker. Due to their abrupt adoption of *-kojt* for this semantic role, in most of the publications of these groups we find a strong disconnect between what is found in earlier sources and the way many of these revitalization groups have determined the language should be. Again, these changes have not occurred across the board, nor have they been promoted with the same fervor among different individuals within the various groups. Yet the tendency to use *-kojt* more and more when referring to humans in the publications of the ALMG and other such groups is rapidly growing.

I am not taking the position that *-kojt* as a NC for humans was strictly an innovation of those working with language revitalization groups, but rather that they facilitated a further semantic extension that had already been taking place during the last half of the twentieth century. This process of generalization does not seem to have taken root to any considerable degree since the data during this time period show overwhelmingly that *-te'* was the preferred classifier for all human referents.

It was the ALMG and other related groups, however, who gave this process a major shot in the arm and propelled *-kojt* into semantic roles even beyond its trajectory at the time.

Let me stress again that from the earliest published sources the NC *-te'* in Ch'orti' appears in nearly all cases referring to humans. For example, according to Wisdom (1950), the NC *-te'* means “a, one, thing, person, someone, the next, a little or a few of”. The two primary meanings here are “thing” and “person”. Indeed, *-te'* is the general classifier for “general things” in Ch'orti' today and in the past. Wichmann's corpus research on Ch'orti' has similarly revealed that *-te'* is the most commonly occurring NC in Ch'orti' (1999: 125). My Ch'orti' data also confirm this observation.

The wide range of categories subsumed by *-te'* can be best described, as Wichmann rightly notes, as comprising “all items not contained in one of the other classes” (1999: 125).¹⁴ After specifying its use for counting humans and objects in general, Schumann (n.d.) similarly concludes that *-te'* is used in other cases “*que no caben en el campo de otros clasificadores* (that do not fit in the field of other classifiers).” Elsewhere Schumann has also observed that the classifier “*-te'*” is used “*para personas y cosas en general* (for people and general things)” (2007: 127). Note that Schumann actually mentions “*personas* (people)” before “things” in this description. He also provides several examples of its use in this regard: *inte' winik* “*un hombre* (a man)” and *inte' ixik* “*una mujer* (a woman)” (Schumann n.d.). Thus, both people and general things are covered by *-te'*, per both Wisdom and Schumann. In neither case is *-kojt* given as an option for use with human beings. Wisdom's (1950) data certainly bear this out, where we find *-te'* being used with a man, a friend, a teething boy, a woman, a workman, a blind woman, a trembling man, a balding man, a girl, a companion, a pregnant woman, and a bearded man. In fact, Wisdom uses *-te'* 100% of the time when the head noun is a human.

Furthermore, in his earlier ethnography (Wisdom 1940), we find precisely the same situation, with *-te'* being used in every case when referring to a human. Indeed, Wisdom unequivocally stated that “Individual kin are referred to by *in-te'* (“a, one”) (1940: 274).

Wisdom's data reveal crucial details regarding the semantic development of both *-kojt* as well as *-te'*. As I have mentioned earlier, the only two cases where Wisdom provided an example of the use of *-kojt* was with “*ux.koht ci*” [*uxkojt tz'i*]

14. Wichmann (1999: 125) notes the broad application of the NC *-te'* with everything from natural phenomena, parts of landscape, and diseases to “plants, emotional or bodily states, instruments and utensils, persons (including kinship terms, persons who possess specific qualities, perform specific functions or hold certain offices), parts of plants, the body or constructions, verbal nouns describing kinds of actions, clothing, illnesses, and expressions of time measurement.”

“three dogs” and “*in koht wakax*” [*inkojt wakax*] “one cow”, both of which are true quadrupeds. However, it seems that the extension of meaning from quadrupeds to other animals had not taken place prior to Wisdom’s fieldwork in the 1930s. Evidence for this can be found in Wisdom’s consistent use of *-te’* with ALL animals, including birds, and even quadrupeds at times. These data show conclusively that there was some looseness in the system in Wisdom’s day that allowed *-te’* to be used for all animals, again even sometimes with quadrupeds.

Girard’s 1949 work (much of which was borrowed from Suárez’s 1892 data) confirms Wisdom’s use of *-te’* for humans and also shows the same flexibility to use *-te’* for quadruped animals at this time, e.g., *Kanin man inté ni čitam* [*K’ani inmani inte’ nichitam*] “I want to buy myself a pig” (1949: 126). Girard’s data contains a number of similar examples:

- (6) a. *inté* [*inte’*] *winik yoopa* [*yo’pa*]
 in-te’ *winik yo’p-a-Ø*
 one-NC(GENERIC) man arrive-THEME-3SG.B
 “A man arrived.” (Girard 1949: 125)
- b. *inté* [*inte’*] *išik* [*ixik*] *yoopa* [*yo’pa*]
 in-te’ *ixik yo’p-a-Ø*
 one-NC(GENERIC) woman arrive-THEME-3SG.B
 “A woman arrived.” (Girard 1949: 125)
- c. *inté* [*inte’*] *tz’i’*
 in-te’ *tz’i’*
 one-NC(GENERIC) dog
 “a dog” (Girard 1949: 126)

As I previously noted, in Oakley’s data humans also always take *-te’* (1966: 249). This same pattern plays itself out in Fought’s texts where *-te’* is consistently used for non-young children: *winik* “man” (1972: 114, 115, 146, 154, 158, 180, 205, 212, 213, 226, 285, 295, etc.) and *ixik* “woman” (1972: 205, 214), *ilama* “wife” (1972: 86, 87), *ijchòk* “girl/daughter” (1972: 309, 310), *primo* “cousin” (1972: 269), *jente* “person” (1972: 311), and *ajk’ajpesyaj* “padrino” (a rain-making priest) (1972: 402). The only time in all of his texts where *-kojt* is used instead of *-te’* for non-young children is the case mentioned earlier of a dead man, who was probably being likend to other dead animals as buzzard food (Fought 1972: 178). Finally, in Fought’s texts *-kojt* is used with *maxtak* “children”, through the conceptual similarity of quadruped animals and babies crawling on all fours as I previously described (cf. Fought 1972: 154).

Thus, the voices from early Ch’orti’ sources are unanimous on three points: (1) the NC *-kojt* is not used for non-young children, (2) *-kojt* appears with quadruped animals and with young children, and (3) *-te’* is the standard NC for humans in general, all animals, and even four-legged animals.

Post-1970 Ch'orti' linguistic sources begin to paint a different picture in terms of the use of *-te'* and *-kojt*. For example, in one dialogue, Lubeck & Cowie (1989: 29) use *-te'* with *winik* "man" as we would expect. In another dialogue in the text the authors use *-te'* with *ijtanob* "younger sisters", but when referring anaphorically to one of those sisters in the next line, they use *-kojt* instead (1989: 35). Similarly, in another section, a man is asked if he plans to plant by himself or with the help of a hired hand, to which he responds, "*C'ani insicb'a chacojt tua' utacren*" [*K'ani insikb'a cha'kojt twa' utakre'n*], "I want to look for two (hired hands) to help me". These two cases are perhaps some of the earlier recorded instances of *-kojt* being used with non-young children. That this practice may not have been widespread at this point is at least inferable from their use of *-te'* and *-kojt* elsewhere in the same publication. Case in point, the authors use *-te'* with *inte' umaxtacob* [*inte' umaxtakob*] "one of their children" (1989: 41), with *inte' uwijchoc* [*inte' uwijchok*] "one of his daughters" (1989: 41), and with *inte' chòm sitz* [*inte' chòm sitz*] "a young boy" (1989: 159). This variability in usage between *-te'* and *-kojt* is indicative of a system in flux, a transitional juncture when several competing forms are simultaneously in use.

The evidence for a movement toward semantic realignment with *-kojt* and *-te'* regarding humans is overwhelming in the contemporary Ch'orti' literature.

Some 25 years later we find clear evidence for both the continual generalization of *-kojt* as well as the persistence of the earlier *-te'*. In 1994, while the PLFM's grammar (Pérez Martínez 1994) expressly stated *-te'* was the correct classifier for *ak'ach* 'chicken', two years later in their 1996 publication on Ch'orti' texts, we see the use of *-kojt* ("*inkojt ak'ach*") (Pérez Martínez 1996: 23). In 1950 Wisdom used *-te'* with *ak'ach* and Fought attributed *-kojt* to "poultry" in 1972, so there is evidence for a continuing overlap in usage into the late 1990s. In *Leyenda maya ch'orti'*, published by the PLFM, we also find the extension from "poultry" only to other birds like *mwon* "hawk" (Pérez Martínez 1996: 28).

Indeed, beginning from the 1990s there seems to be an accelerated advance in the generalization of *-kojt* to include any type of animal, including birds, insects, fish, etc. Thus, in the 1999 *Curso literario idioma ch'orti'* of the ALMG, we find *-kojt* being used with *mwan* "hawk" (ALMG 1999: 12), *pejpem* "butterfly" (ALMG 1999: 27), *chan* "snake" (ALMG 1999: 28, 36), and *chay* "fish" (ALMG 1999: 32). In addition, in *Kanatalwa'r ch'orti'* by CONALFA, *chay* "fish", similarly takes *-kojt* (López de Rosa & Patal 1997: 141). In the ALMG's 2001 dictionary *Ojronerob' ch'orti'* (ALMG 2001a), *-kojt* appears with a wide variety of animals such as *tz'urub* "grasshopper", *tz'ujtz'uj* "Stripe-headed Sparrow", *che'che'* "Bushy-crested Jay", *am* "spider", *mukuy* "dove [various varieties]",¹⁵ *b'eya'* "White-throated Magpie Jay", *usij*

15. In Ch'orti', *mukuy* can refer to a number of species of doves, including Inca Dove, Ruddy Ground Dove, Blue Ground Dove (Hull & Fergus, n.d.)

“black vulture”, and *ajtzo* “turkey”. In the 2001 ALMG translation of the Popol Vuh, *mut* “bird”, *jopop* “crab”, and *arak* “wild animal” also take *-kojt* (ALMG 2001b: 29, 31, 68).

All birds in the *Diccionario ch'orti'* (ALMG 2001a) are given with *-kojt*, without exception. In addition, all animals, including fish, insects, worms, snakes, shellfish, etc. are explicitly said to take *-kojt*.

My data, collected between 1999–2012, a portion of which comes from those who have worked with the ALMG and other such groups, also show this general trend towards using *-kojt* more often with all animals. However, there are still many speakers who do NOT do so and continue using *-te'* with a number of different animals, again indicating the existence of a transitional phase. This intermediate stage of transition can perhaps best be seen in the mixed usage of *-kojt* and *-te'* in the Lubeck's translation of the New Testament into Ch'orti' (Lubeck 1996). Table 2 shows the number of cases of each NC with the same head noun in different passages of the Ch'orti' New Testament.

Table 2. Comparison of use of NCs *-te'* and *-kojt* with animals in the Ch'orti' translation of the New Testament

Term	Translation	# with <i>-te'</i>	# with <i>-kojt</i>
<i>chay</i>	fish	1	13
<i>animal</i>	animal	9	2
<i>pecpec</i> [pekpek]	frog	3	2
<i>mukuy</i>	dove	0	11
<i>cuncuna</i> [kunkuna]	dove [various]	0	1
<i>oror</i>	crow	2	1
<i>chan</i>	snake	4	8
<i>wacax</i> [wakax]	cow	7	6
<i>tz'i'</i>	dog	2	3
<i>arac</i> [arak']	wild animal	28	3
<i>sinam</i>	scorpion	0	4
<i>ac'ach</i> [ak'ach]	chicken	1	2
<i>mis</i>	cat	0	1
<i>mut</i>	bird	2	7

The inconsistency in usage suggests the two forms are in a state of flux. Indeed, in Revelation 16: 13, *pecpec* [pekpek] “frog”, appears both with *-kojt* and *-te'* in the same verse!

It needs to be reemphasized again that the semantic shift with *-kojt* I have described above is not unique to Ch'orti' among Mayan languages. The cognate NC in Ch'ol has undergone a very similar process. Coon & Preminger state: “The positional

koty, for instance, is used to describe not just four-legged animals, but also crawling babies or crouched people. The classifier, in contrast, is extended to many animals that do not have four legs, for example chickens and fish” (2011: 215, fn. 6).

This process of generalization is also found in Akatek Mayan, a language that has a cognate classifier, *-kòn* (*kòn-an*), which is used for animals that derive from the verbal root *-kòn* “bent on four feet” (Zavala 2000: 122). Zavala notes that *-kòn* “classifies nouns that denote not only mammals ‘standing on four feet’ but all animals including serpents, worms, insects, birds, fishes, etc.” Zavala further states that *-kòn* (and other “classificatory suffixes” such as *-wan* “erect on two feet”) have undergone both phonological erosion and semantic generalization (2000: 122). For example *-wan* has gone from referring to beings “standing on two feet” to refer to humans in any position (sleeping, sitting, lying dead, etc.) (2000: 122). He attributes this to the change from lexical classifier to classificatory suffix, parallel to the process of grammaticalization, resulting in “the reduction of an open paradigm” (2000: 122).

Thus, the evolution of *-kojt* in Ch’orti’ towards being more inclusive to all animals has parallels in several other Mayan languages. The shift to human referents, however, is not as commonly attested.

7. The further extension of meaning among human referents

The final area I explore in more detail is the growing tendency to use *-kojt* with any type of human head noun in recent years and to assess the role revitalization groups have played in this process. We have already established the use of *-kojt* in the earliest sources up through the 1980s. Since that time, however, there has been an unprecedented increase in its application to human head nouns, part of which is a direct result of its adoption by language revitalization groups as a broad animacy marker for Ch’orti’. An examination of the use of *-te’* and *-kojt* in post-1990 published sources provides a window into this progressive shift in action.

As PLFM’s *Gramática del idioma ch’orti’* (Pérez Martínez 1994) is the first significant work to appear from one of the Ch’orti’ language revitalization groups, the use of *-te’* and *-kojt* in this work should be of considerable value in determining their function at this critical juncture. In the section of the grammar dealing with NCs, under the description of the NC *-te’*, the first two examples of correct usage are given as *inte’ winik* “un hombre (a man)” and *inte’ ixik* “una mujer (a woman)” (Pérez Martínez 1994: 101), unambiguously establishing the fact that *-te’* was considered a proper classifier for humans at this point in 1994. Under the examples of *-kojt*, the PLFM grammar appropriately gives four examples: *tz’i* “dog”, *ak’ach*

“chicken”, *wakax* “cow”, and *chij* “horse”, but mentions several other four-legged animals in other parts of the text also (Pérez Martínez 1994: 101). Everywhere in the example sentences of this publication humans are consistently modified by *-te'*, never *-kojt* (e.g., *inte' ijch'ok* “a girl”) (Pérez Martínez 1994: 143, 154).

Two years later when the PLFM (Pérez Martínez 1996) published a short compilation of oral traditions in *Leyenda maya ch'orti'*, there was no change in terms of the use of *-te'* for all human head nouns: e.g., *inte' ixik* “a woman”, *cha'te' winikob'* “two men”, *cha'te' ajxujch'ob'* “two thieves”, *cha'te' ch'urkab'ob'* “two babies”, *inte' ijch'ok* “one girl”, *cha'te' uwijtz'inob' tejrom* “two younger brothers”, and *inte' pyar'* “a companion”. Other publications by the PLFM around this time are also valuable for determining the position held by the PLFM as to the use of *-te'* and *-kojt*. A survey of kinship terms in their 1996 *Diccionario ch'orti'* (Pérez Martínez et al. 1996) shows that in nearly every case *-te'* is preferred to *-kojt*. All words referring to humans similarly take *-te'* throughout, such as *ixik* “lady”, *winik* “man”, and even *pak'ab'* “person” (Pérez Martínez et al. 1996) – a significant fact as we will see. Indeed, there is only one term I could find relating to humans where it is explicitly stated that the NC should be *-kojt* and not *-te'*, that being with the term *arib'* “niece” (Pérez Martínez et al. 1996: 13). This single instance, however, may reveal the presence of two competing forms, though *-te'* is obviously the overwhelmingly preferred choice.

Kinship terms were first affected by this change from *-te'* to *-kojt*. Soon, however, this usage would be extended to any type of word that referred to a human, mainly by the ALMG and those who followed suit.

The use of *-te'* with human referents in the *Diccionario ch'orti'* (Pérez Martínez et al. 1996) by the PLFM stands in stark contrast to publications of the ALMG (cf. 2001a; 2001b; 2001c; 2005). Similarly, pedagogical publications by CONALFA (López de Rosa & Patal 1997) show considerably higher rates of usage of *-kojt* than the two major PLFM publications (Pérez Martínez 1994; Pérez Martínez et al. 1996).¹⁶ A comparison of NCs used with human referents (mainly kinship terms) among the *Diccionario ch'orti'* of the PLFM¹⁷ (Pérez Martínez et al. 1996), *Kanatalwa'r ch'orti'* of CONALFA (López de Rosa & Patal 1997), and *Ojronerob' ch'orti'* (ALMG 2001a) of the ALMG shows clearly their contrastive usage (Table 3):

16. In fact, some members of the ALMG and CONALFA have worked for both organizations at different times.

17. Other human referents in the PLFM's dictionary are *béchu* “*patojito* (small child)”, *et'ok* “*compañero (a)* (companion)”, *ixtumin* “*mujer rica en dinero* (woman rich in money)”, *kwach* “*gemelo (a)* (twin)”, *mimi* “*niño de un mes hasta un año* (infant between one month and one year old)”, *menor* “*huérfano* (orphan)”, and *merex* “*humilde* (humble person, poor person).”

Table 3. Comparison of the use of *-te'* and *-kojt* in select published works of the PLFM, CONALFA, and ALMG

Term	Translation	PLFM	CONALFA	ALMG
<i>ixik</i>	woman	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>winik</i>	man	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>ch'urkab'</i>	baby	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>tu'b'ir</i>	mother	<i>-te'</i>		<i>-kojt</i>
<i>chuchu'</i>	baby	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>ijchok</i>	daughter, girl	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>maxtak</i>	child	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>tejrom</i>	boy	<i>-te'</i>	<i>-kojt</i>	<i>-kojt</i>
<i>cha'r</i>	stepchild	<i>-te'</i>		<i>-kojt</i>
<i>sitz'</i>	son	<i>-te'</i>		<i>-kojt</i>

The examples in Table 3 from CONALFA were gleaned from various passages through these works, while the PLFM and ALMG dictionary entries provided unequivocal statements as to their correct usage under each entry. What is remarkably apparent from this comparison is the major swing by CONALFA and the ALMG to their consistent usage of the morpheme *-kojt* when modifying various human head nouns.

The word for “person” (*pak'ab'*) also shows the same progression and inconsistency in its use with NCs. Oddly, the term does not appear in Wisdom’s dictionary (1950). The Spanish loan *gente* “people”, which is commonly substituted for *pak'ab'* today in Ch’orti’, is found in Fought’s texts, where it takes the NC *-te'* (Fought 1972: 178, 311). Even the *Diccionario ch’orti'* lists the correct classifier for *pak'ab'* as *-te'* under the *pak'ab'* entry (Pérez Martínez 1996: 163). In Lubeck’s translation of the New Testament in Ch’orti’ *gente* is used throughout without a single instance to *pak'ab'*. In every case *gente* takes *-te'* (e.g., Luke 17: 34, “*chate' gente*,” meaning “two people”). In Metz’s dissertation, interview transcripts in Ch’orti’ also show the use of *-te'* with “*gente*” (Metz 1995: 301, note 8). Just as we have seen before, however, in the late 1990s and beyond, CONALFA, the ALMG, and others make the conscious shift to modifying *pak'ab'* with *-kojt*. Thus, we find “*inkojt pak'ab'* (one person)” in CONALFA’s 1997 publication (López de Rosa & Patal 1997: 11, 18, 101, 108, 109, 114, 168). In the ALMG’s *Ojronerob' Ch’orti'*, *-kojt* is similarly the default classifier for every case of *pak'ab'* (ALMG 2001a: 16).

Even while the ALMG’s own publications were becoming more and more consistent in the use of *-kojt* with *pak'ab'*, in the ALMG’s 2001 translation of the Popol Vuh *pak'ab'* still appears with *-te'* at times (see ALMG 2001b: 15). On the other hand, the ALMG’s grammar of Ch’orti’, despite containing no clear explanation of the use of *-kojt*, is completely consistent in its use of *-kojt* with *pak'ab'* (ALMG 2004).

In daily conversations in the Ch'orti', however, *-te'* is still considerably more common today than *-kojt*, as this induced shift is struggling to gain traction with those outside of the reach of the ALMG and related groups, especially older speakers.¹⁸ For example, as recently as 2007, Schumann (2007: 240) notes the question “how many people?” is expressed as “*jay-te'*” rather than *jaykojt*.

The core motivation for this modification was not a natural occurrence in the language, but rather a conscious decision among the members of these groups. Thus, the near instantaneous increase in the number of semantic domains of *-kojt* must be contextualized and understood in these terms. Tellingly, however, we can find evidence that this change was not universally applied by all speakers within these organizations. This is especially apparent in publications in which various individuals were responsible for different sections of the text. As might be expected, a certain liminal period might persist in which prior tendencies to use *-te'* might win over more recent directives to use *-kojt*, even in-house. In fact, just such an intermediate stage of use is what is found in the publications of this time.

The next step in the generalization of *-kojt* was to extend its scope to any noun that referred to any living thing. This change took a bit longer; nonetheless it was in part the inevitable result of reinterpretation of *-kojt* taking place during this period. Therefore we see terms like *nuch* “priest” starting to take *-kojt* in literature (ALMG 2001c: 71). In the ALMG's *Ojronerob' ch'orti'* there are sweeping (yet still inconsistent) changes well beyond kinship terms to include essentially any reference to humans such as “*huérfano* (orphan)” and “*ma'n* (hired hand)” (ALMG 2001a).

Inconsistencies in the use of *-te'* and *-kojt*, however, are highly telling and provide a clear view into the transitory period. For example, even with terms that have been using *-kojt* for a much longer time, such as *maxtak* “child”, the *Ojronerob' ch'orti'* dictionary of the ALMG occasionally shows an inconsistent usage between *-kojt* and *-te'*:

18. My consultants who have worked with the ALMG and other such groups regularly switch between both forms, commonly so even in a single story or sentence (I have noted dozens of cases of these in my data). The effects of the ALMG and other language revitalization groups on many speakers' way of viewing *-kojt* became very clear to me on various occasions while transcribing Ch'orti' texts with consultants who had worked with such groups. While transcribing after recording their stories, I was often asked to change a *-te'* to a *-kojt* in order to be ‘more correct’. While speaking extemporaneously, my consultants who had worked with language revitalization groups would often speak in a way that was more natural to them and would forget to use *-kojt* with all living things, as they had been ‘instructed’.

- (7) a. *uxkojt* *umaxtak*
 ux-kojt u-maxtak
 three-NC(ANIM) 3A-child
 “She has three children.” (ALMG 2001a: 168)
- b. *inte’* *maxtak*
 in-te’ maxtak
 one-NC(GENERIC) child
 “a child.” (ALMG 2001a: 159)

Other inconsistencies in the same publication include “*inte’ nyar’*,” “a son-in-law” (2001a: 62–63), and “*inte’ winik*,” “a man” (ALMG 2001a: 56).

The final traceable shift in the semantic range of *-kojt* is its use with ‘agentive’ nouns referring to humans, which are formed in Ch’orti’ by adding the prefix *aj-* “one who”, to verbal or nominal roots. Traditionally, all agentive nouns have taken *-te’* when modified by NCs; for example, Fought’s texts contain a mention of “*inte’ ajk’ajpesyaj*,” “a prayer-sayer” (1972: 402, 417, orthography altered). This pattern accords well with all speakers I have worked with in the field who were not associated with any language revitalization groups (roughly 40 individuals). There is a universal tendency to prefer *-te’* over *-kojt* with agentive nouns. Thus, throughout the entire New Testament in Ch’orti’ all agentive nouns are consistently modified by the NC *-te’* when counted, regardless of whether they refer to humans or not. Accordingly, the *Diccionario ch’orti’* gives *-te’* as the correct classifier for basically all titles for humans prefixed by the ‘agentive’ *aj-* “he/she/one who”, e.g., *inte’ ajtumin* “a rich person”, *inte’ ajk’otor* “an employee”, or *ajmukmayaj* “an undertaker” (Pérez Martínez et al. 1996: 6–9). The use of *-kojt* is nowhere to be found with human referents in this publication. It is clear that the extension of *-kojt* to these agentive human nouns had not been made by this point in 1996 insofar as the PLFM was concerned.

What we see in the 1990s, however, is a gradual pendulum shift to begin including agentive nouns referring to humans within the domain of *-kojt*. Thus, in the ALMG’s 1999 publication *Concurso literario idioma ch’orti’*, we find the first cases of agentive nouns referring to humans taking *-kojt* instead of *-te’*:

- (8) *kochwa’ inkojt* *nixi’ ajk’otorer*
 kochwa’ in-kojt nixi’ aj-k’otor-er
 how one-NC(ANIM) big AG-power-NMLZ
 “like a great authority.” (ALMG 1999: 27)

Other examples include “*inkojt ajmwak*,” “a sick person” (ALMG 1999: 35) and “*inkojt ajk’anpa’r*,” “a soldier” (ALMG 1999: 32). By the time of the ALMG’s 2001 dictionary *Ojronerob’ ch’orti’* this change had spread to an ever-growing number of different human head nouns, though not with complete regularity (ALMG 2001a).

Due to the recent nature of this change, however, there is still inconsistency in its implementation in the publications. For instance, in *Leyenda maya ch'orti'* (Pérez Martínez 1996: 11), we find *cha'te' ajxujch'ob'* “two thieves”. In 2001 the term *ajxujch* [*ajxujch'*] (thief) is given with *-kojt* on page 38 but with *-te'* on pages 36 and 162 of *Ojronerob' ch'orti'* (ALMG 2001a). Other agentive nouns with human references are only given with *-te'*, such as *ajkojk* “guardian (guard)” and *ajtz'ijb'ayaj* “escritor (writer)” (ALMG 2001a: 12, 157). My data from those whom I have worked with in the ALMG, CONMICH, and other groups also contain numerous cases of this extension of the use of *-kojt* with agentive nouns referring to humans: *ajcheyaj otot* “carpenter”, *ajputz'er* “fugitive”, *ajpajrete'* “sculptor”, *ajpakxer* “passer-by”, *ajpetz'si'* “firewood stacker”, and *ajk'ech ojroner* “messenger’ (lit. “one who brings word”). However, in my experience this change from *-te'* to *-kojt* with agentive nouns referring to humans has not been received with much enthusiasm. Indeed, it is uncommon in the speech of those not affiliated with the ALMG and other such groups.

8. Discussion

We have seen to this point that semantic range of *-kojt* has undergone substantial realignment in the last century, the motivations for which are two-fold: (1) the natural generalization of the term, and (2) the direct influence of the ALMG and other revitalization groups. Much of the recent change has been caused by the latter, i.e., by directives and instructions on proper usage by the ALMG and other such groups. The perceived need for an animacy marker proper was a major motivator for the expansion of the usage of *-kojt* with all human head nouns. Again, several leaders of the ALMG have told me directly that they have been promoting it as such since the language lacks such a morpheme found in some other Mayan languages.¹⁹ For example, in the few sources left to us of Ch'olti', the now-extinct language that was most closely related to Ch'orti', there was a NC for “humans”, *-tul*, which was glossed by Morán in his word list as referring to “*hombres*” but likely had a non-gendered reference (Robertson et al. 2010). This is clearly cognate to the *-túul* of Yukatek – a true NC for animate objects (Bricker et al. 1998: 284) – and the *tuhl* of Tzeltal used for “people” (Shklovsky 2005: 40). In Chontal, the cognate NC for both people and animals is *-tu* (e.g., *Ajnon cä mäne' untu piyo' tan caj*, “I went to buy a chicken in town”) (Keller & Luciano 1997: 273). It is the lack just this type of animacy marking classifier that the ALMG felt the need to ‘rectify’.

19. In Chol, the NC for animate objects is *-tikil*, while *-kojt*, as we have seen, is reserved for counting animals (Schumann 1973: 31).

The new position on the question of *-kojt*'s use is expressly stated in *Kanatalwa'r ch'orti'*, a pedagogical book written in Ch'orti' for educating Ch'orti' who are learning the Ch'orti' language (López de Rosa & Patal 1997: 141): "*Tama tuno'r xe' ayan ub'ixirar kak'ampes e ojroner kojt. Kocha' ak'ampestz'a. Inkojt chay, cha'kojt pakab' [pak'ab']*," "For all things that have life we use the word kojt. It is used with one fish, two people" (translation mine). Here the net is cast as widely as possible to include "all things that have life".²⁰ The extension of a NC from animals to humans, while considered insulting in some languages,²¹ was embraced by the Ch'orti' since it was the closest classifier in the language that referred to something living.

9. A change in momentum

The pressing question in this discussion is whether the changes being promoted by the ALMG and other groups will win the day and have a long-lasting impact on the language as a whole. Lucas López de Rosa, a native Ch'orti' speaker, completed his graduate thesis from the Universidad Rafael Landívar entitled, "*Los clasificadores en el idioma maya ch'orti'*" (2001). López de Rosa analyzed the question of the usage of *-kojt* and *-te'* from data secured from conversations, from written surveys in various Ch'orti' communities, and through an analysis of five written oral traditions. Despite some methodological problems associated with his research, his data yield some important results. Based on conversations he conducted with both young and adult Ch'orti' speakers, López de Rosa found that *-kojt* occurred 33% more commonly than *-te'* for young people but that *-te'* was three times more common with adults (2001: 63). Since many of those whom he interviewed worked for or had been taught by the ALMG and similar groups, it is not necessarily surprising that *-kojt* is found more often since it is precisely these groups that have expanded its use²² considerably while concurrently attempting to redefine the semantic range of *-te'*. The fact that adults show the exact opposite tendency, i.e., *-te'* being used 3-to-1 over *-kojt*, points to their longer experience in using the language as it is actually spoken and their slower acceptance of newer models and prescriptive statements.

20. Elsewhere, López de Rosa states *-kojt* is a "*Sufijo / clasificador para enumerar seres vivos y tortillas* (Suffix / classifier for counting living things and tortillas)" (2001: 50).

21. For example, in Khmer (Mon-Khmer) the NC for animals, *kba:l*, is also used in pejorative circumstances for humans (Adams 1989, cited in Aikhenvald 2003: 443).

22. A glance at some of the example sentences of the younger speakers used to formulate his calculations show that it is precisely the type of words the ALMG and other such groups are promoting that took *-kojt* in this case (e.g., *pak'ab'* "person", *ijch'ok* "girl", and *tejrom* "boy") (López de Rosa 2001: 59).

More telling are the responses given in the written surveys conducted by López de Rosa. Out of 50 Ch'orti' speakers, both young and adult, only 10 used *-kojt* in sentences with “*animales, personas y tortillas*” (López de Rosa 2001: 35, Table 7), clearly indicating a preference for *-te'* NC among the Ch'orti' in several of the key semantic domains of *-kojt*: animals and people. Furthermore, his conclusion about *-kojt* “*siendo el clasificador más utilizado* (being the most-used classifier)” in the five stories he recorded and transcribed for his thesis is highly problematic. All of the stories in his appendix are from speakers who had worked or at the time worked for language revitalization groups. Therefore, their use of the classifier *-kojt* had likely been strongly biased. Two of the stories are from Hipólito Ohajaca Pérez, who has also been my main linguistic consultant over the last 13 years, and he has told me outright that he uses *-kojt* with all living things because he was told that was the proper way to speak by the ALMG. Overall, the most revealing part of López de Rosa's research is that it indicates that older speakers have a greater tendency to use *-te'* in many of the situations where the ALMG and similar groups today would use *-kojt*.

As noted earlier, some of these changes, especially the use of *-kojt* with agentive nouns referring to humans, are making little headway at present with the general populace. Indeed, in my experience it is fair to say that many of the decisions made by the ALMG have not been well-received in the hamlets by most speakers. Brent Metz, a leading specialist on the Ch'orti' and someone involved in numerous early meetings with the ALMG in the 1990s, recently made a similar observation that he knows of “plenty of people who resist the Academia's [the ALMG's] instructions” (personal communication, July 2010). With limited resources (human, financial, and physical), the adoption of many of the new ‘rules’ for speaking and the more than 1,000 neologisms they have recently published (see ALMG 2006) may take considerably longer to filter into the daily speech of the majority of the Ch'orti'.

Groups such as the ALMG that are involved in the crucial documentation and promulgation of the Ch'orti' language make up a very small percentage of the Ch'orti'-speaking population. Thus, the impact so far on the language as a whole could be viewed as negligible. On the other hand, the ‘sanitizing’ and replacement of certain forms by activist Ch'orti' groups do represent permanent changes in one important sense: the written word. The amount of literature produced by these groups has been considerable in recent years and has gone far in making important information in Ch'orti' available to Ch'orti' speakers, especially for classrooms where Ch'orti' is taught or used in instruction. The growing body of new Ch'orti' publications making use of ‘purified’ forms could therefore have an incrementally larger impact on the speech of the general Ch'orti' population over time, especially since no other literature is available to the Ch'orti' outside of these publications. Thus, even minor changes in grammatical or orthographical forms could eventually

become the standard for Ch'orti' speakers due to their ever larger presence in the copious published works being produced by the ALMG and others. However, in the short term, it remains to be seen just how serious an impact these innovations have on the general Ch'orti'-speaking population.

10. Conclusion

This diachronic study of the process of semantic generalization of the NC *-kojt* in Ch'orti' shows clear trends in its development as well as change brought about by direct intervention. The data presented above trace this evolution chronologically through the documented sources in Ch'orti'. Of course, as is always the case in languages with numeral classification systems, there can be considerable variation in the use of NCs among the general populace. Some of the variations noted in this study are possibly related to aspects of style at a particular moment (i.e., social setting). NCs are not monolithic; their specificity can vary depending on various social or linguistic factors (cf. Aikhenvald 2003: 446). Additionally, individual levels of maturity in the language and the age of the speakers can also influence their use. Among all of the consultants I have worked with during nearly three years of fieldwork among the Ch'orti', it has consistently been the oldest speakers that have shown a more limited use of *-kojt* and a much broader use of *-te'*.

Agreed upon notions of 'correctness' by the ALMG and other such groups are slowly making their mark on the Ch'orti' language today. This meta-awareness of what constitutes this 'correctness' represents the first step towards the codification of the Ch'orti' language. Part of this process is purely ideologically driven, the result of a sense of inadequacy within the language that needs to be 'fixed', as some among the ALMG have explained it to me. I have shown here that the semantic generalization of the NC *-kojt* in Ch'orti' is part of this ongoing process. However, other aspects of its evolution took place long before these language revitalization groups, such as the change from being used with quadrupeds to all animals in general. At some point, *-kojt* began to modify human head nouns, a development that was picked up on and accelerated by the ALMG and other such groups in order to establish a fully functioning animacy marking NC in Ch'orti'.

Simultaneously, the use of *-te'*, which had been the default NC for all human referents up until this time, had its semantic domain slightly reduced, though only by a small margin so far. The NC *-kojt* is yet to be fully adopted by many speakers with any human head noun. Indeed, in 2010, in a discussion on this topic with workers at the ALMG, Metz (personal communication, July 2010) reports that they all agreed people use *-te'* with *winik* (man) or *pak'ab'* (person), but, as Metz was told, "they shouldn't" (!).

Thus, the semantic generalization of the Ch'orti' NC *-kojt* is the natural result of a broadening of the semantic domain of *-kojt* – parallel to similar changes in other Mayan languages containing a cognate form – as well as a result of ideology-driven linguistic change.

References

- Adams, Karen Lee. 1989. *Systems of Numeral Classification in the Mon-Khmer, Nicobarese and Asian Subfamilies of Austroasiatic*. Canberra: Pacific Linguistics.
- Aikhenvald, Alexandra Y. 2003. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Attinasi, John J. 1973. *Lak T'an: A Grammar of the Chol (Mayan) word*. Ph.D. dissertation, University of Chicago.
- ALMG. 1999. *Concurso literario idioma ch'orti'*. Guatemala: K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2001a. *Ojronerob' ch'orti'*. *Vocabulario ch'orti'*. Guatemala: K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2001b. *Pojp jun*. K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2001c. *Utwa'chir e ojroner ch'orti'*: *Tradición oral ch'orti'*. Guatemala. K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2004. *U'tirach e ojroner maya ch'orti' / Gramatica descriptiva ch'orti'*. Guatemala: K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2005. *Na'tanyaj xe' ayan taka e pak'ab'ob'. Tzijb'arb'ir tama ojronerob' ch'orti' yi e castellano – Tradición oral bilingüe de la cultura ch'orti', ch'orti' – castellano*. Guatemala. K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- ALMG. 2006. *Topop ojroner maya ch'orti' / Actualización lexical del idioma Maya Ch'orti'*. Guatemala: K'ulb'il Yol Twitz Paxil (Academia de Lenguas Mayas de Guatemala).
- Asher, R. E. 1985. *Tamil*. London: Routledge.
- Bricker, Victoria, Eleuterio Po'ot Yah & Ofelia Dzul Po'ot. 1998. *A Dictionary of The Maya Language As Spoken in Hocobá, Yucatán*. Salt Lake City: University of Utah Press.
- Coon, Jessica. 2005. "On the Formation of Intransitives and Numeral Classifiers in Chol". Unpublished manuscript.
- Coon, Jessica & Omer Premiger. 2011 "Positional Roots and Case Absorption". *New Perspectives in Mayan Linguistics* ed. by Heriberto Avelino, Jessica Coon & Elisabeth Norcliffe, 206–241. Cambridge, MA: MIT Working Papers in Linguistics 59.
- Craig, Colette. 1986. "Jacalteco Noun Classifiers: A study in grammaticalization". *Lingua* 70.241–284. doi:10.1016/0024-3841(86)90046-X
- Craig, Colette. 1992. "Classifiers in a Functional Perspective". Layered Structure and Reference in a Functional Perspective: Papers from the Functional Grammar Conference, Copenhagen, 1990 ed. by Michael Fortescue, Peter Harder & Lars Kristofferson, 277–301. Amsterdam: John Benjamins.
- Fought, John G. 1967. *Chorti (Mayan): Phonology, morphonemics, and morphology*. Ph.D. dissertation, Yale University.

- Fought, John G. 1972. *Chortí (Mayan) Texts* ed. by Sarah S. Fought. Philadelphia: University of Pennsylvania Press.
- Fought, John G. N.d.a. Ch'orti' Field Notes, #355–413 CAA005R006I00. www.ailla.utexas.org. (Accessed: 27 October 2017.)
- Fought, John G. N.d.b. Ch'orti' Field Notes, #214–270 CAA005R003I001. www.ailla.utexas.org. (Accessed: 27 October 2017.)
- García de León, Antonio. 1971. *Los elementos del tzotzil colonial y moderno*. México, D. F.: Universidad Nacional Autónoma de México.
- Girard, Rafael. 1949. *Los chortís ante el problema maya: historia de las culturas indígenas de América, desde su origen hasta hoy*, Vol. I. México, D.F.: Antigua Librería Robredo.
- Grinevald, Colette. 2004. "Classifiers". *Morphologie: ein internationales Handbuch zur Flexion und Wortbildung* ed. by Geert E. Booij, Christian Lehmann, & Joachim Mugdan, 1016–1030. Berlin: Walter de Gruyter.
- Haviland, John Beard. 1981. *Sk'òp sotz'leb, el tzotzil de San Lorenzo Zinacantan*. México, D.F.: I.I.F./C.E.M., Universidad Autónoma de México.
- Hull, Kerry M. In preparation. "Ch'orti' Mayan Numeral Classifiers". Unpublished manuscript.
- Hull, Kerry M. 2005. *An Abbreviated Dictionary of Ch'orti' Maya. A final report for the Foundation for the Advancement of Mesoamerican Studies, Inc. (FAMSI)*. www.famsi.org/reports/03031/index.html. (Accessed: 27 October 2017.)
- Hull, Kerry M. 2016. *A Dictionary of Ch'orti': Mayan – Spanish – English*. Salt Lake City: University of Utah Press.
- Hull, Kerry M. & Rob Fergus. N.d. Fieldnotes from ethno-ornothological research in Ch'orti' area in 2006.
- Hurley Vda. de Delgaty, Alfa & Agustín Ruíz Sánchez. 1978. *Diccionario tzotzil de San Andrés con variaciones dialectales* (= *Serie de vocabularios y diccionarios indígenas "Mariano Silva y Aceves"*, 22). México, D.F.: Instituto Lingüístico de Verano, A.C.
- Kaufman, Terrence. 1972. *El proto-tzeltal-tzotzil: fonología comparada y diccionario reconstruido* (= Centro de Estudios Mayas, Cuaderno 5). México: Universidad Nacional Autónoma de México.
- Kaufman, Terrence S. & William M. Norman. 1984. "An Outline of Proto-Cholan Phonology, Morphology, and Vocabulary". *Phoneticism in Maya Hieroglyphic Writing* ed. by John S. Justeson & Lyle Campbell, 77–166 (= *Institute of Mesoamerican Studies*, 9). Albany: SUNY-Albany.
- Keller, Kathryn C. & Plácido Luciano G. 1997. *Diccionario chontal de Tabasco*. Tuzcon, AZ: Instituto Lingüístico de Verano, A.C.
- Lacadena, Alfonso Garcia-Gallo. 2010. "Synchronic and Diachronic Variation of Maya Classifiers: From glyphs to alphabet". Paper presented at "Stability and Adaptation of Classification Systems in a Cross-Cultural Perspective" at the École Normale Supérieure, Paris, May 14th–16th, 2010.
- Laughlin, Robert M. 1975. *The Great Tzotzil Dictionary of San Lorenzo Zinacantan* (= *Smithsonian Contributions to Anthropology*, 19). Washington D.C.: Smithsonian Institution Press.
- López de Rosa, Lucas. 2001. *Los clasificadores en el idioma maya ch'orti'*. M.A. thesis, Universidad Rafael Landívar, Guatemala.
- López de Rosa, Lucas & Flavio Mucía Patal. 1997. *Kanatalwa'r ch'orti': texto integrado para la primera etapa de post-alfabetización bilingüe idioma ch'orti'*. Guatemala: CONALFA/Instituto de Lingüística de la Universidad de Rafael Landívar.

- Lubeck, John E. 1996. *E ojroner xe' imb'utz tua' cawinquirar Jesucristo : el Nuevo Testamento en chortí-español*. Wycliffe Bible Translators, Inc.
- Lubeck, John E., & Diane L. Cowie. 1989. *Método moderno para aprender el idioma chortí'*. Guatemala: Instituto Lingüístico de Verano.
- Membreño, Alberto. 1897. *Hondurñerismos: vocabulario de los provincialismos de Honduras*. Tegucigalpa: Tipografía Nacional.
- Metz, Brent E. 1995. *Experiencing Conquest: The political and economic roots and cultural expression of Maya-Chortí*. Ph.D. dissertation, State University of New York at Albany.
- Metz, Brent, Teo Ramírez, Fred García & Mundo García. 1992. *Español – Chortí – English – Diccionario*. Unpublished manuscript.
- Oakley, Helen. 1966. "Chortí". *Languages of Guatemala* ed. by Marvin K. Mayers, 235–250. The Hague: Mouton.
- Pérez Martínez, Vitalino. 1994. *Gramática del idioma chòrtí'*. Guatemala: Proyecto Lingüístico Francisco Marroquín.
- Pérez Martínez, Vitalino. 1996. *Leyenda maya chòrtí'*. Guatemala: Proyecto Lingüístico Francisco Marroquín.
- Pérez Martínez, Vitalino, Federico, GarcíaFelipe Martínez and Jeremías López. 1996. *Diccionario del idioma chòrtí'*. Antigua, Guatemala: Proyecto Lingüístico Francisco Marroquín.
- Robertson, John S., Danny Law & Robbie A. Hartel. 2010. *Colonial Cho'lti': The seventeenth-century Moran manuscript*. Norman: University of Oklahoma Press.
- Suárez, Ruano Alberto. 1892. "Idioma Chortí". Box 42, Folder 7, Gates Collection, Lee Library. Brigham Young University.
- Schumann, Otto Gálvez. N.d. *Vocabulario chortí-español*. Unpublished document.
- Schumann, Otto Gálvez. 1973. *La lengua chol de Tila (Chiapas)*. México: Universidad Nacional Autónoma de México.
- Schumann, Otto Gálvez. 2007. *Introducción a la morfología verbal del Chortí*. México: Academic Publishers.
- Shklovsky, Kirill. 2005. "Person Marking in Petalcingo Tzeltal". B.A. thesis, Reed College.
- Warkentin, Viola & Ruby Scott. 1980. *Grammatica ch'ol (= Serie de Gramáticas de Lenguas Indígenas de México, 3)*. México, D.F.: Instituto Lingüístico de Verano.
- Wichmann, Søren. 1999. *A Chortí' Morphological Sketch*. Unpublished manuscript.
- Wisdom, Charles. 1940. *The Chortí Indians of Guatemala*. Guatemala: Editorial José de Pineda Ibarra.
- Wisdom, Charles. 1950. *Material on the Chortí Language*. Microfilm Collection of Manuscripts on Meso-american Cultural Anthropology, no. 28. Chicago: University of Chicago. Transcribed by Brian Stross.
- Zavala, Roberto. 2000. "Multiple Classifier Systems in Akatek (Mayan)". *Systems of Nominal Classification* ed. by Gunter Senft, 114–146. Cambridge: Cambridge University Press.

Diachronic and synchronic aspects of the simplification of grammatical gender in an obsolescent language

The case of Irish

Alessio S. Frenda

Institute of Technology Blanchardstown

The comparison of two corpora of spoken Irish – one representative of a traditional variety as spoken in 1964, the other of the language spoken today on Irish-language media – reveals a number of differences in the way gender agreement is marked on different agreement targets. The difference is minimal in terms of article–noun agreement and most conspicuous with pronominal agreement. An intermediate position is occupied by noun–adjective agreement within the noun phrase. This study suggests that the gender system of Irish, historically based on both semantic and morphological assignment, is becoming purely semantic, similar in this to the gender system which characterized the final stages of a number of now-extinct Celtic varieties. In their final stages, these exhibited a purely semantic gender assignment system, whereby feminine agreement was limited to anaphoric pronouns with female antecedents. Within the noun phrase, feminine agreement markers were only found in a few fossilized expressions. Irish currently appears to be at an intermediate stage along the path towards a similar system.

1. Introduction

While it seems uncontroversial that gender assignment always has a semantic core even if sometimes limited (Corbett 1991: 63), its exact nature is language specific; however, it has been observed that when a language has more than one gender for animates, the biological gender distinction constitutes the semantic core of the system (Dahl 1999a, 1999b). In Irish, biological gender determines the assignment of animates (humans, certain higher animals); the gender of inanimates appears to correlate with certain formal properties and some secondary semantic criteria with rather limited predictive power.

Within the Celtic family, to which Irish belongs, the traditional gender system has undergone some changes. This can be observed for example in the final stages of now-extinct Manx and Cornish, where the gender of inanimates ceased to be distinguished and marked (see §5). In present-day Irish and Welsh, uncertainties about the gender of inanimates have been reported in the speech of younger speakers: observations about the virtually complete collapse of the traditional grammatical gender system have been made by, e.g., Jones (1998) and Maguire (1991: 211); but, perhaps more strikingly, similar developments or incipient traces thereof have also been documented in more traditional varieties, e.g. by Jones (1998: 66, 171) for Welsh, Ó Siadhail (1989: 149) for Donegal Irish and Dorian (1976, 1981: 124–129) for East Sutherland Gaelic.

All these studies highlight that anaphoric reference to inanimate antecedents, both masculine and feminine antecedents, tends to be made using masculine pronouns. In East Sutherland Gaelic, the attributive adjective had also virtually ceased to agree with feminine nouns, with the unmarked masculine forms prevailing.

Whether and to what extent such developments are due to the influence of varieties of English has also been discussed: Ó Siadhail (1984: 175) suggests that the overgeneralization of masculine pronouns for inanimate referents could be contact-independent since it has long been attested in Donegal Irish, and Comrie (1979) also points at contact-independent, language-internal motivations (such as phonetic distance) to account for the same phenomenon in Welsh.

Dorian (1977: 100) observes that structures which have a functional counterpart in English seem to resist change better than those which do not – and gender distinctions within the realm of inanimates are almost completely absent from English. The same point is made by Stenson (1993: 122) apropos of “existing constructions with English analogues” and is corroborated by Maguire’s observation that the palatalized/non-palatalized opposition in the consonant system of the neo-Gaeltacht children’s variety is almost completely lost,¹ which undoubtedly brings the phonology of this variety closer to that of English (Maguire 1991: 203). In general, Maguire speaks of the adoption of a “lowest common denominator” approach in the Irish of these bilingual children, who appear to choose the form that most readily suits the communicative needs, often a non-standard one “directly influenced by the English system” (Maguire 1991: 189–190). Developments like these are particularly important, since the phonological opposition of palatality was instrumental to the preservation of the declension system after the loss of the inflectional endings (see McCone 1994: 94).

1. *Gaeltacht* (plural *Gaeltachtaí*) is the Irish term used to officially designate those areas where Irish is still spoken as a community language. The term *neo-Gaeltacht* is used to designate later, self-established communities of non-traditional speakers.

It has been suggested that language change in obsolescent languages is faster than in ‘healthy’ ones.² The difference, according to Dorian (1981: 154), is not one of quality, but of pace:

The types of change in formal language structure ... are not notably different from those well established in the study of language change in general. But the timespan for change seems to be compressed ... and the amount of change seems relatively large.

In other words, obsolescence “precipitates” language change (Jones 1998: 240). In principle, then, the study of obsolescent languages should give us the opportunity to observe, in a compressed timespan, a sequence of changes that would take longer to surface and become generalized in ‘healthier’ languages.³

In this paper, after providing some preliminary information about the sociolinguistic status of Irish in §2.1, I sketch the gender system of traditional Irish in §2.2. Section 3 presents the data sets I used in this study and the methodology I adopted: I compared two data sets, one representative of a traditional variety as spoken in 1964, the other of the language spoken today on Irish-language media. The analysis is both quantitative and qualitative and focuses on the persistence of syntactic agreement with masculine and feminine controllers, to assess whether its rate depends on the controller’s gender and/or on the data set.

The findings are presented in §4, which focuses in turn on agreement with the article (§4.1), the attributive adjective (§4.2) and the personal pronoun (§4.3); a discussion of their significance follows in §5. In §6, I sum up the discussion, highlight some of the limitations of this study and point out directions for future research.

2. Jones (1998: 5–6) defines obsolescence as a “gradual reduction in use, due to domain-restriction, [which] may result in the emergence of historically inappropriate morphological and/or phonological forms together with extensive lexical borrowing”.

3. As pointed out by one reviewer, obsolescence ensues when a linguistic community transitions, for different reasons, to another language; it is a contact-related phenomenon, and the systemic changes that characterize it are typical of contact situations. For this reason, Dorian’s statement that essentially the same type of changes is produced in both obsolescent languages and non-obsolescent, non-contact situations might be a rather strong generalization from a particular to a more general case. My reading of Dorian’s statement, however, is that while linguistic obsolescence is a special case it is not an exceptional one, and changes that take place under obsolescence are subject to the same general tendencies that govern language change in general. For instance, as discussed below, they are in line with generally-observed typological tendencies such as the Agreement Hierarchy.

2. Irish: Sociolinguistic background and the traditional gender system

2.1 Sociolinguistic status of Irish

Irish, the object of this study, is today an obsolescent language. Its speakers are all bilinguals (Ó Murchú 1988: 248) and represent a minority of the population.⁴ The linguistic effects of obsolescence, in terms of historically unexpected forms, can be noticed in various domains. Ó Curnáin (2007a: 36), for instance, observes the change in word order from the traditional [object pronoun + connective particle + verbal noun] (e.g. *é a dhéanamh* “doing it”) to [verbal noun + object pronoun] (*déanamh é*) in the Irish of Iorras Aithneach (in the Galway Gaeltacht); as already mentioned, Maguire (1991: 203) describes the reduction of the phonological inventory following the loss of the traditional opposition of palatality. As regards morphological change, Nic Pháidín (2003: 125) has observed that some speakers use analytical forms ([preposition + pronoun]) instead of the traditional inflected prepositions.

As a spoken language, Irish is characterized by dialectal fragmentation, which coexists with a codified standard variety used for teaching and official purposes (see Ó Baoill 1988). The standard variety (*an Caighdeán*) is an artificially created one, which assembles forms and structures taken from different dialects, and has in the past met the hostility of native speakers: ‘Dublin Irish’ (as it was called) was regarded by many as a spurious, unnatural linguistic invention (Breatnach 1964; Dorian 1994; Ó Baoill 1988). Nevertheless, the education system and the media have managed to win new speakers through more or less standardized forms (Ó Baoill 1988; Ó Murchú 1969). This raised concerns about the loss of local dialectal features as the standardized variety was also used in Gaeltacht schools, instead of the local native varieties (Breatnach 1964; Ó Flatharta 1989). However, these varieties have since been given visibility and promoted, including through radio and TV: *Raidió na Gaeltachta* (“Radio of the Gaeltacht”) was established in 1972 and TG4 (originally *Teilifís na Gaeilge* “Irish-language TV”) in 1996 (Akutagawa 1987: 138). Both stations have chosen to adopt the native Gaeltacht varieties, rather

4. According to figures from the 2006 Census, 40.8% of the population aged three years and over identified themselves as Irish speakers; however, only about 44% of the latter (or 18% of the total) self-reported to ever use Irish outside the education system, and only 3% of them (or 1% of the total) to do so daily (CSO 2006: Tables 1 and 40; Punch 2008). In 2011, the number of self-identified Irish speakers shrank further to 31.3%, while the number of people speaking Irish outside the education system was reported to be 19% of the total; within Gaeltacht areas only, the percentage of those reporting to speak Irish daily outside the education system was 35% (CSO 2011: 40–41).

than the standard, even in formal, ‘corporate-brand’ programmes such as the news (Esslemont 2007). In any case, these different varieties essentially share the same grammatical gender system.

2.2 Gender in traditional Irish

Modern Irish has two genders, masculine and feminine. In what follows, I outline the gender system of traditional Irish.

Gender assignment in Irish can be described as correlating with both semantic and formal features, in terms of Corbett (1991). Gender assignment is normally semantic in the case of animates, i.e. based on their biological gender: female referents are feminine and male referents masculine. There is a very small number of exceptions: if the grammatical gender of a noun is at odds with its semantic gender, the former prevails in local agreement (within the noun phrase) and the latter in pronominal agreement. For example, *cailín* ‘girl’ is masculine: within the noun phrase, the article and adjectives show masculine agreement forms, but for anaphoric reference a feminine pronoun is always used.

For inanimates, gender assignment correlates with formal properties of the noun. At least historically, the relevant correlation was with the inflectional paradigm. Formal assignment has undergone a certain amount of restructuring and there is now a tendency to equate palatal final consonants with the feminine gender and non-palatal ones with the masculine. This way, gender assignment correlates with just one form of the noun (its citation form), rather than with a set of inflected forms. This strong correlation is itself a reflex of the ancient system of inflectional endings, and it has been extended analogically in some dialects (cf. Ó Siadhail 1984).

Gender agreement is systematically found only with the article, the attributive adjective and third-person singular pronouns,⁵ on which we will concentrate in turn. Agreement marking in Insular Celtic crucially depends on a series of morphological alternations collectively known as initial mutations (see Frenda 2011: 5.3.2 for a discussion). Briefly put, three types of initial mutation are relevant to gender marking:

5. Attributive nouns, i.e. nouns which immediately follow the controller within the same noun phrase, are sometimes treated as adjectives for the purposes of agreement, provided that they head an indefinite noun phrase. However, there exists a certain variation across the different dialects as to the extent to which this is the case, as well as several exceptions: cf. de Bhaldraithe (1953: 262) for Connemara Irish and Ó Sé (2000: 61) for Munster Irish; see also Caighdeán (1979: 83–84) for the Official Standard.

1. lenition, orthographically indicated by the letter < h > after the sign of the mutated consonant, is the alternation between a stop and a homorganic fricative; the system is not perfectly symmetrical, and it also applies to a few consonants that are already fricative (for an overview, see Ó Siadhail 1989: 112);⁶
2. /t/-prefixing;
3. /h/-prefixing.

Essentially, agreement in gender (and number) within the noun phrase (article–noun and noun–adjective) is marked by initial mutations. Notably, article–noun agreement can be marked on the controller (by means of initial mutations which affect the noun), as well as on the target (using different forms of the article). In the genitive singular, for instance, gender is marked in both ways: on the target, by two different morphemes (*an* vs. *na*), and on the controller, by initial mutation (lenition after *an*, /h/-prefixing after *na*), as in Example (1).

- (1) a. *cóta* *an* *pháiste*
 coat(M).NOM.SG ART.GEN.M.SG child(M).GEN.SG
 “the child’s coat” (cf. NOM. SG. *an páiste*)⁷
- b. *barr* *na* *haille*
 top(M).NOM.SG ART.GEN.F.SG cliff(F).GEN.SG
 “the top of the cliff” (cf. NOM. SG. *an aill*)

Adjectives mark gender agreement via initial mutations and, to a much more limited extent, morphologically, as can be seen in Examples (2a) and (2b). However, the morphological marking of gender and case in the adjective is now quite rare and tends to be preserved only in set phrases (Mac Eoin 1993: 115–116). As we shall see in § 4.2, it is hardly represented in our data.

An attributive adjective is lenited for instance when it follows a masculine noun in the genitive singular: cf. *bhig* /v’ig/ in Example (2b), compared to the unlenited form *beag* /b’og/ in (2a).⁸

6. Historically, for the sonorants /N, R, L/, lenition indicated their alternation with non-tense /n, r, l/, which is not represented orthographically. However, the [± TENSE] opposition is today virtually lost (cf. de Bhaldraithe 1953: 257 fnn. 1–3; Ó hUiginn 1994: 560, 562; Ua Súilleabháin 1994: 488; Hickey 2003: 264; Ó Curnáin 2007a: 188–201) and lenition in these contexts was therefore not investigated as part of the present study.

7. Lenition: < p > = /p/, < ph > = /f/.

8. Lenition: < b > = /b/, < bh > = /v/; also note the alternation between the nominative singular morpheme *beag* and the genitive singular morpheme *big*.

- (2) a. *cailín* *beag*
 girl(M).NOM.SG little.NOM.M.SG
 “a little girl”
- b. *cóta* *an* *chailín* *bhig*
 coat(M).NOM.SG ART.GEN.M.SG girl(M).GEN.SG little.GEN.M.SG
 “the little girl’s coat”

Agreement with personal pronouns is limited to the third person singular. Pronouns agree morphologically, e.g. masculine *sé* vs. feminine *sí*, with the exception of possessive pronouns, which mark agreement by initial mutation of the following word, as shown in (3).

- (3) a. *a* *chóta*
 POSS.3SG.M coat(M).NOM.SG
 “his coat”
- b. *a* *cóta*
 POSS.3SG.F coat(M).NOM.SG
 “her coat”⁹

3. Methodology

3.1 The data

This study is based on the analysis of linguistic data from two corpora of spoken Irish. The first corpus, referred to below as the older generation corpus (OC), contains material recorded in 1964 in Ros Muc, a small village located in the Connemara Gaeltacht, west of the city of Galway. It was recorded during a number of informal recording sessions, in which people from the area – one or two at a time – would engage in conversation with a younger member of the community, who acted as a facilitator on behalf of the researcher; they would mostly recount anecdotes and stories about their lives. The resulting corpus was later edited for publication by Wigger (2000).

Only a portion of this corpus was used in this study. The selection I analyzed totals about 12,200 words. All seven speakers in it were born between 1896 and 1937. Five of them were born between 1896 and 1914 and were either fishermen or farmers. The youngest two (an electrician and a teacher) were both born in 1937; the teacher was the facilitator and his contribution to the corpus is negligible, limited to a few interspersed questions asked to either start the session or invite

9. Lenition: < c > = /k/, < ch > = /x/.

another speaker to intervene. One advantage of using this corpus is that the original recordings are available online,¹⁰ so that it was possible to listen to them and check for the accuracy of the transcriptions.¹¹ Excerpts from this component are identified by a bibliographical reference (Wigger 2000) followed by the three-number code given to it by the corpus editor, e.g. 4-01-02.¹²

The second corpus, referred to as the contemporary corpus (CC), is comprised of 11 texts and contains approximately 20,000 words.¹³ It includes seven texts (approximately 13,000 words) from Raidió na Gaeltachta, the national Gaeltacht-based broadcaster (RnaG), and four texts from non-Gaeltacht sources (approximately 7,000 words).¹⁴ The shows were broadcast between 1997 and 2007; they were recorded or retrieved from existing archives and transcribed with the assistance of three Irish speakers. There are 26 speakers in total, of whom 11 can be placed in the 30–40 age group (at the time of recording), seven in the 40–50 group and four in the 50–60 group. Another three can be estimated (based on the content of the interview and contextual information) to be younger than 30 and one to be older than 60. The type of material included is non-scripted and conversational in nature and normally takes the form of a one-on-one interview. The composition of the corpus is such that it includes material from both areas where Irish is still supposed to be spoken as a community language and those where this is not the case; the aim

10. They can be downloaded from the website of the Dublin Institute for Advanced Studies at <https://www.dias.ie/celt/celt-publications-2/caint-ros-muc/>. (Accessed: 30 January 2018.)

11. The transcriptions had been somewhat ‘polished’ for better readability, i.e. repetitions and false starts were often omitted. The facts pertaining to agreement, and in particular to initial mutations, were mostly accurately represented.

12. The first number (4 in this example) refers to the conversational session (referred to as *caint* by Wigger 2000) in which the excerpt is found; the second (01 in the example) to a specific part in the *caint* (each part has a title which refers to the topic of the conversation); the third number (02 in the example) identifies the stretch of discourse in which the example is found (this subdivision is found in Wigger (2000) and refers to stretches generally not exceeding one minute in duration).

13. The size and design of this corpus were determined to match a parallel corpus of contemporary broadcast Welsh used in a wider study of gender agreement in Celtic (Frenda 2011). Also note that this is roughly the size of the broadcast interviews component in the International Corpus of English, which is available at <http://ice-corpora.net/ice/design.htm>. (30 January 2018.)

14. The latter include two Dublin-based radio stations (Raidió na Life and Newstalk) as well as a show from the (now discontinued) podcast series *An Líonra Sóisialta* (“The Social Network”). <http://www.anlionra.com/2006/11/14/lionra-82-wikiie/>. (Accessed: 20 January 2010.)

was to represent the language that today's speakers of Irish are exposed to by the media. Excerpts from this corpus are identified by citing the source (radio station and programme name) and the date of the recording in the format dd/mm/yyyy (e.g. RnaG "Seal Aneas" 21/10/1997).

3.2 Procedure

Each text was manually searched for any context in which gender agreement could be marked. Agreement rate was then calculated for each agreement target as the number of instances in which agreement was found divided by the total number of contexts in which it could have been found. Given that rather small counts had been recorded for certain target types, all tests of statistical significance for the difference between proportions were Fisher's exact tests (cf. Baayen 2008: 113); statistical significance was assumed for p-values smaller than 0.05.

4. Findings

In this section I present the results of the analysis, dealing in turn with each target type (article, adjective and pronouns) in § 4.1–§ 4.3, and summarizing the findings in § 4.4.

4.1 Agreement with the article

In the older generation corpus, the rate of article–noun agreement is almost the same for both masculine (98%) and feminine controllers (97%), with no statistically significant difference, as shown in Table 1.

Table 1. Agreement with the article in the older generation corpus

	Masc.	Fem.
Agreeing	227	107
Total	232	110
%	98	97
	p = 0.7154	

In the contemporary corpus (Table 2), the article–noun rate of agreement is significantly lower with feminine nouns (88%) than with masculines (97%).

Table 2. Agreement with the article in the contemporary corpus

	Masc.	Fem.
Agreeing	176	136
Total	182	155
%	97	88
	p = 0.0028	

One difference between the older and the contemporary corpus is that in the former, article–noun agreement is as frequent with masculines as it is with feminines, whereas in the latter it appears to be significantly less frequent with feminines.

The two corpora also differ in terms of the likelihood of agreement with feminine nouns. As shown in Table 3 (“Overall”), this is significantly greater in the older than in the contemporary corpus (97% vs. 88% of the time, respectively). The figures in Table 3 (under the heading “Excl. proper names”) also address the question whether the occurrence of agreeing articles within proper names, including place names and personal names, inflated agreement rates. It may be argued that if an agreement form is contained within a noun phrase that constitutes a proper name, as in Examples (4) and (5), then it should not be considered as actually representing productive agreement, under the hypothesis that such noun phrases are stored in the lexicon as unanalyzed units.

Table 3. Article agreement with feminine controllers, older (OC) vs. contemporary corpus (CC)

	Overall		Excl. proper names	
	Fem. (OC)	Fem. (CC)	Fem. (OC)	Fem. (CC)
Agreeing	107	136	107	123
Total	110	155	110	142
%	97	88	97	87
	p = 0.0059		p = 0.0029	

- (4) RnaG “Seal Aneas” (21/10/1997)
Mac Na Mara
 son(M).NOM.SG ART.GEN.F.SG sea(F).GEN.SG
 (a surname)
- (5) RnaG “I Measc na nDaoine” (22/02/2007)
An Ghráig
 ART.NOM.F.SG village(F).NOM.SG
 (a place name)

Therefore, the above comparison was carried out a second time after proper names had been removed from the samples (13 such noun phrases with feminine heads were found in the contemporary corpus, none in the older one). This, however, did not significantly alter the picture, as shown in the two rightmost columns of Table 3; we can conclude that the contribution of this factor was also marginal.

Since gender assignment in Irish can have formal or semantic motivation, it makes sense to ask whether we would observe different agreement rates if we did not consider cases of semantic agreement, that is, if we removed from the sample all the feminine nouns that have female referents. However, only two such nouns were found (one in each corpus), both showing syntactic (feminine) agreement. The contribution of the semantic factor is therefore negligible.

4.2 Agreement with the adjective

As we have seen, gender agreement with attributive adjectives can be marked both via initial mutation and through inflection. The latter, however, is only found in the genitive singular and only with certain morphologically-defined classes of adjectives. In the data set I used, morphological marking was possible in just six occurrences (five in the older and one in the contemporary corpus); gender marking was in line with syntactic agreement in all six, but it ought to be noted that five of them were counted as proper names. Therefore, no meaningful generalization about the regularity of inflectional agreement could be made based on these data, and what follows only refers to agreement via initial mutations.

In the older generation corpus (Table 4), masculine agreement is not significantly more frequent than feminine agreement (94% and 92% respectively). In the contemporary corpus, on the other hand, masculine agreement (observed 91% of the time) is significantly more likely than feminine agreement (45%), as shown in Table 5.

Table 4. Agreement with the adjective in the older generation corpus

	Masc.	Fem.
Agreeing	49	33
Total	52	36
%	94	92
	p = 0.6853	

Table 5. Agreement with the adjective in the contemporary corpus

	Masc.	Fem.
Agreeing	51	9
Total	56	20
%	91	45
	p < 0.0001	

If we compare the two corpora, we see that adjective agreement with feminine controllers is significantly more likely in the older than in the contemporary corpus, as shown in Table 6.¹⁵

Table 6. Adjective agreement with feminine controllers

	Fem. (OC)	Fem. (CC)
Agreeing	33	9
Total	36	20
%	92	45
	p = 0.0002	

A few examples of proper names were counted, like the two place names in (6) and (7).

- (6) Wigger (2000) (1-01-06)

Aill *Bhuí*
 cliff(F).NOM.SG yellow.NOM.F.SG
 “Aille-wee [lit. Yellow Cliff]”

- (7) Wigger (2000) (1-01-09)

Cuan *an* *Fhír* *Mhóir*
 Harbour(M).NOM.SG ART.GEN.M.SG man(M).GEN.SG big.GEN.M.SG
 “Greatman’s Bay”

All of these show the expected agreement forms. However, to exclude them from the count does not significantly alter the picture, as shown in Table 7: in the contemporary corpus, agreement with masculine controllers is still more frequent than with feminine ones. The difference between the two corpora in terms of agreement rates with feminines also remains significant, as shown in Table 8.

15. Only two occurrences of semantic agreement, both in the contemporary corpus, were found. The contribution of the semantic factor to the observed agreement rates is therefore negligible.

Table 7. Agreement with the adjective, excluding proper names

	Older generation corpus		Contemporary corpus	
	Masc.	Fem.	Masc.	Fem.
Agreeing	47	28	49	8
Total	50	31	54	19
%	94	90	91	42
	p = 0.67		p < 0.0001	

Table 8. Adjective agreement with feminine controllers, older vs. contemporary corpus, excluding proper names

	Fem. (OC)	Fem. (CC)
Agreeing	28	8
Total	31	19
%	90	42
	p = 0.0007	

Even though only adjectives capable of marking gender agreement were counted,¹⁶ we can observe that attributive adjectives are not too frequent in the data: no more than 100 occurrences in either corpus. This is however not altogether unexpected and appears consistent with Mac Eoin's (1993: 116) remark that the use of the adjective in attributive position is perceived as 'bookish' and generally avoided in spoken Irish.

4.3 Pronominal agreement

Pronouns were divided into three groups, based on (a) the different agreement-marking devices employed (inflection for types 1 and 2, initial mutations for type 3) and (b) Ó Curnáin's (2007b: 1261) observation that the generalization of masculine anaphors with feminine inanimate antecedents is more likely to happen when the pronoun is prepositional.¹⁷ The three groups are the following:

16. That is, only adjectives whose initial consonant is subject to initial mutation.

17. As regards the distinction between anaphora and cataphora, one might expect it to have a bearing on agreement rate. Ó Curnáin (2007b: 1260–1261), for instance, reports that in the Irish of Iorras Aithneach (County Galway) the cataphoric position is a favourite context for the over-generalization of masculine pronouns. However, cataphoric relations with feminine antecedents are extremely rare in our corpora (only two were found in the contemporary corpus and just one in the older generation corpus). None of them, in line with Ó Curnáin's observation, shows agreement with the co-referential noun phrase.

1. simple pronouns (e.g. *sé* [3SG.M] ‘he/it’);
2. prepositional pronouns (*aige* [at.3SG.M] ‘at him/it’);¹⁸
3. possessive pronouns (*a^L* [POSS.3SG.M] ‘his/its’).

One limit to the investigation of gender agreement with prepositional pronouns is that it is often impossible to distinguish between the gender-inflected pronoun *ann* (*i* ‘in’ + 3SG.M) and the invariable form *ann* ‘there’, the latter representing adverbial usage and not involving agreement, as shown in (8).

- (8) *níl aon duine ann*
 NEG.be.PRS one person in.3SG.M/there
 ‘no one is in it’ / ‘there is no one’

It is almost always possible to interpret *ann* as meaning simply ‘there’. The only structure in which *ann* can be unambiguously identified as a pronoun appears to be the predicative construction [NP + *atá ann/inti*], exemplified in (9).

- (9) a. Wigger (2000) (1-01-18)
Nár mhaith an fear_i a bhí ann_i?
 NEG.COP.PST good ART.NOM.M.SG man(M).NOM.SG REL
 be.PST in.3SG.M
 ‘Wasn’t he a good man?’
- b. Wigger (2000) (7-01-06)
[oíche Sathairn]_i a bhí inti_i
 night(F).NOM.SG Saturday(M).GEN.SG REL be.PST in.3SG.F
 ‘it was a Saturday night’

Therefore, it was decided that no occurrence of *ann* would be counted unless as part of this type of structure.

Also excluded from the count were all instances of complex anaphora (in the sense of Consten et al. 2007), i.e. expressions with a propositionally structured antecedent and a conceptually complex referent. By default, complex anaphora is realized in Irish using masculine pronouns, but feminine pronouns may be found if the predicate is a feminine noun, as in the following example (taken from the older generation corpus).

18. The term ‘prepositional pronouns’ refers to forms of the prepositions inflected for person, number and gender.

(10) Wigger (2000)(1-01-01)

T: *Bhoil anois, is faoi is mó a bhí muid ag caint*
 well now it is about this mostly that we were talking
an oíche cheana faoin mbádóireacht, ar ndóigh,
 the other night about boating of course
'sí is gaire do do chroí.
 it is that which is closest to your heart.

M: *'Sí ar ndóigh an chéad* turn ...
 It is of course the first thing [?]

T: *Agus, t'athair, agus is dóigh d'athair mór romhat.*
 And your father and of course your grandfather before you

M: *Muise 'sin í an fhírinne.*
 Indeed (is) that.F the truth(F)

“T: Well now, that’s what we were talking about the other night for the most part, boating, of course you’re very fond of it. – M: I am, of course, it’s the first thing ... – T: And your father, and of course your grandfather before you. – M: Indeed, that’s the truth.”

In (10), *sin* “that” is a complex anaphor: it is used by the speaker to refer to the statement that boating has been a central activity in his family for generations. The feminine pronoun *í* is semantically pleonastic but syntactically required here because the predicate is expressed by a definite noun phrase (*an fhírinne* “the truth”);¹⁹ its form is feminine because the predicate NP is headed by a feminine noun (*fírinne*). According to de Bhaldráithe (1953: 138), who describes a dialect of the same area, it is largely an idiosyncratic matter of lexical distribution whether a masculine or feminine pronoun is used in these cases: feminine pronouns are traditionally found with certain feminine nouns (e.g. *ceist* “question”, *cúis* “cause”, *fírinne* “truth”) but not with others (e.g. *caoi* “way”, *tuairim* “opinion”). It is reasonable to assume that frequency of use and formularity have played an important role in defining the norm for each lexical item. Other studies of traditional varieties of the Galway Gaeltacht have highlighted that masculine pronouns are increasingly common in this type of constructions, regardless of the gender of the predicate noun phrase (cf. Ó Curnáin 2007b: 1260f.).

19. On the insertion of this pleonastic pronoun see Ó Nualláin (1911: 34), Carnie (1995: 165) and Doherty (1996).

4.3.1 *The older generation corpus*

As shown in Table 9, while almost all masculine pronouns in the older generation corpus have masculine antecedents, only 43% of the feminine pronouns have feminine antecedents; almost 60% of them are co-referential with masculine nouns. This is due in part to semantic agreement and in part to historically well-established idiosyncrasies in pronominal agreement, as follows:

- a. Ten of the 57 feminine pronouns in question are instances of semantic agreement: as we have seen, masculine nouns sometimes have female referents, in which case agreement is normally syntactic within the noun phrase and semantic with anaphoric pronouns.
- b. Another 42 feminine pronouns co-refer with nouns that belong in a very special lexico-semantic category, which includes masculine nouns like *bád* “boat”, *soitheach* “vessel” and *púcán* “(a type of fishing boat)”. According to Ó Siadhail (1984), members of this lexico-semantic category are traditionally referred to by feminine pronouns, irrespective of grammatical gender, in all dialects.

Table 9. Gender of pronouns and of their antecedents in the older generation corpus

	Total	With masc. antecedent	With fem. antecedent
Masc. pronouns	359	351 (98%)	8 (2%)
Fem. pronouns	100	57 (57%)	43 (43%)

These 52 (10+42) pronouns were excluded from the count because their use with masculine antecedents cannot be considered a recent innovation. As Table 10 shows, agreement with masculine controllers is significantly more frequent than with feminine ones: overall, the agreement rate is 99% with masculine antecedents and 84% with feminines. Excluding semantic agreement, the picture is essentially unchanged: the agreement rate is 98% with masculine controllers, 83% with feminine controllers.

Table 10. Pronominal agreement in the older generation corpus

	Overall		Excl. sem. agr.	
	Masc.	Fem.	Masc.	Fem.
Agreeing	351	43	124	40
Total	354	51	127	48
%	99	84	98	83
	p < 0.0001		p = 0.0016	

When we break these figures down by pronoun type (simple, prepositional and possessive), similar proportions are obtained for simple and prepositional pronouns, as shown in Table 11; however, the corresponding counts get smaller and the difference appears to be only borderline statistically significant ($p = 0.0407$ and $p = 0.0675$, respectively); as regards possessive pronouns, the figures must be treated as merely suggestive due to the fact that very few of them were observed with feminine antecedents. It cannot therefore be determined whether possessive pronouns, which mark agreement via initial mutation, show a different agreement pattern compared to simple and prepositional pronouns, which do not.

Table 11. Pronominal agreement in the older generation corpus by pronoun type, excluding semantic agreement

	Simple		Prep.		Poss.	
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
Agreeing	82	20	32	20	10	0
Total	85	24	32	23	10	1
%	96	83	100	87	100	0
	p = 0.0407		p = 0.0675			

4.3.2 *The contemporary corpus*

Whereas in the older generation corpus there was a rather high incidence of feminine pronouns whose antecedent was grammatically masculine, this is not the case in the contemporary corpus. Here, as shown in Table 12, only four feminine pronouns out of 23 (17%) have masculine antecedents.²⁰

Table 12. Gender of pronouns and of their antecedents in the contemporary corpus

	Total	With masc. antecedent	With fem. antecedent
Masc. Pronouns	229	212 (93%)	17 (7%)
Fem. pronouns	23	4 (17%)	19 (83%)

For masculine pronouns with grammatically feminine antecedents, the figures in Table 12 tell us that they represent about 7% of them (17 pronouns). Among these, we find:

20. One is a straightforward case of semantic agreement with the masculine epicene *cara* ‘friend’, whose referent is a woman in this context.

1. simple and prepositional pronouns co-referential with feminine *áit* “place”;
2. simple pronouns co-referential with feminine *argóint* “argument”;
3. simple pronouns co-referential with feminine *áilleacht* “beauty” and *timpeallacht* “environment”;
4. simple pronouns co-referential with feminine *Gaelainn* “Irish language”;
5. possessive pronouns co-referential with feminine *cuid* “portion”, *obair* “work”, *ceist* “question”.

For some of these occurrences a few considerations may be added. To begin with, *áit* co-occurring with a masculine pronoun is not a recent innovation: De Bhaldraithe (1953: 135–136) lists it as one of the few feminine nouns taking masculine anaphors in Cois Fhairrge Irish, and the same is observed by Ó Sé (2000: 87) about the Corca Dhuibhne variety.²¹ Nevertheless, at least one feminine pronoun as the antecedent of *áit* is recorded in the RnaG section of the contemporary corpus. Furthermore, no oscillation in the gender assignment of *áit* is ever recorded within the noun phrase.

Áilleacht and *timpeallacht* may be grouped together for morphological reasons: the derivative suffix *-acht* (/əxt/) is normally associated with feminine gender, but non-palatalized consonant endings such as this are otherwise associated with the masculine. Moreover, *áilleacht* (like *argóint*) is abstract and non-concrete, a category which has been linked to the overgeneralization of masculine pronouns in Irish (cf. Ó Curnáin 2007b: 1259).

As can be seen in Table 13, the agreement rate is 98% with masculine controllers and only about 53% with feminine ones. If semantic agreement is excluded, the gap is even wider: 98% with masculine controllers as opposed to just 26% with feminine ones. In both cases, the difference is significant.

Table 13. Pronominal agreement in the contemporary corpus

	Overall		Excl. sem. agr.	
	Masc.	Fem.	Masc.	Fem.
Agreeing	212	19	145	6
Total	216	36	148	23
%	98	53	98	26
	p < 0.0001		p < 0.0001	

Only six feminine pronouns agree with inanimate antecedents, four of which co-refer with the same lexical antecedent (feminine *móin* “turf”), one with the feminine place name *Inis Oírr* (the name of an island). This suggests that the occurrence

21. Cois Fhairrge and Corca Dhuibhne are two Gaeltacht areas representative of the Connacht and Munster varieties, respectively.

of feminine pronouns with inanimate referents is less frequent in contemporary spoken Irish than it used to be.

Agreement rates with masculine antecedents, on the other hand, are very high for all three pronoun types, as shown in Table 14: 99% for simple pronouns (98% not counting semantic agreement), 98% for prepositional pronouns (100% not counting semantic agreement)²² and 92% for possessive pronouns (80% not counting semantic agreement).²³

Table 14. Pronominal agreement with masculine controllers in the contemporary corpus

	Overall			Excl. sem. agr.		
	Simple	Prep.	Poss.	Simple	Prep.	Poss.
Agreeing	141	60	11	104	37	4
Total	143	61	12	106	37	5
%	99	98	92	98	100	80

4.3.3 *Older and contemporary corpus compared*

Feminine pronouns are not very frequent in the contemporary corpus on the whole, and they are quite rare with inanimate antecedents. Syntactic agreement with feminine antecedents is significantly more frequent in the older than in the contemporary corpus: on the whole, the agreement rate with feminine antecedents is 84% in the older generation corpus and 53% in the contemporary one; excluding semantic agreement, it is 83% in the older generation corpus and just 26% in the contemporary one. Pronominal agreement with masculine controllers, on the other hand, is stable across the two components whether or not semantic agreement is counted.

4.4 Summary

In relation to the article, agreement in the older generation corpus is as frequent with masculine controllers as it is with feminine ones; in the contemporary corpus, on the other hand, agreement with masculine controllers is significantly more

22. The higher agreement rate recorded with prepositional pronouns when semantic agreement is not counted is due to one instance of agreement clash (where the noun is grammatically masculine but the referent is a female, and pronominal agreement is semantic). When semantic agreement is excluded, all prepositional pronouns match the grammatical gender of their antecedents.

23. For pronominal reference to feminine controllers, the hypothesized difference in agreement rate with the three different types of pronouns could not be assessed for the contemporary corpus, since 29 out of the 36 pronouns with feminine antecedents are simple.

frequent than with feminine ones. Finally, while the likelihood of syntactic agreement with masculine controllers is the same in both corpora, syntactic agreement with feminine controllers is significantly more likely in the older generation corpus.

With adjective agreement, as was the case for the article, there is no difference in the older generation corpus between masculines and feminines: both genders display comparable agreement rates. The same does not apply, however, to the contemporary data: here, syntactic agreement is significantly more frequent with masculine than with feminine nouns. Again, the likelihood of syntactic agreement with masculine nouns is the same in both corpora, whereas syntactic agreement with feminines is significantly higher in the older generation corpus.

Outside the noun phrase, gender agreement in Irish is found in third-person singular pronouns. Pronominal agreement with feminine antecedents was found to differ from local agreement in that it was significantly less frequent than with masculines in both corpora. Finally, while the likelihood of syntactic agreement with masculine antecedents was the same in both corpora, feminine antecedents in the contemporary corpus were more likely to be referred to by masculine pronouns, especially when the referent was not semantically feminine.

Figure 1 summarizes our findings. Overall agreement rates are shown in the top row, while in the bottom one are those calculated without counting semantic agreement. (As we have seen, including or excluding semantic agreement only makes a difference in the case of pronominal agreement.)

A diachronic comparison can be made by observing the difference between the darker- and lighter-coloured bars in each barplot, representing older-generation and contemporary corpus rates, respectively. What clearly emerges from Figure 1 is the striking diachronic difference between patterns of agreement rates with controllers of different genders: an apparently stable system in the case of masculine controllers as opposed to an evolving one in the case of feminine controllers.

No between-corpus difference in agreement rates with masculine controllers (the two barplots on the left-hand side) is statistically significant, and as we observed they are all well over 80%. As regards agreement with feminine controllers (on the right-hand side of Figure 1), on the other hand, all diachronic differences are statistically significant for all target types, with the higher agreement rates observed in the older generation corpus. Synchronically, i.e. within each corpus, the tendency is for agreement rates with feminine controllers to be higher with the article and lower with attributes and pronouns. When semantic agreement is ignored, pronominal agreement with feminine controllers shows the lowest agreement rates in both corpora. Moreover, pronouns represent the only target type for which a statistically significant difference in agreement rate between masculine and feminine controllers shows up in both corpora.

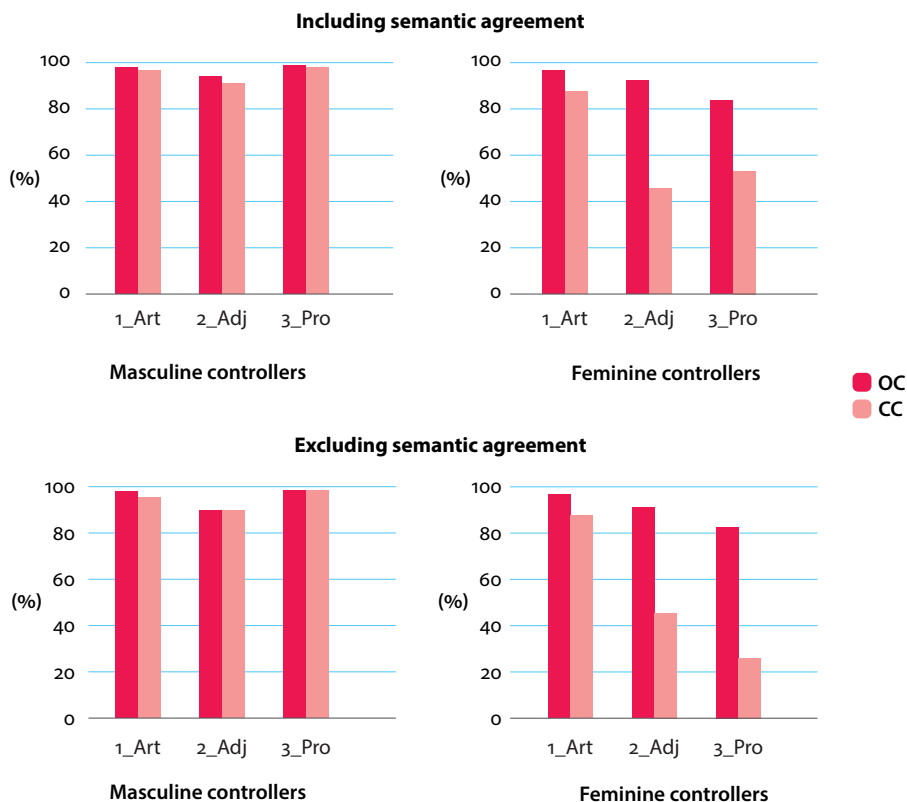


Figure 1. Agreement rates in Irish by corpus, gender and target type

5. Discussion

Throughout this paper the unqualified term ‘agreement’ (in the phrase “agreement rate”) has been used to refer to what Corbett (1991) calls ‘syntactic’, and others ‘formal’ or ‘grammatical’, agreement. As noted, syntactic agreement may occasionally differ from semantic agreement, as is the case of masculine *cailín* ‘girl’ in Irish. The possibility for nouns such as *cailín* to systematically take syntactic and semantic agreement forms depending on the agreement target can be referred to as split agreement and the probability of different agreement forms with different

targets is described by the Agreement Hierarchy (see e.g. Corbett 1991: 226).²⁴ In our corpora, such cases are marginal and hardly represented.²⁵ There is however another type of clash, or competition, namely that between the traditional gender system, which tends to yield syntactic agreement consistently for all target types, and an emerging one, in which feminine agreement forms are much less frequent, especially in terms of pronominal agreement and notably with inanimates. Pronominal agreement in the new assignment system appears to be semantic only, with feminine forms consistently appearing only when the antecedent has a female referent; there is therefore a different kind of split, whereby agreement within the noun phrase reflects traditional gender assignment while pronominal agreement indicates semantic reassignment of feminine inanimates. Our data also show that within the noun phrase syntactic agreement with the article tends to be more frequent than with the adjective. This difference between article and the adjective cannot be described in terms of the Agreement Hierarchy, which only has a position for attributive agreement, and is likely to involve issues of frequency and lexical collocation: nouns are learned, and frequently heard, in conjunction with the article (cf. Friedmann & Biran 2003; Taft & Meunier 1998); the traditional agreement-marking mutation is therefore heard and produced quite frequently. The same can happen when a noun and an adjective appear in highly frequent collocations: in the contemporary corpus (RnaG “Barrscéalta” 23/10/1997), the same speaker produces the phrase *cuid mhaith* “a good amount” (twice), leniting the adjective *maith* after the feminine noun, and the phrase *tuiscint maith* “a good comprehension” (once), without lenition; although both *cuid* and *tuiscint* are (historically) feminine nouns, *cuid mhaith* is a frequent collocation and is found as such in standard dictionaries (cf. Ó Dónaill 1977: s.v. *cuid*). Although there were not enough observations to allow a more thorough investigation of this hypothesis, the above suggests that agreement within the noun phrase is on its way to becoming fossilized, i.e. no longer a productive phenomenon but a mere repetition

24. The Agreement Hierarchy (Corbett, 1991: 226) takes the form ATTRIBUTIVE < PREDICATE < RELATIVE PRONOUN < PERSONAL PRONOUN and predicts that if multiple agreement forms are allowed, semantic agreement will be more likely for targets on the right-hand side of the hierarchy. It also implies that there will be a single cut-off point at which the split may occur: we would not for instance expect to find a language in which semantic agreement is possible with predicates only, while attributive adjectives and pronouns must agree syntactically.

25. There are only 11 instances of assignment clash, five with masculine *cailín* “girl”, whose referent can only be a female human, five with masculine *madadh* “dog”, whose referent can be of either sex, but happens to be a female dog in these five instances, and one with masculine *cara* “friend”, whose referent again can be of either sex but happens to be a woman in this specific occurrence.

of historical mutation patterns largely dependent on frequency, given that it is strongest with frequent collocations and weakest with infrequent ones: structures that are more entrenched (in the sense of Langacker 1987), or “routinized” (Croft 2000: 72), tend to be more resistant to change (cf. Croft 2000; Bybee 2006). In other words, if agreement within the noun phrase were a productive phenomenon, we would not expect that there would be a significant difference between the likelihood of agreement of the article and that of the attributive adjective. If the reason why today’s speakers of Irish mutate certain nouns after the article, and certain adjectives after certain nouns, were that they ‘knew’ that initial mutation applies to feminine nouns, then they would mutate all feminine nouns and all adjectives following a feminine noun. The fact that mutation is applied differentially to articles and adjectives, being applied almost constantly to the former but significantly less so to the latter, strongly suggests that noun mutation after the article on the one hand and adjective mutation after the noun on the other are now independent of each other. The patterns of pronominal substitution observed in the contemporary corpus suggest that where productive, agreement is now semantic.

The comparison with other Celtic varieties, now no longer spoken natively or approaching extinction, is illuminating. Manx, generally thought to have died out in the second half of the last century (Thomson 1992: 101–102), belonged to the same branch of Celtic as Irish and Scottish Gaelic (so-called Q-Celtic) and used to have the same gender-marking devices and assignment system as Irish (Kneen 1973: 49–50). In its final stages, however, agreement within the noun phrase was extremely inconsistent while pronominal agreement had ceased to reflect the grammatical gender of inanimates (Thomson 1992: 118) and was always semantically determined (feminine for female animates, masculine for everything else; cf. Broderick 1993: 239; Kneen 1973: 50). Cornish, which had become extinct by the end of the eighteenth century (George 1993: 414), belonged to the other branch of the Celtic family (P-Celtic), like Welsh. Gender assignment and agreement marking in P-Celtic is not substantially different from that of Q-Celtic. In Late Cornish, however, gender assignment had become exclusively semantic, that is, based on the referent’s sex (Jenner 1904: 78), although some fossilized remnants of lenition in phrases headed by feminine singular nouns survived (Thomas 1992: 368). In East Sutherland Gaelic, which Dorian described more than thirty years ago as rapidly approaching extinction, gender marking with the adjective was almost completely absent, while the use of the historically masculine third-person singular pronoun was the norm “even when mutation mark[ed] the noun as feminine within the NP” (Dorian 1976: 280). As the comparison shows, these three Celtic varieties share the same evolution path leading from a mixed formal–semantic system to a semantic one.

6. Conclusion

In terms of grammatical gender, there is a difference between the variety of Irish spoken as a community language in Connemara until as late as the 1960s and the variety heard today on Irish-language media: in the former, syntactic agreement was consistently observed in both local and pronominal agreement, while in the contemporary variety this is no longer the case: feminine forms of the pronouns appear to be used almost exclusively with female antecedents. The process of simplification whereby a formal gender system changes into a semantic one (by losing formal assignment rules and only retaining a core set of semantic assignment rules) is typologically well established (Corbett 1991). The final stages of related languages like Cornish and Manx provide examples of erstwhile formal assignment systems having lost their formal component and transformed into semantic assignment systems.²⁶ Insofar as it can be demonstrated that the conservative norm still applying within Irish noun phrases is on its way to becoming a fossilized relic, contemporary spoken Irish can be said to be at an intermediate stage along the path from a partially formal to an only semantic assignment system, as suggested by the observation that the persistence of the conservative norm in today's spoken Irish is not categorical and might depend on factors like frequency and entrenchment.

The difference between the older and the contemporary corpus is smallest in terms of agreement rates within the noun phrase and most conspicuous in terms of pronominal agreement; within the noun phrase, feminine agreement is by far more consistent with the article than with attributive adjectives. It is therefore possible to find cases in which the historical gender of a feminine noun is only evident in article–noun agreement, with masculine agreement forms showing up increasingly more often with the attributive adjective and virtually always with pronouns.

To conclude, I point out some directions for further research. The size of the data set on which this analysis is based limited the investigation of issues like frequency and collocations. Since the power of the statistical test crucially depends on sample size, the effect of a number of factors could not be assessed because of an insufficient number of observations. A study based on a larger corpus of contemporary spoken Irish might also reveal a higher number of statistically significant differences where the data set I used only suggested borderline significance. For instance, it could not be determined whether different pronoun types are in fact characterized by different agreement rates. Nor could it be assessed whether particular consonantal onsets tend to be mutated less often, thus differentially affecting

26. One can also compare the case of English, whose gender system was also reduced to a semantically motivated, pronominally marked one (Curzan 2003: 43, 47–48).

agreement rates for targets like the adjectives. Adjectives appear to occupy the middle ground between conservative article agreement and innovative pronominal agreement, in the sense that – when semantic agreement is ignored – they agree with feminine controllers less often than the article does but more frequently than pronouns. As the overgeneralization of masculine agreement spreads from pronominal reference to the noun phrase, adjectives seem to be the next affected target type. It would be interesting to learn what factors are at play in the process: given that the semantic factor seems to be irrelevant within the noun phrase, can the impact of other factors such as collocation frequency, lexical frequency and consonantal onset be weighed?

Acknowledgments

This research was funded by a Government of Ireland Exchange Scholarship and an IRCHSS Postgraduate Scholarship. I am grateful to John Saeed, Brian Nolan and three anonymous reviewers for their comments; thanks are also due to Eamonn Mullins for his assistance with the statistical analysis of the data. I take full responsibility for any errors or oversights.

References

- Akutagawa, Michie. 1987. “A Linguistic Minority under the Protection of its Own Ethnic State: A case study in an Irish Gaeltacht”. *Third International Conference on Minority Languages: Celtic papers* ed. by Gearóid Mac Eoin, Anders Ahlqvist & Donncha Ó hAodha, 125–146. Clevedon: Multilingual Matters.
- Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511801686
- Breatnach, R. A. 1964. “The Irish Revival Reconsidered”. *Studies: An Irish quarterly review* 53(209).18–30.
- Broderick, George. 1993. “Manx”. *The Celtic languages* ed. by Martin J. Ball & James Fife, 228–285. London: Routledge.
- Bybee, Joan. 2006. “From Usage to Grammar: The mind’s response to repetition”. *Language* 82:4.711–733. doi:10.1353/lan.2006.0186
- Caighdeán. 1979. *Gramadach na Gaeilge agus litriú na Gaeilge: An caighdeán oifigiúil* (6th ed.). Baile Átha Cliath: Oifig an tSoláthair.
- Carnie, Andrew Hay. 1995. *Non-Verbal Predication and Head-Movement*. Ph.D. dissertation, MIT.
- Comrie, Bernard. 1979. “Morphophonemic Exceptions and Phonetic Distance”. *Linguistics* 17(215/216).51–60.
- Consten, Manfred, Mareile Knees & Monika Schwarz-Friesel. 2007. “The Function of Complex Anaphors in Texts: Evidence from corpus studies and ontological considerations”. *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference* ed. by Monika Schwarz-Friesel, Manfred Consten & Mareile Knees, 81–102. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/slcs.86.09con

- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9781139166119
- Croft, William. 2000. *Explaining Language Change: An evolutionary approach*. Harlow: Longman.
- CSO. 2006. *Census 2006*, Vol. IX: *Irish language*. Central Statistics Office Ireland. <http://www.cso.ie/en/census/census2006reports/census2006-volume9-irishlanguage/>. (Accessed: 8 September 2014.)
- CSO. 2011. This is Ireland. Highlights from Census 2011, Part 1. *Central Statistics Office Ireland*. http://www.cso.ie/en/media/csoie/census/documents/census2011pdr/Census2011_Highlights,Part,1,web,72dpi.pdf. (Accessed: 24 October 2014.)
- Curzan, Anne. 2003. *Gender Shifts in the History of English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486913
- Dahl, Östen. 1999a. "Animacy and the Notion of Semantic Gender". *Gender in Grammar and Cognition, Part I: Approaches to Gender* ed. by Barbara Unterbeck, 99–115. Berlin & New York: Mouton de Gruyter.
- Dahl, Östen. 1999b. "Elementary Gender Distinctions". *Gender in Grammar and Cognition, Part II: Manifestations of Gender* ed. by Matti Rissanen, Terttu Nevalainen & Mirja Saari, 577–594. Berlin & New York: Mouton de Gruyter.
- de Bhaldráithe, Tomás. 1953. *Gaeilge Chois Fhairrge*. Baile Átha Cliath: Institiúid Árd-Léinn Bhaile Átha Cliath.
- Doherty, Cathal. 1996. "Clausal Structure and the Modern Irish Copula". *Natural Language & Linguistic Theory* 14:1.1–46. doi:10.1007/BF00133402
- Dorian, Nancy C. 1976. "Gender in a Terminal Gaelic Dialect". *Scottish Gaelic Studies* 12.279–282.
- Dorian, Nancy C. 1977. "A Hierarchy of Morphophonemic Decay in Scottish Gaelic Language Death: The differential failure of lenition". *Celtic Linguistics* 1976 (=Word, 28), 96–109.
- Dorian, Nancy C. 1981. *Language Death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.
- Dorian, Nancy C. 1994. "Purism vs. Compromise in Language Revitalization and Language Revival". *Language in Society* 23:4.479–494. doi:10.1017/S0047404500018169
- Esslemont, Alan. 2007. *Telephone Interview with Alan Esslemont, Director of Television, TG4*, 22 January 2007.
- Frenda, Alessio S. 2011. *Gender in Insular Celtic: A functionalist account of variation and change in Irish and Welsh*. Ph.D. dissertation, Trinity College, Dublin. <http://goo.gl/8tVI0F>. (Accessed 4 February 2018.)
- Friedmann, Naama & Michal Biran. 2003. "When is Gender Accessed? A study of paraphasias in Hebrew anomia". *Cortex* 39:3.441–463. doi:10.1016/S0010-9452(08)70258-2
- George, Ken. 1993. "Cornish". *The Celtic Languages* ed. by Martin J. Ball & James Fife, 410–468. London: Routledge.
- Hickey, Raymond. 2003. "Reanalysis and Typological Change". *Motives for Language Change* ed. by Raymond Hickey 258–278. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511486937.016
- Jenner, Henry. 1904. *A Handbook of the Cornish Language chiefly in Its Latest Stages with Some Account of Its History and Literature*. London: David Nutt.
- Jones, Mari C. 1998. *Language Obsolescence and Revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Oxford: Clarendon Press.
- Kneen, John Joseph 1973. *A Grammar of the Manx language*. *Yn Cheshaght Ghailckagh* [The Manx Gaelic Society].

- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*, Vol. I: *Theoretical prerequisites*. Stanford: Stanford University Press.
- Mac Eoin, Gearóid. 1993. "Irish". *The Celtic Languages* ed. by Martin J. Ball & James Fife, 101–144. London: Routledge.
- Maguire, Gabrielle. 1991. *Our Own Language: An Irish initiative*. Clevedon: Multilingual Matters.
- McCone, Kim. 1994. An tSean-Ghaeilge agus a réamhstair [Old Irish and its prehistory]. *Stair na Gaeilge* [The history of Irish] ed. by Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach, 61–219. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Nic Phóidín, Caoilfhionn. 2003. 'Cén fáth nach?' – Ó chánúint go criól ['Why not?' – From dialect to creole]. *Idir lúibíní: Aistí ar an léitheoireacht agus an litearthacht* [In parentheses: Essays on reading and literacy] ed. by R. Ní Mhianáin, 113–130. Baile Átha Cliath: Cois Life.
- Ó Baoill, Dónall P. 1988. "Language Planning in Ireland: The standardization of Irish". *Language Planning in Ireland* ed. by Pádraig Ó Riagáin (= International Journal of the Sociology of Language, 70), 109–126. Berlin: Mouton de Gruyter.
- Ó Curnáin, Brian. 2007a. *The Irish of Iorras Aithneach County Galway*, Vol. I. Dublin: Dublin Institute for Advanced Studies.
- Ó Curnáin, Brian. 2007b. *The Irish of Iorras Aithneach County Galway*, Vol. II. Dublin: Dublin Institute for Advanced Studies.
- Ó Dónaill, Niall. 1977. *Fócloir Gaeilge-Béarla*. Baile Átha Cliath: Roinn Oideachais agus Eoiláiochta.
- Ó Flatharta, Peadar. 1989. "The Teaching of Irish as a First Language in Primary Schools in the Gaeltacht". *Teanga* 9.75–83.
- Ó hUiginn, Ruairí. 1994. "Gaeilge Chonnacht". *Stair na Gaeilge* ed. by Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach, 539–609. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Ó Murchú, Máirtín. 1969. "Common Core and Underlying Forms". *Ériu* 21.42–75.
- Ó Murchú, Máirtín. 1988. "Diglossia and Interlanguage Contact in Ireland". *Language, Culture and Curriculum* 1: 3. 243–249. doi:10.1080/07908318809525043
- Ó Nualláin, Gearóid. 1911. "Gadelica minora". *The Irish Ecclesiastical Record*, 4th series, Vol. XXX, 33–42, 137–144, 361–368, 501–508, 609–619.
- Ó Sé, Diarmuid. 2000. *Gaeilge Chorca Dhuibhne*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.
- Ó Siadhail, Mícheál. 1984. "A Note on Gender and Pronoun Substitution in Modern Irish Dialects". *Ériu* 35.173–177.
- Ó Siadhail, Mícheál. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620621
- Punch, Aidan. 2008. "Census Data on the Irish Language". *A New View of the Irish Language* ed. by Caoilfhionn Nic Phóidín & Seán Ó Cearnaigh, 43–54. Dublin: Cois Life.
- Stenson, Nancy. 1993. "English Influence on Irish: The last 100 years". *Journal of Celtic Linguistics* 2.107–128.
- Taft, Marcus & Fanny Meunier. 1998. "Lexical Representation of Gender: A quasiregular domain". *Journal of Psycholinguistic Research* 27.23–45. doi:10.1023/A:1023270723066
- Thomas, Alan R. 1992. "The Cornish language". *The Celtic Languages* ed. by Donald MacAulay, 346–369. Cambridge: Cambridge University Press.
- Thomson, Robert L. 1992. "The Manx Language". *The Celtic Languages* ed. by Donald MacAulay, 100–136. Cambridge: Cambridge University Press.

- Ua Súilleabháin, Seán. 1994. "Gaeilge na Mumhan". *Stair na Gaeilge* ed. by Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach, 479–538. Maigh Nua: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Wigger, Arndt, ed. 2000. *Caint Chonamara: Bailiúchán Hans Hartmann. Imleabhar IX. Ros Muc*. Bonn: Universität Bonn.

Numeral classifier systems in the Araxes-Iran linguistic area

Don Stilo

Max Planck Institute for Evolutionary Anthropology

This corpus-based study discusses numeral classifiers (NCs) in neighboring languages of disparate origins: Azerbaijani (Turkic) and within Indo-European both colloquial Armenian and Vafsi (NW Iranian). Colloquial Tehran Persian and some other smaller Iranian languages, an additional Armenian dialect, a peripheral dialect of Azerbaijani, two Neo-Aramaic dialects (Semitic), and colloquial Georgian (Kartvelian) are also marginally included. This study shows that languages of very different origins have developed typologically similar classifier systems with very parallel behaviors. Although NCs in these languages are robust in usage, they represent simple systems, generally consisting of only two members: (1) a universal NC meaning ‘seed, grain’ used for all noun types (including human), and (2) an NC meaning ‘person’ optionally used for humans. The statistics of frequency of each classifier in these languages are tabulated. The diachrony of NCs in the area, their relationship to Greenbergian typology, their areality, and their fade-out phenomena are also discussed.

1. Introduction

In this article I discuss the areal distribution of numeral classifier systems based on my current work on the *Atlas of the Araxes-Iran Linguistic Area* (Stilo in preparation), henceforth the *Atlas*. This area, with the Araxes river flowing more or less through its center, includes northern and central Iran, the Republic of Azerbaijan, Armenia, Georgia, eastern Turkey, and northern Iraq. The specific subset of these languages dealt with in this paper are presented towards the end of this section. The goal of this *Atlas* is to show the multitude of isoglosses shared by the languages of this zone, how languages of different families have converged even in small details within these isoglosses, and how far-reaching some of these areal patterns are.

Below, I would like to discuss the similarities and differences in the classifier systems in the Araxes-Iran zone – they are robust in about half these languages – and to demonstrate the areal aspects of their usages. Languages of very different

origins here have developed the same type of small-inventory classifier systems, mostly with very similar behaviors. We shall see that the local Turkic languages, Azerbaijani and Turkish, have borrowed the forms and their usage from Iranian languages and have completely integrated them into their noun phrase domains. Aramaic dialects of the area have also been affected by Iranian classifier systems, but much more marginally than Turkic.

Another areal aspect that I present in the *Atlas* is a discussion of the diminishing of various types of linguistic features at the periphery of any given isogloss, a phenomenon I call ‘fade-out’. Numeral classifiers are one interesting example of this process. Some languages of the Araxes-Iran zone located on the western and northern periphery of the classifier isogloss show a pronounced diminishing of the category of numeral classifier and are transitional to areas where there are no vestiges of classifiers.

Except for the few examples of Colloquial Persian, provided from personal knowledge, the present study is corpus-based and presents statistics of frequency of classifiers in various contexts in some of the relevant languages. While these corpora are not very large, they still prove to be helpful in discovering the classifier patterns of the languages of the area.

This article treats the following issues regarding numeral classifiers: their forms and lexical meanings in the various languages affected and their spread into other languages through borrowing (§2.1), the similarities and differences in their usages as well as their frequency in the languages involved (§3.1–§3.3–§3.5), their diachrony in the area (4), Greenbergian typology of numeral classifiers (§5), and their reality and fade-out phenomena (§6).

The languages most prominently discussed in this chapter are representatives of Turkic – Colloquial Azerbaijani (also Azeri) of Tabriz (Iran) – as well as of two genera of Indo-European – Colloquial Yerevan Armenian and Vafsi (NW Iranian, Tatic, Tati). To a lesser extent, Colloquial Tehran Persian (SW Iranian), Muslim Caucasian Tat (SW Iranian, spoken in the north of the Republic of Azerbaijan), Gilaki, Mazanderani (both: NW Iranian, Caspian), and Masulei (NW Iranian, Tatic, Southern Talyshi) are also included. Two other families – Semitic (Jewish Neo-Aramaic dialects of Sanandaj and Urmia) and Kartvelian (Colloquial Georgian) – plus Gawraji (NW Iranian, a mixed Hawrami-Central Kurdish language of Iran) and one additional dialect each of Armenian (Lorri), Tati (Koluri), and Azerbaijani (Bayat, Iran) are also marginally discussed. It is impossible to provide a comprehensive, in-depth study of the role of classifiers within the grammars of the 15 or so languages discussed here, and I have had to exclude many language varieties of the families introduced here that also have classifiers.

Although numeral classifiers in the Araxes-Iran zone are robust in usage, they represent simple systems, generally consisting of only two classifiers: (1) a universal

numeral classifier (UNC) meaning ‘seed, grain’ used for all noun types (including human), and (2) a numeral classifier meaning ‘person’ sometimes used for humans (HNC). My investigation only includes sortal numeral classifiers and not mensural classifiers (see Gil 2005: 226).

The issue of numeral classifiers also involves a separate isogloss from the *Atlas*, which I occasionally mention here as a general corollary to classifiers: the use of singular forms of nouns after numerals, except for head nouns that are definite – “these five girls”, etc. – but even in such cases plurality is not obligatory in these languages. See §5 for Greenberg’s discussion of this correlation. Armenian, Vafsi, and Gilaki, however, do occasionally use a plural after a numeral (\pm classifier), even when the noun is not definite (see (8) below).

2. Overview of numeral classifiers

2.1 Numeral classifiers: forms and meanings

2.1.1 *The UNC for inanimate, human, and other animate head nouns*

The most common form of the universal numeral classifier in Iranian is *dane*, *danæ*, *dune* “seed, grain/kernel (of wheat and other cereals)”. (1) shows its original lexical sense:

- (1) Vafsi (Stilo 2004: 118, #38)
váxd-e dánæ besd-æn-i, in gændem æ-čuzen-e.
 time-LINK grain close-INF-OM this wheat DUR-burn-3s
 “At the time to form grains (i.e., ripen), this wheat burns.”

Vafsi has two variants: (1) feminine *dánæ* “grain of wheat (barley, etc.); chicken/bird feed; numeral classifier” and (2) *dané* “numeral classifier”:

- (2) Vafsi (*southernmost of Tati group, Tatic, NW Iranian*): *dánæ* ~ *dané*
 (Stilo 2004: 42, #120)
do dané zendani h-e.
 two UNC prisoner exist-3s
 “There are two prisoners.”

All dialects of Talyshi and a few neighboring Tati languages use the root *g(ə)læ*, which also means “seed, grain/kernel (of cereals)”. In these languages, the combination of *g(ə)læ* with the numeral *i* “one” usually yields a lexicalized form *illæ* (< *i-glæ*): The Gawrajui numeral classifier *gilæ* possibly also comes from this root. Example from Talyshi:

- (3) Masulei (*Southern Talyshi, Tatic, NW Iranian*): g(ə)læ (Lazard 1979: 62)
 ... čəmən se=glæ æsb... ba se=glæ kəla... ba
 my three=UNC horse with three=UNC hat with
 se=glæ šəmšir...
 three=UNC sword
 “... my three horses ... and the three hats ... and the three swords ...”

The form *dane/æ* has been borrowed into Turkish and Azerbaijani from the local Iranian (but non-Persian) substratum of Azerbaijan. The Turkish *tane* “kernel, grain (of a cereal plant; of sand, salt, sugar); single thing, item, piece” (Redhouse 2007: 827) has undergone a Turkish sound change involving devoicing of initial stops. Azeri dialects have three variants, *danæ*, *dana*, and *dænæ*, the latter two derived via Turkic rules of vowel harmony:

- (4) Colloquial Azerbaijani (*Tabriz dialect*): dana (Kıral 2001: 172, #93)
 iki dana ayri məyulæ=di-lær.
 two UNC other category=COP-PL
 “They are two different categories.”

The Armenian¹ universal numeral classifier is *haṭ* “grain, seed, piece” (Martirosyan 2009: 392):

- (5) Colloquial Armenian: haṭ (eanc: OSD Polylogue 060)
 yerku haṭ aman piṭi Ø-lvana-yir.
 two UNC plate NECESSITATIVE SBJV-wash-2s:PAST
 “You were supposed to wash two dishes.”

A second universal numeral classifier: In some Iranian languages, there is an additional universal numeral classifier: *ta*, usually glossed as “item, piece”. This universal numeral classifier appears in Colloquial Persian and Muslim Caucasian Tat (*ta* ~ *tæ*) only for numerals above 1 – while the numeral 1 takes the other universal numeral classifier, *dune* in colloquial Persian, *dənə* in Tat. Caspian languages, a group of NW Iranian including Gilaki and Mazanderani, have no equivalent of *dane* and, instead, use only *ta* with all numerals as well as those quantifiers that take classifiers, irrespective of the animacy or number of the noun.

1. My transcription for Armenian and Georgian differs from standard practice for these languages: I mark both the voiceless aspirated (*pʰ, tʰ, kʰ*; etc.) and glottalized stops (*p̚, t̚, k̚*; etc.) with special diacritics. In standard practice, the latter are left unmarked (*p, t, k*; etc.). My transcription of *i, ʒ, j* also differs from the *ə, j, j̄* of standard practice.

- (6) Mazanderani, Kordkheyl dialect (*Caspian, NW Iranian*): = ta
(Borjjan & Borjjan 2008: II 03)

debis=ta gō dašt-εmε.
200=UNC cow have:PAST-1S
“I had 200 cows.”

2.1.2 *The human numeral classifier (HNC)*

In addition to the universal numeral classifiers introduced above, which are the preferred choice with human nouns, a dedicated human numeral classifier is often used for human nouns in the Araxes-Iran zone. In most Iranian languages this classifier is *næfær* “person”, a loanword from Arabic. In Azeri *næfær* is also common, but we encounter the Turkic word *kiši* “person, man” as a classifier as well.

- (7) Vafsi (*Tatic, Tati, NW Iranian*): *næfær* (Stilo 2004: 124, #3)
in se næfær hæzræt-e musa-y ræfeq-an ær-goā...
this three HNC saint-LINK Moses-OM friend-OP DUR-want:PAST
“these three friends of Moses wanted ...”

The equivalent Armenian HNC, *hok'i* “person” has the basic meaning “soul”.

- (8) Colloquial Armenian: *hok'i* (eanc: OSD 09–1)
yerku hok'i p'esa-ner un-en.
two HNC son-in-law-PL have-3P
“They have two sons-in-law.”

2.1.3 *Numeral classifiers and the formal written vs. colloquial registers*

An important point to keep in mind in examining classifiers for this article is that I have concentrated on the colloquial registers of the languages investigated for various reasons:

1. The universal numeral classifiers for Persian and Armenian are often avoided in the formal written registers. Mache (2012: 29) notes that “the language that is used in the literature, newspaper articles, lectures, is heard on the radio, etc, excludes certain forms of numeral classifier expression that are, in fact, quite common in colloquial speech ...”.
2. The classifier system of Written Persian differs significantly from modern colloquial even when spoken in formal situations among educated speakers with a certain social distance. Formal Persian clearly has a very complex classifier system: Mache (2012: 73–87) lists 129 sortal numeral classifiers, but I find that some 85 of these are either mensural classifiers (“glass” > “a GLASS of wine”), are semantically non-mensural, non-sortal words (*xuše* “a CLUSTER (of grapes)”,

implying a shape² rather than a quantity), or are problematic in other ways. This still leaves us with an impressive 45 sortal classifiers.

3. The full classifier system of Persian is rather analogous to the collective nouns specific to certain animals in English. That is, while most speakers of English know and use “a GAGGLE of geese”, “a SCHOOL of fish”, or “a PRIDE of lions”, most well-educated speakers, in realistic terms, will not know “a SLEUTH of bears”, “a LEASH of foxes”, or “a KNOT of toads”. Wikipedia³ lists 236 such terms. While the 45 sortal classifiers of Persian and the 236 collective animal nouns of English are valid terms in these languages, they do not constitute common knowledge among the general populace on the same level as the number of classifiers in Chinese do for native speakers of Chinese. The majority of classifiers on Mache’s list are really only known to highly educated people. Even Mache states in various places that some of the classifiers she lists are considered archaic.
4. If we include the system of numeral classifiers from the formal written registers of the language, Persian stands out as being completely different from all the languages of its neighbors. On the other hand, as we shall see below, the classifier system in the modern colloquial registers proves to be very similar in usage and inventory to most of the surrounding languages, both those directly related, i.e., Iranian languages, as well as those not directly related, particularly Armenian and Azerbaijani.

For an excellent, exhaustive study of classifiers in both Written and Colloquial Persian, I highly recommend Mache (2012).⁴ This valuable study came to my attention only after having completed most of the present article. My goal here, however, in contrast to Mache’s work, her Ph.D. dissertation, is not to provide a comprehensive exposition of classifiers in any one language of the area. My goal, rather, is to discuss the local areal distribution of classifier systems and to show their parallelisms, similarities, and differences.

Four formal Persian classifiers occur more often than the other sortal classifiers mentioned above in modern written Persian: *tæn* “body” for persons, *jeld*

2. Note the cognate word in Vafsi and a derivative from it focusing on shape: *hušæ* “cluster, bunch (of grapes)” vs. *hušaleng* “vine segment after grapes have been removed” (Stilo fieldwork). Clusters of grapes can vary greatly in size and number of grapes they carry and can themselves be counted.

3. http://en.wikipedia.org/wiki/List_of_animal_names. (Accessed: 25 April 2017.)

4. See also Lazard (1957: 91–92), Lambton (1953: 43–44), and Thackston (1978: 112) for discussions of classifiers. Gebhardt (2009) and Mahootian (1997), however, give somewhat fuller discussions of numeral classifiers for Persian, as does Schroeder (1999) for Turkish, but Mache’s work seems to be the most complete and up-to-date study.

“volume” for books, *parče* “cloth” for villages, and *ædæd* “number” for various nouns. Note that Vafsi, an unwritten language, also (but rarely) uses the classifier *parče* for counting villages as well as children. Colloquial Persian also sometimes uses *jeld* for books.

3. Behavior of numeral classifiers

3.1 Usages of numeral classifiers

In most of the relevant languages of the area – but not all – classifiers are optional even in colloquial registers. (See §3.5 for a discussion of their obligatory use as heads of NPs.) The following pairs of sentences with the same head nouns show the optionality of classifiers:

- (9) Colloquial Armenian (eanc: OSD 09–4)
yerku haṭ yerexa un-en.
 two UNC child have-3P
 “They have two children.”
- (10) Colloquial Armenian (eanc: OSD Polylogue 118)
yerku yerexa un-i.
 two child have-3s
 “She has two children.”
- (11) Colloquial Azerbaijani (*Tabriz dialect*) (Kiral 2001: 144, #53)
tæk=o tænha bi dana otax-da zendegi el-ir-Ø.
 sole=and alone one UNC room-LOC life make-DUR-3S
 “She lives all alone in one room.”
- (12) Colloquial Azerbaijani (*Tabriz dialect*) (Kiral 2001: 198, #30)
iki otay-i var=ïdi-Ø.
 two room-3s:POSS exist=COP:PAST-3S
 “It had two rooms.”
- (13) Azerbaijani (*peripheral dialect, Bayadistan*) (Bulut 2006: 337, #1)
on-næn ikki dæne qiz var-Ø.
 s/he-ABL two UNC girl exist-3s
 “He has two daughters from her.”
- (14) Azerbaijani (*peripheral dialect, Bayadistan*) (Bulut 2006: 334, #3)
altï qiz=ide-g.
 six girl=COP:PAST-1P
 “We were six girls.”

- (15) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 146, #32)
šeš otáq-i kellæ (æ)d-do-Ø leyle-y
 six room-OM key DUR-give-3S boy-OM
 “He gives the boy the keys of the six rooms.”
- (16) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 146, #30)⁵
hæfd dánæ otaq=es dærdæ.
 seven UNC room=3S:OBL have:PAST
 “He had seven rooms.”
- (17) Masulei (*Tatic, Southern Talyshi, NW Iranian*) (Lazard 1979: 62)
padeša se=glæ kellæ... æ pir-æ merdæk-e
 king three=UNC girl that old-LINK man-OBL
vær-e mæn-en=æ.
 near-OBL BE₂-3P=COP:3S
 “The king’s three daughters ... are at that old man’s house.”
- (18) Masulei (*Tatic, Southern Talyshi, NW Iranian*) (Lazard 1979: 60)
padeša se kellæ=š av-e æ dem-e=ko
 king three girl=3S:OBL water-OBL that side-OBL=ABL
bær-oærd=æ.
 PV-bring:PAST=COP:3S
 “He got the king’s three daughters out from the other side of the water.”

5. The interlinear abbreviation OBL in this article refers to a secondary set of person agreement markers (PAMs) in the following functions: (A) encoding of pronominal possessives (including the pronominal adjunct of an adposition, especially those of nominal origin); (B) encoding of pronominal objects (Patients) in the verb; (C) encoding of pronominal Experiencers in experiencer verbs; (D) cross-indexing of Agents in languages that have ergative marking (or vestiges of it in the verbal paradigms); and (E) certain other usages not appearing in this article (e.g., Recipient).

These oblique PAMs do not encompass all the above functions in all languages, e.g., Armenian and Azerbaijani (A; occasionally C), Georgian (B and D not appearing in this article; for C see (77)), Masulei and Koluri (C not appearing in this article; D) but Vafsi, Gawrajui, Neo-Aramaic (A–E). In languages that have tense-based split alignment (Vafsi, Masulei, Koluri, Gawrajui, Neo-Aramaic), functions B and D only appear in the relevant tenses. These oblique PAMs in the Agent and often the Experiencer functions are generally leftwardly mobile in Iranian languages (but not Aramaic). Caspian languages (Gilaki, Mazanderani) have no oblique PAMs in any domain.

Additionally, in some Iranian languages the interlinear abbreviation OBL also refers to an Oblique nominal case that encodes exactly the same functions and the oblique PAMs. For this article these languages include Vafsi, Masulei, and Caucasian Tat (the latter with a case-like oblique enclitic).

Note both options with similar noun types in the same Masulei sentence:

- (19) Masulei (*Tatic, Southern Talyshi, NW Iranian*) (Lazard 1979: 64)
hær se bəra padeša se=glæ kellæ bær-ən.
 each three brother king three=UNC girl take-3P
 “All three brothers take the king’s three daughters (in marriage).”

Classifiers, however, are not optional throughout the area. This corpus-based study indicates that the use of classifiers is fully obligatory in Mazanderani, Gilaki, and Muslim Caucasian Tat. They make exclusive use of *ta*, even with human nouns (but see (40) for a type of exception in Muslim Caucasian Tat). The Borjjan & Borjjan (2008) Mazanderani texts show no tokens of the human numeral classifier *næfær*, and its usage was rejected by a speaker of urban Mazanderani who only accepted, e.g., *dε=ta tajer* “two merchants” instead (Forough Ebadian, p.c.).

- (20) Mazanderani, Kordkheyl dialect (*Caspian, NW Iranian*) (Borjjan: II 04)
ε=ttā rika pier-e pæli Ø-mundess-ε.
 one=UNC boy father-LINK by DUR-stay:PAST-3S
 “One son would stay with father.”
- (21) Mazanderani, Kordkheyl dialect (*Caspian, NW Iranian*) (Borjjan: II 04)
on mōqe hæmε ε=ttā ja dæ-imi.
 that time all one=UNC place PV-1P
 “Those times we were all in one place.”
- (22) Gilaki, Rashti dialect (*Caspian, NW Iranian*) (Rastorgueva 1971: 264, #12)
nimsaat nú-guzəšt-ə=bu-Ø ki du=ta duzd
 half.hour NEGv-pass:PAST-PP=AUX:PAST-3S SUBR two=UNC thief
úya b-amo-Ø.
 there PU-come:PAST-3S
 “A half hour had not passed when two thieves came there.”
- (23) Gilaki, Rashti dialect (*Caspian, NW Iranian*) (Rastorgueva 1971: 236, #38)
un-ə=ja yəkmahə dú=ta utay kərayə bu-kúd-im.
 s/he-LINK=ABL one.monthly two=UNC room rent PU-make:PAST-1P
 “We rented two rooms from him for one month.”
- (24) Muslim Caucasian Tat, Daykušču dialect (SW Iranian) (Grjunberg 1963: 126, #98)
injæ amarəgar-ha sə čar=ta ærməni bu-Ø.
 here comer-PL three four=UNC Armenian be:PAST-3S
 “The ones who came here were three (or) four Armenians.”

- (25) Muslim Caucasian Tat, Daykušču dialect (SW Iranian) (Grjunberg 1963: 151, #17)
mæn=æ bæ qatür=mæn dü=tæ xasiyét híst-ü.
 I=OBL to mule=1S:POSS two=UNC feature EXIST-3
 “My mule has two (special) features.”

Human numeral classifier for human nouns: While numerals with human nouns mostly take the universal numeral classifier, as in (2), (9), (13), (17), etc., the human numeral classifier is also a valid option for human nouns, except for the three languages mentioned in the preceding paragraph:

- (26) Colloquial Azerbaijani (*Tabriz dialect*) (Kıral 2001: 214, #63)
šah döwræ-sin-dæ beš on næfær adam gızil-i
 king around-3S:POSS-LOC five ten HNC person gold-ACC
čiy-ar-dı-lar.
 come.out-CAUS-PAST-PL
 “The Shah (and) five (or) ten people around him got the gold out.”
- (27) Colloquial Armenian (eanc: Polylogue 127)
yes ink^ç-is un-ec^ç-el=em 22 hok^çanoc xorhurd u mer
 I self-1S:POSS have-PAST-PP1=1S 22 person:ADJ council and our
faķulteṭ=i 220 hok^çi usanoy.
 department=DEF 220 HNC student
 “I myself had a 22-person council and our department, 220 students.”
- (28) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 88, #39)
se næfær tajér-i ke=san æ-čappa.
 three HNC merchant-OM house=3P:OBL DUR-ransack:PAST
 “They would ransack three merchants’ houses.”
- (29) Masulei (*Tatic, Southern Talyshi, NW Iranian*) (Lazard 1979: 42)
əm kellæ əm se næfær ræfey-e=ko kæm i-llæ ræs-ə?
 this girl this three HNC friend-OBL=ABL which one-UNC arrive-3S
 “Of these three friends, which one does this girl go to?”

3.2 Disallowed uses of numeral classifiers

In all languages of the Araxes-Iran zone, sortal classifiers never occur with units of money, time, or measurements. For examples, see (35) (“one day”) and (38) (“one year”). This point is also noted by Greenberg (1990: 186) as a cross-linguistic generality. Additionally, classifiers are not licensed with demonstratives in NPs without a numeral in these languages. Numerals and the ‘one’ quantifier (see §3.4) are the only elements that trigger the use of classifiers.

Greenberg (1990: 186) notes another case also valid here: the word for ‘person’ never functions as its own classifier, e.g., Persian *do næfær næfær > *do næfær* ‘two people’. The only acceptable pattern in this vein occurs in languages with more than one word for ‘person’, e.g., Armenian *100 hok’i mart* ‘[we invited] 100 people’ (eanc: Polylogue OSD 10–12) and Azerbaijani *otuz næfær kíši* ‘thirty people’ (Kiral 2001: 208, #37); see also Example (26)).

3.3 The numeral ‘1’

‘1’ differs from other numerals with regard to classifiers in three ways: (1) ‘1’ also serves as an indefinite article, mostly accompanied by a classifier in those languages that have classifiers (§3.3.1); (2) the form of the classifier with ‘1’ vs. other numerals may differ (§3.3.2); and (3) the frequency of the classifier with ‘1’ vs. other numerals may differ (§3.3.3).

3.3.1 ‘1’ as an indefinite article

In this zone, ‘1’ often serves as an indefinite article, which can be obligatory, optional, or rare, depending on the language. ‘1’ as a numeral is usually formally indistinguishable from ‘1’ as an indefinite article, but the frequencies may differ drastically between these two usages. Some languages have a formally distinct indefinite article.⁶ This article may be optionally reinforced by the numeral ‘1’, in which case a classifier may be used. Also, since indefinite articles in some languages are themselves optional, indefinite NPs may occur as bare nouns without ANY article or classifier, ‘1’ or otherwise. Gil (2005: 227) notes that such bare nouns “may be understood as either mass or count, and as either singular or plural.”⁷

(30) Colloquial Persian

sib mí-xor-i?

apple DUR-eat-2s

‘Do you eat apples? ~

Would you like (lit: eat) an apple/some apples?’

Use of UNCs with indefinite articles: The examples below compare ‘1’ as an indefinite article and ‘1’ as a numeral, both accompanied by a classifier in some Araxes-Iran languages. In all cases, the universal numeral classifier may be optionally deleted, but I have not provided these examples due to space restrictions.

6. Dedicated indefinite articles are usually diachronically derived from the numeral ‘1’ in these languages but the two elements are still formally distinct in the contemporary languages.

7. See also Corbett (2000: 9–19) for a discussion relevant to this issue.

- (31) Colloquial Armenian (eanc: OSD 10–5)
 ‘1’ as an indefinite article, UNC *haṭ*, ± human
mi haṭ arajark un-em.
 one UNC proposal have-1s
 “I have a proposal.”
- (32) Colloquial Armenian (eanc: Polylogue 124)
 ‘1’ as a numeral, UNC *haṭ*, ± human
mi haṭ ṭya un-em.
 one UNC boy have-1s
 “I have one son.”
- (33) Colloquial Azerbaijani (*Tabriz dialect*) (Kiral 2001: 146, #76)
 ‘1’ as an indefinite article, UNC *dana*, ± human
tæbæye-ye dovvom-da bi dana otax hazir
 story-LINK second-LOC one UNC room ready
el-il-læx bu-lar-a.
 make-DUR-PL s/he-PL-DAT
 “They prepare a room for them on the second floor.”
- (34) Colloquial Azerbaijani (*Tabriz dialect*)
 (Kiral 2001: 144, #53) (same as Example (11))
 ‘1’ as a numeral, UNC *dana*, ± human
tæk=o tænha bi dana otax-da zendegi el-ir-Ø.
 sole=and alone one UNC room-LOC life make-DUR-3S
 “She lives all alone in one room.”
- (35) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 138, #12)
 ‘1’ as an indefinite article, with UNC *dánæ*, non-human
yey ru vérg-i yey danæ ogéj=es b-árdæ=ve.
 one day wolf-OM one UNC sheep=3S:OBL PU-bring:PAST=AUX:PAST
 “One day the wolf had brought (him) a sheep.”
- (36) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 34, #27)
 ‘1’ as a numeral, UNC *dánæ*, ± human
yey dánæ zene n-ðw-æ.
 one UNC woman NEGv-be:SBjV-3S
 “There should not be (even) one woman (left here).”
- (37) Masulei (Lazard 1979: 44)
 ‘1’ as an indefinite article, UNC *gələ* (> *i-llæ*), ± human
i-llæ hekayæt əm-on=ra bə-va.
 one-UNC tale s/he-OP=for PU-say
 “Tell them a tale.”

- (38) Masulei (Lazard 1979: 44)
 ‘1’ as a numeral, UNC *gəɫæ* (> *i-llæ*), ± human
i-llæ bəro-ær pæs æz i sal doktori
 one-UNC brother:-OSII after from one year doctoring
dær-mot=æ.
 PV-learn:PAST=COP:3S
 “After one year one brother (of the three) learned medicine.”
- (39) Koluri (*Tatic, Central Tati, NW Iranian*) (Yarshater 1959: 65)
 ‘1’ (*i*) as an indefinite article, UNC *gəɫæ* (> *i-llæ*), inanimate
i-llæ xas-æ dæsmal-e=ku dæ-váess=eš=e.
 one-UNC fine-LINK kerchief-OBL=LOC PV-tie:PAST=3S:OBL=COP:3S
 “He wrapped it up in a fine kerchief.”

In Muslim Caucasian Tat, ‘1’ as a numeral requires a universal numeral classifier, but a classifier usually does not appear with ‘1’ as an indefinite article:

- (40) Muslim Caucasian Tat, Daykušču dialect (Grjunberg 1963: 147, #639)
 indefinite article, no UNC
bo-w suyrá yæ xakærglæ hist-Ø.
 in-that bowl one egg exist-3s
 “There’s an egg in that bowl.”

‘1’ as an indefinite article, HNC for human nouns: The use of a human numeral classifier is also permissible with indefinite articles in the languages of the area as well:

- (41) Vafsi (Stilo 2004: 68, #20)
 ‘1’ as an indefinite article, HNC *næfær*, human
yey næfær dózd-i=æm eqbál-i=s æt-árd-e
 one HNC thief-INDEF=FOC fate-OM=3S:OBL DUR-bring:PAST-PP
biæban=dæ (æ)r-gerda-Ø.
 desert=in DUR-wander:PAST-3S
 “As fate would have it, a thief was walking around in the field.”
- (42) Masulei (Lazard 1979: 40)
 ‘1’ as an indefinite article, HNC *næfær*, human
se næfey=enæ, i næfær næjar, i næfær xəyat, i
 three friend=COP:3P one HNC carpenter one HNC tailor one
næfær molla
 HNC mullah
 “There are three friends: a carpenter, a tailor, and a mullah”

As noted, indefinite articles are indistinguishable from the numeral ‘1’. Clear examples of the latter pattern are, in fact, quite rare in the corpora consulted. There is also an additional complication with the classifier *næfær*: the indefinite article ‘1’ usually combines with the human numeral classifier as a lexified indefinite pronoun in most of these languages, e.g., Vafsi *yey næfær*, Armenian *mi hok’i*, and the equivalents in other languages, mean “someone, somebody”. There is no parallel pronoun formation for the universal numeral classifier.

3.3.2 *Different numeral classifiers for ‘1’ vs. ‘higher than 1’*

As far as we know, only Colloquial Persian and Muslim Caucasian Tat have two UNCs in complementary distribution, one used with the numeral ‘1’ (*dune* and *dæné*, respectively) and *ta* (Persian) ~ *ta/tæ* (Caucasian Tat) with numerals above ‘1’ and the quantifier *čænd* (see §3.4).

- (43) Muslim Caucasian Tat, Daykušču dialect (Gryunberg 1963: 124, #26)
yé dæné šillé mi-zæn-úm bæ tü...

one UNC slap FUT-hit-1s to you

“I’ll give you a slap ...”

(no examples of ‘1’ as a numeral encountered in corpus; for numerals over ‘1’, UNC *ta* ~ *tæ*, see (24) and (25);

- (44) Colloquial Persian

‘1’ as a numeral, UNC *dune*, ± human

ye dune bærädær dar-æm=o do=ta xahær.

one UNC brother have-1s=and two=UNC sister

“I have one brother and two sisters.”

3.3.3 *Different frequencies for ‘1’ vs. ‘higher than 1’*

As previously mentioned, tokens of ‘1’ as a numeral are rare in the texts consulted, and I can only make tentative statements about the differences in frequencies of the former vs. ‘1’ as an indefinite article. In the discussion below, the numeral ‘1’ is first contrasted with numerals over ‘1’; ‘1’ as a numeral vs. indefinite article is discussed separately after that.

Table 1 shows the frequencies of the two main classifier types with human, animal (where available), and inanimate nouns in the three languages prominently discussed in this article. Since this study is corpus-based, no elicitation was done for the cells with 0 tokens in the table. In the percentage rows in the table, every horizontal block of two cells for inanimates and animals adds up to 100%, as does every block of three cells for human nouns.

Table 1. Frequencies of numeral classifiers in Azerbaijani, Armenian, and Vafsi

Tokens	Tabriz Azerbaijani					Colloquial Armenian						Vafsi							
	Inanim.		Human			Inanim.		Animal		Human		Inanim.		Animal		Human			
	150					211						348							
	UNC	Ø	UNC	HNC	Ø	UNC	Ø	UNC	Ø	UNC	HNC	Ø	UNC	Ø	UNC	Ø	UNC	HNC	Ø
'1' num	4	10	2	0	0	0	10	–	–	8	0	9	–	–	–	–	2	0	1
%	28.6	71.4	too few tokens			0	100	–	–	47.1	0	52.9	–	–	–	–	too few tokens		
'1' art.	83	33	28	2	6	127	28	24	2	118	4	62	26	135	24	11	16	5	65
%	71.6	28.4	77.8	5.6	16.6	81.9	18.1	92.3	7.7	64.1	2.2	33.7	16.1	83.9	68.6	31.4	18.6	5.8	75.6
2/2 +	16	4	10	2	0	11	35	0	3	24	26	54	8	21	4	2	7	2	19
%	80.0	20.0	83.3	16.7	0	25.9	76.1	too few		23.1	25	51.9	27.6	72.4	66.7	33.3	25.0	7.1	67.9

Numerals: Table 1 shows the low frequency of use of classifiers with inanimate nouns for the numeral '1' in Azerbaijani (28.6%) and even lower in Armenian at 0%. There are no tokens in the Vafsi corpus to be able to address this point. There is a slight increase in classifier use for inanimates with numerals over '1' in Armenian at 25.9%. The Vafsi figures at 27.6% for this category are very similar. Azerbaijani, however, shows a huge jump to 80% in this case. Thus, except for the last figure (Azerbaijani), there is a general preference, but not a requirement, for the non-use of classifiers for inanimate nouns with numerals in Armenian and Vafsi.

There are too few tokens of classifiers with '1' as a numeral for human nouns in Azeri and Vafsi to establish frequency, but with numerals over '1' the Azeri figures rise substantially from 80% for inanimates to 100% (83.3% UNC, 16.7% HNC) for humans. Use of Armenian classifiers with numerals above '1' for human nouns starts out low for inanimates (25.9%) and almost doubles to 48.1% (23.1% UNC, 25.9% HNC) for humans. Vafsi inanimates also start out low at 27.6% for numerals over '1' and only rise to 32.1% (25% UNC, 7.1% HNC) for human nouns.

Indefinite articles: In contrast to the corresponding figures for '1' as a numeral, inanimates in Azerbaijani and Armenian both show a high use of classifiers with '1' as an indefinite article – numeral: 28.6 > article: 71.6% and 0% > 81.9%, respectively. Conversely to these languages, Vafsi has a very low frequency of classifiers with '1' as an indefinite article (16.1%). Comparing these three languages for indefinite articles with human nouns, we find the reverse progression, i.e., a decrease of classifier usage from Azeri 83.4% (77.8% UNC, 5.6% HNC), to Armenian 66.3% (64.1% UNC, 2.2% HNC), to Vafsi 24.4% (18.6% UNC, 5.8% HNC). Vafsi is low in classifier frequency for both inanimate and human nouns with '1' as an indefinite article. Strikingly, however, the situation for NON-HUMAN ANIMATES (animals) in Vafsi is quite the reverse of both inanimate and human nouns, as discussed next.

Numeral classifiers and the Vafsi animacy hierarchy: Various domains in Vafsi grammar lead to establishing the following animacy hierarchy (only a subset is given here): human > animals > trees/large plants > inanimates. I had already determined that this hierarchy affected object marking, adjuncts of adpositions, possessor constructions, and verb agreement. Table 1 now leads me to add the frequency of the classifier *dánæ* with animals vs. both humans and inanimates to this list. As Table 1 shows, with '1' as an indefinite article, inanimate and human nouns overwhelmingly favor the NULL MARKING of numeral classifiers – 83.9% and 75.6%, respectively. For animals, however, the reverse situation is true: Ø marking of the classifiers with '1' as article falls to 31.4%.⁸ A parallel contrast in classifier

8. A similar situation holds for Armenian. Further investigation into numerals over '1' with animals is needed.

marking is also found for numerals higher than ‘1’ with animals (33.3%) vs. inanimate (72.4%) and human (67.9%) nouns. One point of interest here, to my mind, is that the two opposite poles of the animacy hierarchy are in agreement but contrast with an intermediate category (animals).

3.4 Quantifiers and numeral classifiers

Most languages in this area have two types of quantifier: (1) a main group (‘many’, ‘some’, ‘all’, ‘each’, ‘every’, etc.), that never take classifiers and, depending on the quantifier and language, usually require a plural head noun; and (2) a group with one quantifier (‘a few/several’) taking optional classifiers with head nouns usually in the singular. Only the latter type is of interest here. Since this quantifier means both ‘a few/several’ and ‘how many?’ and, since ‘how many (things)?’ mirrors the answer (‘numeral-classifier-HEAD’), this word usually exhibits a similar behavior to that of numerals and may thus take a classifier in most of these languages:

- (45) Colloquial Persian (UNC, always obligatory in purely colloquial registers)

čæn(d)=ta moælleṃ dar-i?
 how.many=UNC teacher have-2s
 “How many teachers do you have?”

- (46) Colloquial Persian

(ye) čæn(d)=ta ketab dar-æṃ bæra=t.
 one how.many=UNC book have-1s for=2s:POSS
 “I have a few books for you.”

While there is a connection between numerals and the quantifier/WH-word, there also is a sharp distinction in the use of classifiers with these two categories. In some languages, as we just saw in Persian (45)–(46) above, the classifier is much more obligatory with the quantifier. In other languages, such as Vafsi and Armenian, it is not obligatory with the quantifier. The corpus for Azerbaijani does not provide enough information to make a full judgment for this category, although my hunch is that it is obligatory with the quantifier.

- (47) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 138, #16)

WH-word *čænd*, no numeral classifier
čænd ogæj=es ær-go?
 how.many sheep=3s:OBL DUR-want
 “How many sheep are needed (lit: does it want)?”

- (48) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 76, #1)
 Quantifier *čænd*, no numeral classifier
v-is-di ke bəle čænd esfəhán-i
 PU-3S:OBL-see:PAST SUBR yes how.many Esfahani-OM
lutí-e æt-á-nde.
 wise.guy-PL DUR-come-3P
 “He saw, yeah, a couple of wise guys of Esfahan coming.”
- (49) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 36, #50)
 Quantifier *čænd* + UNC *danæ*
čænd danæ dozd=e in šæ:r=dæ.
 how.many UNC thief=COP:3S this city=LOC
 “There are a few thieves in this country.”
- (50) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 78, #28)
 Quantifier *čænd* + HNC *nəfær*
čæn nəfær moan dir-am.
 how.many UNC guest have-1P
 “We have a few guests.”
- (51) Colloquial Azerbaijani (*Tabriz dialect*) (Kıral 2001: 142, #13)
 Quantifier *nečæ* + UNC *dana*
jəlal nečæ dana=da uşay-i var=idi.
 Jalal how.many UNC=also child-3S:POSS EXIST=COP:PAST
 “And Jalal had a few children.”
- (52) Koluri (*Central Tati*) (Yarshater 1959: 66)
 Quantifier *čæn* + UNC *gələ*
čæn gələ pilə merdək-an æga=ku neštæ b-imbe.
 how.many UNC big man-PL there=in seated be:PAST-3P
 “Several great men were sitting there.”

Table 2 shows that spoken Armenian tends not to use a classifier with the WH-word usage of *k’ani* (26 tokens), but a minority of tokens (6) in the corpus indicates that it is a permissible pattern. With the non-WH meaning ‘a few’, there is also the additional possibility for an optional indefinite article *mi* ‘1’ to precede the quantifier – as in most languages of the area, e.g., Persian example (46) above. The table

Table 2. The Armenian quantifier as “how many?” and “a few”

	“How many?”	“A few”	
	<i>k’ani?</i>	<i>k’ani</i>	<i>mi k’ani</i>
+ UNC + head	6	2	1
+ Ø + head	26	7	23

also shows that Armenian even has a preference (24 tokens vs. 9) for the use of the indefinite article with the non-WH sense of this quantifier. Examples follow below.

- (53) Colloquial Armenian (eanc: OPD A-1)
 WH-word *k'ani*, no numeral classifier + head (\pm human)
k'ani yerk'=es nerḡayac'n-elu?
 how.many song=2s present-FUT.PTCP
 "How many songs are going to perform?"
- (54) Colloquial Armenian (eanc: OSD 04–3)
 WH-word *k'ani* + UNC *haṡ* + head (\pm human)
k'ani haṡ diplom un-es?
 how.many UNC diploma have-2s
 "How many diplomas do you have?"
- (55) Colloquial Armenian. (eanc: OSD 9–5)
 Quantifier *mi k'ani* + UNC *haṡ* + head (\pm human)
mi k'ani haṡ=el axč'ik=a heṡ uyark-el.
 one how.many UNC=FOC girl=3s back send-PP
 "He has sent a few girls back"
- (56) Colloquial Armenian (eanc: Polylogue 026)
 Quantifier *k'ani* + UNC *hok'i* + head (\pm human)
k'ani hok'i mart' meṛn-um=a.
 how.many HNC man die:PRS=3S
 "A few men die."

3.5 Obligatory use of numeral classifiers as heads of NPs

If the head noun of a numeral phrase is deleted, i.e., in an "anaphoric construction of Q-CL without overt expression of the noun" (Greenberg 1990: 186), the classifier almost universally in this area (even if optional in the full NP) obligatorily becomes the head of the NP. Thus, the classifier is obligatory when items are counted but not overtly mentioned. (See also (29): "which one?"). In (57), the use of *čar* alone by Speaker B would be ungrammatical:

- (57) Colloquial Persian
 A: *čæn(d)=ta šo:be dar-æn(d)?*
 how.many=UNC branch have-3P
 B: *čar=ta.*
 four=UNC
 A: "How many branch offices do they have?" B: "Four."

Counting out items creates a similar situation for obligatory classifiers. Suppose, for example, that someone whose peach tree has borne fruit starts counting the fruit by quickly scanning:

- (58) Colloquial Persian
bé-bin-æm, do=ta, čar=ta, hæft=ta, dæh=ta, dævazdæh=ta...
 SBJV-see-1s two=UNC four=UNC seven=UNC ten=UNC twelve=UNC
 “Let me see, two, four, seven, ten, twelve ...”
 (N.B. *do, *čar, *hæft, *dæh, *dævazdæh are ungrammatical here.)

To add to the tendency to treat ‘1’ differently from other numerals in Colloquial Persian and Muslim Caucasian Tat (see § 3.3.2), the number ‘1’ as the head of an NP may alternate (although not in all contexts) between the usual ‘1’ + universal numeral classifier *dune* and a special form of ‘1’ (*yéki*) with no classifier: *ye dune ~ yéki, do=ta, čar=ta, hæft=ta*, etc.

- (59) Colloquial Persian
čænd=ta mí-xa-y? yéki bæs=e?
 how.many=UNC DUR-want-2s one:HEAD enough=COP:3s
 “How many do you want? Is one enough?”

- (60) Colloquial Persian
yéki=šun xærab bud-Ø, endáxt-æm dur.
 one:HEAD=3P:OBL ruined be:PAST-3s throw:PAST-1s far
 “One of them was rotten, (so) I threw (it) out.”

- (61) Muslim Caucasian Tat, Daykušču dialect (Grjunberg 1963: 129–130, 178)
yéki bæ xarí ništ-Ø.
 one:HEAD LOC ground sit:PAST-3s
 “One of them sat down on the ground.”

When a classifier after numerals or the quantifier/WH-word ‘how many, a few’ serves as head of an NP and includes the referent noun (62)–(65) or pronoun (66) in a partitive-type sense (‘one/two/a few OF HIS FRIENDS’; ‘one/two/a few OF THEM/US/YOU’), two patterns are used:

A. The referent NOUN/PRONOUN is encoded in an ablative-marked partitive construction:

- (62) Colloquial Azerbaijani (*Tabriz dialect*) (Kıral 2001: 166. #14)
geysær bi dana [yaxči irani film-lær-dæn]=di-Ø.
 Geysar one UNC good Iranian film-PL-ABL=COP-3s
 “Geysar (Caesar) is one [of the best Iranian films].”

- (63) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 70, #39)
hâr=es=da yey dânae [æz in æmame-ye
 PV=3S:OBL=give:PAST one UNC from this turban-LINK
isbi]=s bæ-sátte.
 white=3S:OBL PU-build:PRF
 “He has made one [of those white turbans].”
- (64) Colloquial Persian
čæn(d)=ta [æz moælle-m-a] inglisi mî-dun-æn(d)?
 how.many=UNC from teacher-PL English DUR-know-3P
 “How many [of the teachers] know English?”
- (65) Colloquial Armenian (eanc: OSD Polylogue 35)
in̄tr-el=a [eʔ yer-kʻ-er-icʻ] yerekʻ haʔ vorpeszi k̄runk
 choose-PP1=3S that song-PL-ABL three UNC so.that crane
p̄araʔon-i=n Ø-yerk-i. ʻ
 festival-DAT=DEF SBJV-sing-3s.
 “He chose three [of those songs] in order to sing Krunk (“the Crane”) at the festival.”
- (66) Colloquial Armenian (eanc: OSD Polylogue 123)
[nrancʻ-icʻ] mi hokʻi tʻoʔ Ø-ga-Ø.
 s/he:PL-ABL one HNC let SBJV-come-3s
 “Let one [of them] come.”
- B. A referent PRONOUN may optionally be indexed by a possessive enclitic or an oblique enclitic in its possessive sense. (See also (60) and compare the full form (66) vs. the short form pronoun (67).)
- (67) Colloquial Armenian (eanc: OSD Polylogue 55)
yerku haʔ=[i] baraḳ=a, mnacʻaç=i haʔ=a.
 two UNC=3S:POSS thin=COP:3s rest=3S:POSS thick-COP:3s
 Two [of them] are thin, the rest are thick.”
- (68) Colloquial Azerbaijani (*Tabriz dialect*) (Kıral 2001: 144, #40)
bir-[i]=dæ mun-un ev-i-n
 one-3S:POSS=also s/he-GEN house-3S:POSS-ACC
ist-ir-di-Ø al-a-Ø.
 want-DUR-PAST-3S buy-SBJV-3S
 “One [of them] wanted to buy his house.”
- (69) Colloquial Persian
čæn(d)=ta=[šun] hæmuz jævab né-dad-æn(d).
 a.few=UNC=3P:POSS yet answer NEGV-give:PAST-3P
 “A few [of them] haven’t answered yet.”

- (70) Vafsi (*Tatic, Tati, NW Iranian*) (Stilo 2004: 78, #20)
yéki=[s] æt-ar-e ke=s=dæ (æ)d-ærz-e.
 one:HEAD=3S:OBL DUR-bring-3s house-3S:OBL=in DUR-put-3s
 “He brings [one of them] and puts it in the house.”

When the numeral classifier serves as the head of an NP, it may then assume any NP role including the case marking or case-like enclitics for that role in the given language:

- (71) Colloquial Armenian (eanc: OSD Polylogue 105)
mi haṭ-i=n harc‘r-ink‘ as-ec‘-Ø č‘-i ġin-um.
 one UNC=DAT-3S:POSS ask:PAST-1P say-PAST-3S NEGV-3S go-DUR
 “We asked one of them; he said he wouldn’t go.”
- (72) Colloquial Persian
ná-tunest-æm un do=ta=ro bé-foruš-æm.
 NEGV-can:PAST-1s that two=UNC=ACC SBJV-sell-1s
 “I wasn’t able to sell those two.”

4. Diachrony

For most of the languages discussed above, little or nothing is known of the diachrony, origins, or evolution of numeral classifiers. Significantly more, however, can be said about Persian. The usage, as well as the form, of at least *ta* as a numeral classifier goes back to Middle Persian. Mo‘in (1963: 451) notes that *ta* comes from Middle Persian, for which MacKenzie (1971: 81) lists *tāg* “item, unit; alone, single”. Abrahamyan (1965: 244), who uses the more conservative transcription *tāk*, not only lists the English translation as “a unit”, but – interestingly for our study of classifiers – also gives both the Modern Persian translation as *dāne*, *tā*, as well as the Armenian as *mi haṭ*, the classifier forms cited throughout this article for these languages. Brunner (1977: 45–46) says the following for Middle Persian: “A counting word sometimes accompanies the numeral; the substantive which is dependent on this phrase MAY [emphasis mine] occur in the plural”:

- (73) Middle Persian (Brunner 1977: 46)
*tansrd‘al-‘n ḥpt GBR‘ W PRŠT-‘n ḥm mr.**
 officer-PL seven HNC:man and cavalry-PL same number
 “seven officers and the same number of cavalry”
 (*Transliteration is traditional for Middle Persian including Aramaic ideograms in capitals.)

Brunner (1977: 45) notes that inanimate nouns after numerals are singular (see also §5, points 1 and 2 below regarding Mazanderani), whereas animate nouns may be singular or plural; for the use of classifiers, however, he only provides one example, (73). While Rastorgueva (1966) gives examples of numerals and head nouns, no mention is made of classifiers. The following is one of two examples of the use of this classifier Mache provides for Middle Persian:

- (74) Middle Persian (Mache 2012: 171)
čand tā(k) dānāg-ān ī hindūg-ān
 a.few UNC wiseman-PL LINK Indian-PL
 “some Indian wise men”

While no traces of classifiers are found in any stages of Iranian BEFORE Middle Persian, i.e., Old Persian or Avestan,⁹ Mache’s work shows us clearly that classifiers SUBSEQUENT to Middle Persian became notably more robust, at least in class membership, in all stages of Persian, beginning with many examples from the 11th century onwards. Her work points out in very convincing terms how integral these classifiers had become, and are now, in Persian – at least in the written language.

There are still three unresolved issues with these points on diachrony:

1. We have no frequency counts of classifiers from corpora of either Middle Persian or Classical Persian to trace classifier robustness from there to modern Colloquial Persian. Since Brunner (1977: 45–47) shows only one use of a classifier out of 22 examples of numerals with head nouns, we suspect that their use in Middle Persian was optional, as they are in the modern language. They were clearly optional in Classical Persian, as well.
2. It is impossible to determine from historical corpora whether written Middle Persian or Classical Persian also reflect the reality of their spoken registers. Perhaps there was, as with Modern Persian, a smaller inventory in the spoken language of the time and the full classifier system, especially of Classical Persian, was only used by a small literary elite.
3. How did classifiers emerge in Iranian as a whole, and specifically *tāg* in Middle Persian? Mache (2012: 178–190) offers some hypotheses, but at present they are still conjectural. Gil (2005: 227) notes, “The main concentration of numeral classifiers is in a single zone centered in East and South-East Asia, but reaching out both westwards and eastwards ... [and NCs] pick up again, albeit in

9. While Old Avestan, the oldest known form of Iranian, is generally classified as an early form of Eastern Iranian, it can for all intents and purposes also be considered virtually identical to Proto-Iranian (D. Neil MacKenzie, p.c.).

optional usage, in parts of western Asia centring on Iran and Turkey; it is not clear whether this should be considered as a continuation of the same large though interrupted isogloss, or as a separate one.” Given the areal bent to my own work, I think the most promising route may be that of a larger isogloss, but it is far too early to make any secure proposal in this vein. If we can find a substantial number of other interesting isoglosses with this distribution, we might be on slightly firmer ground.

Mache does not mention two other related glosses of the classifier *ta* relevant to the evolution and diachrony of classifiers in Iranian: (1) “single as opposed to a pair”, i.e., one of a pair; and (2) “half of a load”, i.e., one side of a double saddlebag on both sides of a pack animal (Mo’ in 1963: 451). These senses parallel the Georgian classifier *chali* (see §6) also glossed “one of a pair” (Harrell et al. 2002: 801). This correlation may provide a clue to the origins of classifiers in this area.

5. Greenbergian typology

As mentioned in §1, Greenberg (1990: 177) discusses the correlation between two patterns, the use of classifiers and number in the head noun: “Numeral classifier languages generally do not have compulsory expression of nominal plurality, but at most facultative expression.” There are four types of patterns involving numeral classifiers in the Araxes-Iran zone:

1. A pattern that requires classifiers between numeral and head noun and does not license a plural form of the head noun: Although it was stated in § 1 that definite nouns may be pluralized in many Araxes-Iran languages, Mazanderani does not allow this pattern with inanimate nouns. Plural forms of inanimates (**me se=tta kitab-un* > *me se=tta kitab* “my three books”) were rejected by a native speaker in elicitation (but see next point).
2. A pattern that requires classifiers between numeral and head noun and do license a plural form of the head noun: In Mazanderani this pattern only occurs with animate head nouns. The Mazanderani typology mentioned in points 1 and 2 here seems to be the same as noted for Middle Persian by Brunner (see §4).
3. A pattern that has optional classifiers between numeral and head noun and licenses a plural form of the head noun: MOST LANGUAGES in the area allow an optional plural with a definite head noun after a numeral + classifier, the latter also optional. Armenian, Vafsi, and Gilaki may also license a plural form of non-definite head nouns after numerals.

4. A pattern that has optional classifiers between numeral and head noun and requires a plural form of BOTH the classifier AND the head noun: This pattern exists in Jewish Urmi Neo-Aramaic (Khan 2008) and Jewish Sanandaj Neo-Aramaic (Khan 2009) of Iran. Since this type has not yet been discussed in this article, nor have I found it in Greenberg's typology, examples of singular (75) and plural (76) types are given here for Urmi where classifiers are only a minor pattern (about 33% in my corpus):

(75) Jewish Urmi Neo-Aramaic (Khan 2008: 404)¹⁰

... + *əl-amart-í* *gal=xá=danka* *tapúk*
 ...OBJMK-palace(F)-1s:POSS with=one-UNC:SG kick:SG
ma-pərx-át-ta *b=šəmm-e?*
 CAUS-fly-2SM-3SF:OBL to=sky-PL

“(How can you) make my palace fly in the air with one kick?”

(76) Jewish Urmi Neo-Aramaic (Khan 2008: 430)

g=urmí *tré=dank-e* *knəšy-è* *it-wá.*
 in=Urmia two=UNC-PL synagogue-PL exist-PAST

“In Urmi there were two synagogues.”

In my ongoing *Atlas* work, other patterns may be discovered. Arabic dialects of Kurdistan may exhibit additional typologies, e.g., Greenberg's (1990: 178) Omani Arabic type, which has pluralized classifiers with numerals above '3' and a head noun in the singular.

6. Areality and fade-out

Recently I have been fascinated by the phenomenon of fade-out of various types of linguistic features at the periphery of any given isogloss and have been working on quantifying this fade-out along a geographic cline (see Stilo 2012). Numeral classifiers are another interesting example of such a fade-out process. That is, while classifiers are very robust at the 'epicenter' of this isogloss in north and central Iran, as we progress away from this area, the numeral classifiers fade out in three ways: frequency, class membership, and/or domain.

Some Araxes-Iran languages on the western and northern periphery of the numeral classifier isogloss show fade-out effects in frequency and class membership. With regard to frequency, while some languages such as Lorri Armenian (Shnogh subdialect along the border with Georgia) and Jewish Sanandaj Aramaic (Khan

10. In Neo-Aramaic both direct and oblique PAMs encode gender in the singular, hence the interlinears 2SM, 2SF:OBL.

2009) have optional classifiers, Table 3 shows that their actual frequency is quite low and the majority of nouns with numerals occur without classifiers. Gawrajui shows an approximate 50/50 split. Class membership of the numeral classifier category is also reduced since the corpora for these three languages show no human numeral classifier for human nouns, only the one universal numeral classifier.

Table 3. Three languages with low to medium frequency of numeral classifiers

	Gawrajui					Lorri Armenian					J. Sanandaj Aramaic				
Total tokens	23					19					22				
Split	12 <i>gilæ</i> /11 Ø					4 <i>haṭ</i> /15 Ø					8 <i>dana~danka</i> /14 Ø				
	Inanim.		Human			Inanim.		Human			Inanim.		Human		
	UNC	Ø	UNC	HNC	Ø	UNC	Ø	UNC	HNC	Ø	UNC	Ø	UNC	HNC	Ø
'1' num.	–	–	–	–	2	1	5	–	–	–	–	–	–	–	4
'1' article	3	–	2	–	–	2	1	–	–	2	7	–	–	–	2
2 or more	2	4	5	–	5	–	7	1	–	–	1	5	–	–	3
Source	Mahmoudweysi et al. (2012)					Mkrтч'yan et al. (1977)					Khan (2009)				

I show the fade-out of classifiers in two directions with Tables 1 and 3. First, as we progress westward from obligatory classifier use in Caspian languages to the robust but optional use in Persian and Vafsi (see Table 1) and from there on to the diminished use of classifiers in Gawrajui (50%) and Jewish Sanandaj Aramaic ($\pm 33\%$) on Table 3, we eventually come to the area in the west where there is no classifier use: Central Kurdish, other forms of Aramaic, and possibly Arabic. Kurmanji¹¹ also has extremely marginal use of two classifiers.

The second areal fade-out is the northwesterly cline starting with Caspian languages as the epicenter where classifiers are obligatory throughout. From there we come to Azerbaijani and Armenian where classifiers are robust but highly optional, especially in specific domains, differing according to language. At the next to last

11. Chyet (2003) has two numeral classifiers for Kurmanji: “*Heb* ‘(1) ... seed, kernel, grain; (2) small quantity, small amount ...; (3) general counting word (for things and people)’” (2003: 268) and “*lib* ‘(1) ... a single grain or seed or seed ...; (2) counting word for fruits, nuts, etc.: *bîst-sih lib gwîz* ... ‘20–30 walnuts’” (2003: 355). Since MacKenzie (1961) makes no mention of these words or any type of classifier, their usage may be as marginal in Kurmanji as they are in Georgian (next paragraph). More work is needed to determine the robustness (in frequency and domain) and dialectal distribution of these classifiers. Geoff Haig (p.c.) reports that he has not found these classifiers in texts from Amadiya or Djabal Sinjar of Iraq, or in his Erzurum (Turkey) corpus, where *tane*, a borrowing from Turkish, sometimes occurs. While MacKenzie does not mention any such patterns in Central Kurdish of Iraq and Iran either, similar marginal classifier usage may show up in further corpus investigation.

point to the northwest, but still within the classifier zone and very peripheral to it, we find Georgian, which has not yet been discussed. That is, Georgian¹² seems to be the last step before we reach the area where there are no vestiges of sortal classifiers. The element of interest in Georgian is the classifier,¹³ *c^hali*. In the following example we see that this word functions both as classifier between numeral and head noun (Speaker A) as well as head of an NP when the referent noun is deleted (Speaker B):

(77) Colloquial Tbilisi Georgian (Lia Shartava, personal communication)

A: *ramdeni c^hali p^hank^hari g-neb-av-t^h?*
 how.many UNC pencil 2:OBL-would.like-THEME-PL

B: *xut^hi c^hali, t^hu šeiʒleba*
 five UNC if it.is.possible
 “How many pencils would you like?” “Five, please.”

The element *c^hali*, however, is highly reduced in Colloquial Georgian in all three categories mentioned above for fade-out phenomena: frequency, class membership, and domain:

– Frequency

While examples of its use were provided by a native speaker, there is not even one token of it in my corpus of about 6–7 hours of spoken Georgian;

– Class membership

Georgian only has this one element functioning as a true classifier;

– Domain

c^hali only occurs with inanimates, never with animals or animates; even many types of inanimates do not permit it. It probably also occurs only in very restricted situations, e.g., shopping or other types of parceling out.

I have not investigated the occurrence of classifiers progressing eastwards from central Iran toward Afghanistan and Central Asia or toward southeast Iran and the Indian subcontinent.

12. Georgian is fully included, however, within the isogloss of singular form of the head noun after numerals.

13. I thank Alice Harris for pointing out this word and its numeral classifier usage in Georgian and Lia Shartava for providing me with multiple examples as well as grammaticality judgments on its occurrence.

7. Conclusion

In this article I have first discussed the similarities and differences in sortal numeral classifier usage in certain languages in the Araxes-Iran zone and, on the other hand, have viewed the phenomenon of classifiers from an areal perspective. I have shown that languages of very different origins have eventually acquired the same type of small-inventory classifier system, mostly with the same or very similar behaviors. None of these languages would have had them in their proto-languages. In the case of Persian and Armenian, for example, Proto-Indo-European is certainly not thought to have had any sortal classifier pattern, and as we saw in § 4 above, even the oldest stages of Iranian show no traces of them. Local Turkic languages and relevant Aramaic dialects of the area borrowed the forms and their usage much later from Iranian languages – in the case of Turkic, probably not from Persian but from the indigenous Iranian languages of Azerbaijan by incoming Oğuz Turkic tribes in the 11th century, who then introduced them into Anatolia (see Stilo 2014, 2016).

As mentioned earlier, the corpora for this study are small, but they help determine the classifier patterns of the languages under discussion. Detailed, in-depth examination of various linguistic categories in these small corpora can bring up new topics and unexpected conclusions. Such work often reveals new and surprising facts about these underinvestigated languages, as we saw in §3.3.3 with the issues of numeral classifiers and the animacy hierarchy in Vafsi and with a typological pattern of plurality for both classifiers and head nouns demonstrated in §5 that has not been noted by Greenberg.

It is my hope that scholars working on numeral classifier typology will find areas of intriguing commonalities and interesting differences between the classifier systems of the languages investigated in this article and those of languages of other parts of the world.

Acknowledgments

I would like to thank Avazeh Mache and one anonymous reviewer for their helpful comments and suggestions on this article.

References

- Abrahamyan, Rouben. 1965. *Pahlaveren-parskeren-hayeren-řuseren-angleren bařaran*. Yerevan: Mitk'.
- Borjian, Habib & Maryam Borjian. 2008. "The Last Galesh Herdsman: Ethno-linguistic materials from South Caspian rainforests". *Iranian Studies* 41: 3. 365–402.
doi:10.1080/00210860801981336

- Brunner, Christopher J. 1977. *A Syntax of Western Middle Iranian*. Delmar, NY: Caravan Books.
- Bulut, Christiane. 2006. *On Culture and Language in the Turkic Region of Bayâdistân*. Habilitationsschrift, Johannes Gutenberg-Universität Mainz.
- Chyet, Michael L. 2003. *Kurdish-English Dictionary*. New Haven: Yale University Press.
- Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9781139164344
- ALMG. eanc = Eastern Armenian National Corpus. Ongoing. <http://www.eanc.net/>. (Accessed: 25 April 2017.)
- Gebhardt, Lewis. 2009. *Numeral Classifiers and the Structure of the DP*. Ph.D. dissertation, University of Illinois.
- Gil, David. 2005. "Numeral Classifiers". *The World Atlas of Linguistic Structures* ed. by Martin Haspelmath, Matthew Dryer, David Gil & Bernard Comrie. 226–229. Oxford: Oxford University Press, 226–229.
- Greenberg, Joseph. 1990. "Numeral Classifiers and Substantival Number: Problems in the genesis of a linguistic type" *On Language: Selected writings of Joseph H. Greenberg* ed. by Keith Denning & Suzanne Kemmer. Stanford: Stanford University Press, 166–193.
- Grjunberg, Alexandr L. 1963. *Jazyk severoazerbajdžanskix tatov*. Leningrad: Akademija Nauk SSSR.
- Harrell, S. J., Mala Koupounla, Mala Tsitsishvili & Ketevan Gabounla. 2002. *Georgian-English Dictionary*. Hyattsville: Dunwoody Press.
- Khan, Geoffrey. 2008. *The Jewish Neo-Aramaic Dialect of Urmi*. Piscataway: Gorgias.
- Khan, Geoffrey. 2009. *The Jewish Neo-Aramaic Dialect of Sanandaj*. Piscataway: Gorgias.
- Kiral, Filiz. 2001. *Das gesprochene Aserbaidžanisch von Iran*. Wiesbaden: Harrassowitz.
- Lambton Ann K. S. 1953. *Persian Grammar*. Cambridge: Cambridge University Press.
- Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Paris: Klincksieck.
- Lazard, Gilbert. 1979. "Textes en tâleši de Mâsule". *Studia Iranica* 8:133–66.
- Mache, Avazeh. 2012. *Numeral Classifiers in Persian*. Munich: Lincom.
- MacKenzie, Donald Neil. 1961. *Kurdish Dialect Studies*, Vol. I. London: Oxford University Press.
- MacKenzie, Donald Neil. 1971. *A Concise Pahlavi Dictionary*. London: Oxford University Press.
- Mahmoudweysi, Parwin, Denise Bailey, Ludwig Paul & Geoffrey Haig. 2012. *The Gorani Language of Gawraju, a Village of West Iran: Texts, grammar, and lexicon*. Wiesbaden: Reichert.
- Mahootian, Shahrzad. 1997. *Persian*. London: Routledge. doi:10.4324/9780203192887
- Martirosyan, Hrach. 2009. *Etymological Dictionary of the Armenian Inherited Lexicon*. Leiden: Brill.
- Mkrtč'yan, M. et al., eds. 1977. *Hay žolovrdakan hek'iat'ner* [Armenian Folk Tales], Vol. VIII. Yerevan: Haykakan SSH Gitut'yunneri Akademiayi Hratarakch'uty'un.
- Mo'in, Moḥammad, ed. 1963. *Borhân-e qāṭe'*. Vol. I. Tehran: Mo'assese-ye Amir Kabir.
- Rastorgueva, Vera Sergeevna, ed. 1971. *Ġiljanskij jazyk*. Moskva: Nauka.
- Rastorgueva, Vera Sergeevna. 1966. *Sredne-persidskij jazyk*. Moscow: Nauka.
- Redhouse, James W., Sir. 2007. *The Redhouse Turkish-English Dictionary*. Istanbul: Redhouse.
- Schroeder, Christoph. 1999. *The Turkish Nominal Phrase in Spoken Discourse*. Wiesbaden: Harrassowitz.
- Stilo, Donald L. 2004. *Vafsi Folk Tales*. Wiesbaden: Reichert Verlag.
- Stilo, Donald L. 2012. "Intersection Zones, Overlapping Isoglosses, and 'Fade-out/Fade-in' Phenomena in Central Iran". *Iranian Languages and Culture: Essays in honor of Gernot Ludwig Windfuhr* ed. by Behrad Aghaei & M. R. Ghanoonparvar. Costa Mesa: Mazda Publishers, 134–155.

- Stilo, Donald L. 2014. "Further Notes on the Iranian Substratum of Azerbaijani Turkish". *Turkic Language in Iran – Past and Present* ed. by Stein Heidi. Wiesbaden: Harrassowitz Verlag, 271–293.
- Stilo, Donald L. 2016. "On the Non-Persian Iranian Substratum of Azerbaijan". *Turks and Iranians: Interactions in language and history* ed. by Éva Á. Csató, Lars Johanson, András Róna-Tas & Bo Utas. Wiesbaden: Harrassowitz Verlag, 185–234
- Stilo, Donald L. In preparation. *Atlas of the Araxes-Iran Linguistic Area*.
- Thackston, Wheeler M., Jr. 1978. *An Introduction to Persian*. Tehran: Soroush.
- Yarshater, Ehsan. 1959. "The Dialect of Shāhrud (Khalkhāl)". *Bulletin of the School of Oriental and African Studies* 22:52–68. doi:10.1017/S0041977X00076126

The diachrony of Oceanic possessive classifiers

Frantisek Lichtenberk[†]

University of Auckland

With just a small number of exceptions, Oceanic languages have multiple types of attributive possessive constructions that involve a more or less elaborate system of possessive classifiers. Constructions with possessive classifiers usually serve to express alienable possession. A different construction type, one that does not involve possessive classifiers, is typically used to express inalienable possession. Proto-Oceanic had three possessive classifiers. In some present-day languages the original system of classifiers has been expanded considerably, while in some others it has been reduced or eliminated altogether. In some languages possessive classifiers exist alongside numeral classifiers, but the two systems operate on different principles. The paper investigates the system of possessive constructions in Proto-Oceanic, its emergence and its subsequent developments. It also critically assesses the claim that the Oceanic system with different constructions to express alienable and inalienable possession is the result of contact with Papuan (non-Austronesian) languages.

1. Introduction

The Oceanic languages form a subgroup within the Austronesian family. Oceanic languages are spoken in New Guinea, Island Melanesia, Micronesia and Polynesia, but not all of the Austronesian languages in the first three areas are members of the Oceanic group. The position of Oceanic in Austronesian is shown in Figure 1.

The areas where the Oceanic and the non-Oceanic Austronesian languages are spoken are shown in Map 1. Note that both major areas also contain non-Austronesian (Papuan) languages.

A feature characteristic of nearly all Oceanic languages is the existence of more than one type of attributive possessive construction. In what may be considered the typical Oceanic pattern there is a basic binary distinction between **DIRECT** and **INDIRECT** possessive constructions. In a direct construction the possessum noun carries an affix (normally a suffix, but a few languages have prefixes) that indexes

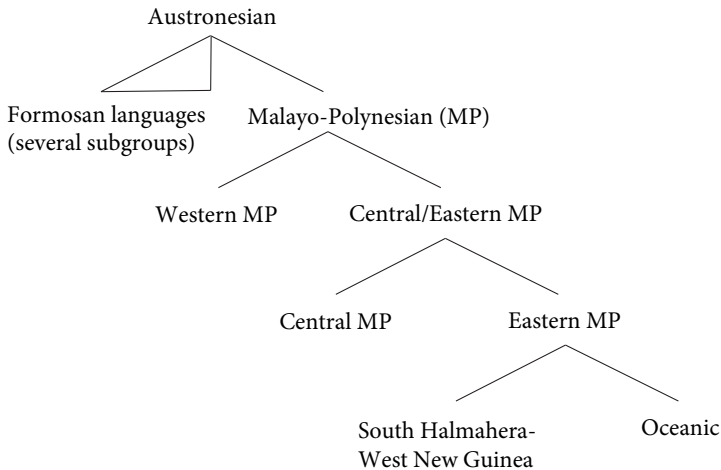


Figure 1. Higher-level subgrouping of Austronesian (after Blust 1983/1984, 2009)

the possessor. Example (1) illustrates a direct construction in Manam (Papua New Guinea):¹

- (1) *tamá-gu*
 father-1SG.POSS
 “my father” (Lichtenberk 1983a: 278)

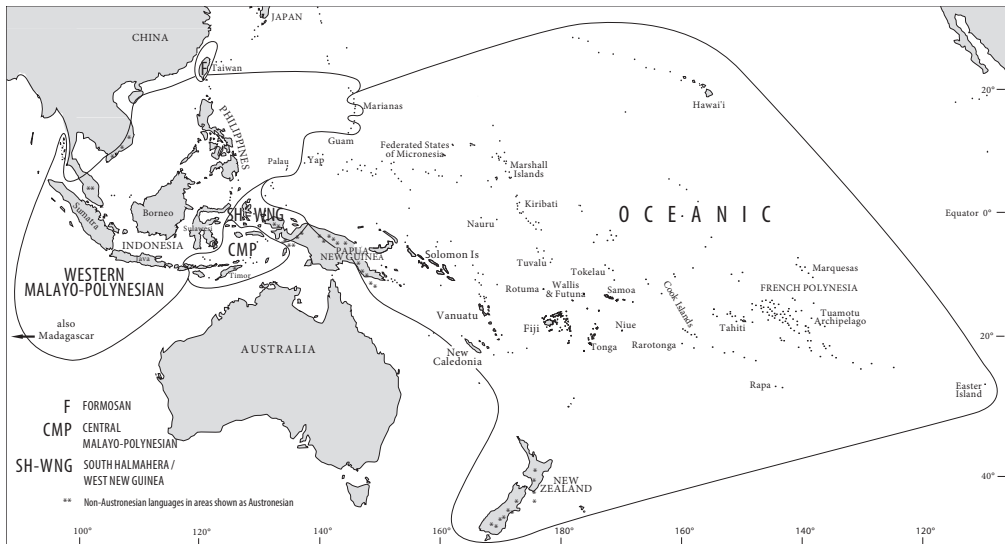
In an indirect possessive construction the affix indexing the possessor is attached not to the possessum noun but to a possessive classifier.² This is illustrated in (2) from Manam:

- (2) *síoti né-gu*
 shirt POSS.CLF(GENERAL)-1SG.POSS
 “my shirt” (Lichtenberk 1983a: 283)

In addition to being indexed on the possessum noun or on a possessive classifier, the possessor may be expressed by a noun phrase. This is illustrated for an indirect possessive construction in (3), also from Manam:

1. In some cases original glosses have been modified for the sake of uniformity.

2. Palmer & Brown (2007) have suggested that in at least some languages the elements called ‘possessive classifiers’ are really nouns. However, I have argued (Lichtenberk 2009a) that there is considerable evidence against such an analysis. The elements in question are treated here as possessive classifiers.



Map 1. The Austronesian language family and major subgroups (Ross et al. 2011: xxv; reproduced courtesy of Pacific Linguistics)

- (3) *nátu sítí né-di*
 child shirt POSS.CLF(GENERAL)-3PL.POSS
 “the children’s shirts”

(Lichtenberk 1983a: 294)

For another example from Manam, this time with a pronominal possessor phrase, see (6) in §2.

The present paper has two central concerns. The first one is the paths of development of the systems of possessive classifiers from Proto-Oceanic to present-day languages. As will be seen, the number of possessive classifiers ranges from zero to large, difficult-to-determine numbers. The second concern has to do with the rise of the system of possessive classifiers in Proto-Oceanic. Here it will be necessary to go beyond the Oceanic group and to consider the existence of possessive classifiers in some non-Oceanic Austronesian languages as well as the existence of multiple possessive construction types in some non-Austronesian/Papuan languages spoken in areas where Austronesian languages are also spoken. The terms ‘non-Austronesian’ and ‘Papuan’ are interchangeable in the present context; the term ‘Papuan’ is used in this paper. Note that ‘Papuan’ does not signal genetic unity. There is no evidence that the non-Austronesian languages form a single family.

It has been suggested that the presence of possessive classifiers in Austronesian languages is due to contact with Papuan languages (Klamer et al. 2008; Donohue & Schapper 2008). In the absence of historical records of early Papuan–Austronesian contact, this paper critically assesses the hypothesis of multiple possessive types in Oceanic being ultimately due to Papuan influence. While the impetus for the development of an alienable–inalienable possession contrast may have come from Papuan languages, the more complex Proto-Oceanic system with possessive classifiers for alienable possession is clearly a later development, post-dating the emergence of the alienable–inalienable contrast.

Because, by definition, direct possessive constructions do not employ possessive classifiers, they do not figure prominently in the discussion. The paper is structured as follows. Section 2 provides more detail on the typical Oceanic pattern of possessive constructions, with both their formal and semantic/pragmatic aspects taken into account. Section 3 focuses on the diachrony of the Oceanic possessive classifier systems, from Proto-Oceanic to present-day languages. Section 4 deals with multiple possessive constructions in non-Oceanic Austronesian languages that belong to a sister group of Oceanic. Section 5 examines the claim that the existence of multiple possessive constructions in Austronesian, and specifically in Oceanic, is due to Papuan influence. Finally, §6 provides a summary and conclusions.

2. The typical Oceanic pattern of attributive possessive constructions

The presentation that follows is based on Lichtenberk (2009b), where more detail and references to other works can be found. Here only a brief overview is given by way of necessary background for the subsequent discussion of possessive constructions in Oceanic and outside Oceanic.

As mentioned in §1, Oceanic languages typically exhibit a formal contrast between direct and indirect possessive constructions. In terms of their semantics, possessive constructions express various types of real-world relations between the possessum and the possessor, of which literal possession – ownership – is only one. At the cost of some oversimplification, one can say that direct possessive constructions typically express inalienable possession; normally, though not exclusively, part-whole relations; and kinship relations. Indirect possessive constructions typically exhibit other types of possessum–possessor relations that can be loosely characterized as alienable possession. With some exceptions, languages that have a direct–indirect construction contrast have more than one subtype of indirect constructions distinguished by different possessive classifiers. The different subtypes express different kinds of possessum–possessor relations. Languages may exhibit various types of exceptions, or what appear to be exceptions. For example, only some kinship relations may be expressed by the direct construction, with others expressed by an indirect construction. However, on the whole, there is a robust link between direct constructions and inalienable possession, and indirect constructions and alienable possession.

As discussed in §3, the number of possessive classifiers in a language varies from one to more than 20.³ Most commonly, the number is two, three or four. In some languages with a large number of classifiers, their exact number is not evident. As a rule, all but one of the classifiers express a fairly restricted range of possessum–possessor relations, while the remaining one functions as a ‘general’ classifier, used when none of the other classifiers or the direct construction is appropriate. For example, Manam has two possessive classifiers. One is used when the possessum serves as an item of food or drink for the possessor or is metonymically linked to food or drink for the possessor; see (4) and (5), respectively:

- (4) *níu ʔaná-gu*
 coconut POSS.CLF(ALIM)-1SG.POSS
 “my coconut (as my food or drink)” (Lichtenberk 1983a: 300)

3. There are also a few languages without any classifiers; see §3.2.

- (5) *ʔilólo ʔaná-ŋ*
 string POSS.CLF(ALIM)-2SG.POSS
 “your fishline” (Lichtenberk 1983a: 292)

Besides the alimentary, food/drink classifier, Manam has a general classifier. This classifier is present in Examples (2) and (3) in §1, where the possessum is a shirt, and in (6) below, where the possessum is a canoe:

- (6) *ʔáiʔo ʔáti ne-ŋ*
 2SG canoe POSS.CLF(GENERAL)-2SG.POSS
 “your canoe” (Lichtenberk 1983a: 294)

More information on the types of possessum–possessor relations expressed by means of possessive classifiers is given in §3.

A feature of Oceanic possessive systems frequently commented on is fluidity, the possibility of a noun to occur in the head/possessum position in more than one type of possessive construction, with different types of possessum–possessor relations. There may be a contrast between the direct and the indirect constructions, as in (7) and (8), again from Manam:

- (7) *paŋaná gu*
 head 1SG.POSS
 “my head (head as part of my own anatomy)” (Lichtenberk 1983a: 302)
- (8) *paŋána ʔaná-gu*
 head POSS.CLF(ALIM)-1SG.POSS
 “my head (head meant for my consumption, e.g., a fish head)”
 (Lichtenberk 1983a: 302)

Alternatively, there may be a contrast between two or more types of indirect possessive constructions, illustrated in the next three examples from Fijian.⁴ Unlike Manam, where the same possessive classifier is used for items of food and drink for the possessor, Fijian has one classifier for food possession and another classifier for drink possession; see (9) and (10), respectively. Note that juicy, suckable food falls in the drink category.

- (9) *na ke-na maqo*
 ART POSS.CLF(FOOD)-3SG.POSS mango
 “his mango for eating (i.e., green mango)” (Pawley 1973: 168)
- (10) *na me-na maqo*
 ART POSS.CLF(DRINK)-3SG.POSS mango
 “his mango for sucking (i.e., ripe, juicy mango)” (Pawley 1973: 168)

4. ‘Fijian’ here designates Standard Fijian, sometimes also referred to as ‘Bauan’.

Fijian also has a general classifier, shown in (11). There the possessum, a mango, is neither food nor 'drink' for the possessor.

- (11) *na no-na maqo*
 ART POSS.CLF(GENERAL)-3SG.POSS mango
 "his mango (as property, e.g., which he is selling)" (Pawley 1973: 168)

The following examples from Lolovoli (Vanuatu) show fluidity between two indirect constructions and the direct construction. In the situation expressed in (12) the water will serve as drink for the possessor, and it is the drink classifier that is used:

- (12) *Na=ni utu na me-mu wai*
 1SG.SBJ=IRR draw.water ACC POSS.CLF(DRINK)-2SG.POSS water
 "I will draw you some water (to drink)." (Hyslop 2001: 181)

Water can be used to wash clothes, and in such cases the general classifier is used:

- (13) *Na=ni utu na no-mu wai.*
 1SG.SBJ=IRR draw.water ACC POSS.CLF(GENERAL)-2SG.POSS water
 "I will draw you some water (to wash with, or use for some purpose)."
 (Hyslop 2001: 181)

And water can be used for the possessor to bathe in, in which case the direct construction is used:

- (14) *Na=ni utu na wai-mu.*
 1SG.SBJ=IRR draw.water ACC water-2SG.POSS
 "I will draw you some water (to bathe with)." (Hyslop 2001: 181)

There is another semantic dimension relevant to possessive constructions that merits mention, the distinction between 'passive' (or 'subordinate') possession and 'active' possession. In passive possession the possessor is Undergoer-like in the associated situation, while in active possession the possessor is Actor-like. As a rule, passive possession is expressed either by means of the same classifier that expresses food or alimentary possession or by means of the direct construction, while active possession, when in contrast with passive possession, is expressed by means of the general classifier. (For more information on passive possession, see Lynch 2001.)⁵ There is often fluidity between passive and active possession.

In Fijian, passive possession is expressed by means of the same classifier that expresses food possession (see (9)):

5. Lynch mentions two other strategies for the expression of passive possession: (i) a dedicated classifier; and (ii) the general classifier.

- (15) *ke-mu* *i-vacu*
 POSS.CLF(FOOD)-2SG.POSS NMLZ-punch
 “your punch (you receive)” (Schütz 1985: 462)

Compare (15) with (16) the general classifier, expressing active possession:

- (16) *no-mu* *i-vacu*
 POSS.CLF(GENERAL)-2SG.POSS NMLZ-punch
 “your punch (you give)” (Schütz 1985: 462)

In Aroma (Papua New Guinea), passive possession is expressed by means of the direct construction, also used to express inalienable possession; see (17) and (18), respectively:

- (17) (*thau*) *rauparaupa-ku*
 1SG picture-1SG.POSS
 “my picture (depicting me)” (Lynch 2001: 196)
- (18) (*thau*) *ama-ku*
 1SG father-1SG.POSS
 “my father” (Lynch 2001: 196)

Active possession is expressed by means of the general possessive classifier:

- (19) (*thau*) *ye-ku* *rauparaupa*
 1SG POSS.CLF(GENERAL)-1SG.POSS picture
 “my picture (which I have/took/painted)” (Lynch 2001: 197)

Fluidity in Oceanic possessive constructions is by no means rare. It is for this reason that the use of the different types of possessive constructions is better analyzed as being primarily based on types of possessum–possessor relations rather than identifying noun classes (Lichtenberk 1983b). It needs to be recognized, however, that there are language-specific cases where the choice of a possessive construction is not (fully) predictable on such semantic/pragmatic grounds. (See Pawley & Sayaba 1990 for a perceptive discussion of these issues with respect to the possessive system of Wayan [Fiji].)

Finally, a language may have both possessive classifiers and noun/numeral/sortal classifiers. The two systems operate on different principles. The possessive classifier system is based, by and large, on types of possessum–possessor relations, while the noun/numeral/sortal classifier system is based on criteria such as shape, size, animacy, etc. of the referent of the noun phrase. Example (20) from Chuukese (Micronesia) shows the cooccurrence of the two systems within one overall noun phrase. *-fóc* is a numeral classifier (CLF) used for long, cylindrical objects, and *wínúmwó-* is a possessive classifier used for nonsolids for oral consumption.

- (20) *e-fóc wúnúmwo-mw suupwa*
 one-CLF POSS.CLF(NONSOLID)-2SG.POSS cigarette
 “your one cigarette” (Benton 1968: 136)

Benton (1968) provides an interesting discussion of the interplay between numeral and possessive classifiers in Chuukese. For more on numeral classifiers, see §3.2 below.

3. The diachrony of the possessive classifiers systems within Oceanic

3.1 The Proto-Oceanic possessive system

It is generally agreed that in Proto-Oceanic (POc) there was a basic binary distinction between direct and indirect possessive construction types (see, e.g., Pawley 1973; Lichtenberk 1985; Lynch et al. 2002). Within the indirect type, there was (at least) a tripartite distinction involving (at least) three possessive classifiers. The POc system is shown in Table 1, based on the discussion in Lynch et al. (2002). Only rough characterizations of the types of possessum–possessor relations are given in the table, and the position of possessor phrases is disregarded there, because it is not relevant to the discussion that follows.

Table 1. Proto-Oceanic system of possessive constructions (based on Lynch et al. 2002)

direct possessive construction: possessum NOUN-POSS.SUFF

inalienable possession

indirect possessive constructions (with possessive classifiers)

a. *ka-POSS.SUFF POSSESSUM.NOUN

food possession

b. *m^(w)a-POSS.SUFF POSSESSUM.NOUN

drink possession

c. *na-POSS.SUFF POSSESSUM.NOUN

general possession

Lynch et al. (2002) point out that besides the *na- form of the general classifier the following forms are also reconstructible: *a-, *ni-, *ne-, *i- and *e-. While these are, in principle, reconstructible, it is doubtful POc had six variant forms of a single classifier (however, see the discussion below). Lynch et al. also suggest that POc may have had other classifier-like forms, but the exact grammatical status of those forms is uncertain.

As far as passive possession (see §2) is concerned, I have suggested (Lichtenberk 1985) that in POc certain kinds of passive possession may have been expressed by means of the direct construction, while other kinds may have been expressed by means of the classifier *ka-, also used to express food possession. On the other hand, Lynch (2001) has argued that passive possession was (exclusively) expressed by the direct construction.

Early in the history of Austronesian there was no distinction between direct and indirect constructions. There was a possessive construction type that later continued as the direct type, but there were no constructions of the indirect type, with possessive classifiers. This is still the case in most, but not all, non-Oceanic Austronesian languages; see §4.

As to the etymology of the POc classifiers, only partial answers can be offered at present. All of the POc classifiers are monosyllabic/monomoraic. On the other hand, as discussed by Blust, lexical bases/roots in POc tended to be disyllabic, which was also the case in Proto-Austronesian (PAN) and in Proto-Malayo-Polynesian (PMP). Blust (2005: 547) speaks of “the predominant disyllabism of PAN, PMP, and POc lexical bases” (see also Blust 1977, 2009).⁶

If at least some of the POc possessive classifiers were the outcome of grammaticalization from lexical sources, the process would have probably taken place some time earlier, given the likely phonological erosion from a disyllabic source. As for the food classifier *ka-, it is generally agreed that it had some connection with the transitive verb *kani “eat”, possibly via a nominalization, for which Lynch et al. (2002: 70, 78) give two forms, *kani-an and *kan-an “food” (see also Lynch 1973, 2001; Lichtenberk 1985; but see (46b) in §6 below for another nominalized form.)

The etymology of the drink classifier *m^(w)a- is not clear, although, by analogy with the food classifier, one would expect it to have some connection with the POc verb *inum “drink” (Lynch 1973; Lichtenberk 1985), here too possibly via a nominalization.

The history of the general classifier *na- seems to be considerably different from those of the food and the drink classifiers. In any case, no verbal or nominal candidate has been identified as its (possible) source. Rather, Lynch et al. (2002) suggest that the classifier may be related to the common, apparently definite, article *na reconstructed for POc. In fact, Crowley (1985) has reconstructed two forms of the POc common article, *na and *a, and, as mentioned above (after Table 1), Lynch et al. say that the form *a- is also reconstructible as a possible form of the

6. In fact, in some present-day Oceanic languages ALL lexical roots must be at least disyllabic or bimoraic, for example in Fijian, Saliba and Toqabaqita, while in some others most lexical roots are at least disyllabic or bimoraic; see Lichtenberk (2009a) for discussion.

general classifier. About the other reconstructible forms – *ni-, *ne-, *i- and *e- (see above) – Lynch et al. (2002: 78) say the following:

These may be modified versions of the articles, but they also recall the forms of the demonstratives (§3.2.3). It seems semantically plausible that the ‘general’ classifier should be derived from either articles or demonstratives, but at present we cannot be sure which.

The demonstrative forms that Lynch et al. say are reconstructible for POc are shown in Table 2, although, as they point out, it is unlikely that all of them did, in fact, occur in POc.

Table 2. Demonstratives reconstructible for POc (after Lynch et al. 2002: 72)

“here”, “this”, “near speaker”	*i, *e	*ni, *ne	
“there”, “that”, “near hearer”	*a	*na	*ri
“yonder”, “away from both interlocutors”	*o, *u	*no, *nu	*rai

The general classifier *na- and the possible variant *a- would then be related to the common article forms *na and *a, respectively, which in turn are related to two of the near-hearer demonstratives. The other possible variants of the general classifier – *i-, *e-, *ni- and *ne- – would be related to the proximal demonstratives, although, as mentioned above, it is unlikely there were six variants of the general classifier in POc. It is more likely that there have been multiple parallel developments in Oceanic (proto-)languages that have given rise to the apparent reconstructibility.

The development of demonstratives into articles is a common historical process (e.g., Greenberg 1978, Mulder & Carlier 2011). It is not implausible that articles are historical antecedents of the general possessive classifier. As discussed in §3.3 below, articles have been implicated in the development of a possessive system in the Polynesian languages, a system that is a considerable departure from the typical Oceanic pattern.

3.2 Reductions and expansions in Oceanic possessive systems

This section considers in detail the kinds of possessive constructions found in Oceanic. Because of the large number of Oceanic languages (around 450), only some representative cases are discussed here. In some languages the Proto-Oceanic system of possession contrasts has continued unchanged in its basics. In some others, it has been reduced, while in others it has become more complex, even considerably so.

POc had direct possessive constructions and three subtypes of indirect constructions for food possession, drink possession and other possession. This kind of system still exists in some present-day languages, such as Fijian. Direct possession is illustrated in (21), and food, drink and general possession in (22), (23) and (24), respectively, with the last two providing another illustration of fluidity (see §2):

- (21) *na tama-na*
 ART father-3SG.POSS
 “her father” (Schütz 1985: 447)
- (22) *na ke-na uvi*
 ART POSS.CLF(FOOD)-3SG.POSS yam
 “her yam” (Schütz 1985: 450)
- (23) *na me-qu yaqona*
 ART POSS.CLF(DRINK)-1SG.POSS kava
 “my kava (which I drink)” (Milner 1972: 66)
- (24) *na no-qu yaqona*
 ART POSS.CLF(GENERAL)-1SG.POSS kava
 “my kava (which I grow or sell)” (Milner 1972: 66)

In some languages the direct–indirect contrast continues, but there is only a binary contrast in the indirect category between alimentary and general possession. The food–drink contrast has disappeared. This is the case in, for example, Manam; see (1) in §1 (direct possessive construction), (4) in §2 (alimentary possession) and (2) in §1 (general possession).

There are languages where the direct–indirect contrast also continues but there is only one possessive marker in an indirect construction. This is the case in, for example, Houaïlou (New Caledonia). Example (25) contains the direct construction, and Example (26) the indirect one:

- (25) *pani-na*
 mother-1SG.POSS
 “my mother” (La Fontinelle 1976: 300)
- (26) *tawa yi-na*
 dog POSS-1SG.POSS
 “my dog” (La Fontinelle 1976: 300)

In languages such as Houaïlou, the sole element that carries the possessor-indexing suffixes in the indirect construction is unlike the possessive classifiers in languages with two or more classifiers, because it does not contrast with any other. In Houaïlou too there is scope for fluidity. In (27) the noun *mwā* functions as the possessum in the indirect construction, and in (28) as the possessum in the direct construction, with a semantic difference:

- (27) *mwā yi-na*
house POSS-1SG.POSS
“my house” (La Fontinelle 1976: 172)
- (28) *mwā-e*
contents/shell-3SG.POSS
“its contents”, “its shell” (La Fontinelle 1976: 172)

Some languages have no possessive classifiers. Some do maintain a formal binary distinction between expressions of inalienable possession and alienable possession. This is the case in Toqabaqita (Solomon Islands) and its near relatives. Toqabaqita has a possessive construction that continues the POC direct construction, used to express inalienable possession:

- (29) *maa-ku*
eye-1SG.PERS⁷
“my eye(s)” (Lichtenberk 2008: 399)

In the other construction type the possessor can only be expressed by means of a separate phrase; there is no indexing by a personal suffix on the possessum or elsewhere. This construction is used to express alienable possession, with a few kinship terms, and it replaces the direct/suffixing construction when the possessum is individuated (see Lichtenberk 2008, §8.1.7 for detail). In (30) the ‘bare’ (non-suffixing) construction expresses alienable possession. The possessor is indexed by means of the first person singular independent pronoun, not by means of a possessive suffix:

- (30) *waqi nau*
basket 1SG
“my basket” (Lichtenberk 2008: 405)

And in (31) it is used even though the possession is inalienable, because the possessum has been individuated: “my left eye (as opposed to the right one)”; cf. (29), where the possessum is not individuated (either eye, or both eyes).

- (31) *maa mauli nau*
eye be.on.left.side 1SG
“my left eye” (Lichtenberk 2008: 399)

There is at least one language where even the binary alienable–inalienable possession contrast has been obliterated and there is only one type of possessive construction, regardless of the type of possessum–possessor relation, Tobati (Papua New Guinea) (Donohue 2002). However, alienable and inalienable possession

7. The personal suffixes correspond to the possessive suffixes in other Oceanic languages but have other functions outside of possessive constructions; hence the term ‘personal’ suffixes.

may, optionally, be distinguished in cases of intended possessors: there is a dative construction, ‘the item that is for X’, available for alienable possession but not for inalienable possession.

We can now consider types of possessive systems that are expansions of the POC system through the development of additional possessive classifiers. There has been no development of further distinctions within the direct possessive construction category (although in a few languages there are classifiers for certain kinship terms; see the discussion of Chuukese below). In a group of Vanuatu languages a fourth possessive classifier has been added to the original three. The new classifier had the form *bula in the proto-language ancestral to these languages, and it was used when the possessum was conceptualized as valuable property for the possessor. A reflex of this classifier is found in, for example, Lolovoli. Hyslop (2001: 176) says that Lolovoli *bula* signals “natural or valued possession”. It is used primarily to express “ownership of animals ... and ... crops”; see (32), where the classifier occurs twice.

- (32) *Bula-na* *boe mo gani na*
 POSS.CLF(VAL)-3SG.POSS pig REAL eat ACC
bula-da *toa tamwere*.
 POSS.CLF(VAL)-1NONG(INCL).POSS chicken always
 “His pig is always eating our chickens.” (Hyslop 2001: 178)

The same classifier is also used when reference is being made to “some items introduced by Europeans,” such as radios or watches, and to “objects of adornment,” such as earrings and bracelets (Hyslop 2001: 179).

For the etymology of the valuable possession classifier, Pawley (1973: 165) suggests that “[t]he Bauan [Fijian] verb *bula* ‘be alive’ may be cognate.” And Alexandre François has informed me (personal communication, 31 May 2010) that three languages spoken in northern Vanuatu have nouns with the meaning “estate, property” that reflect earlier *bula. As examples he gives Löyöp *n-pələ-k*, Mwotlap *na-mle-k* and Lemerig *n-pələ-k*, all of which are regular reflexes of *na-bula-gu (ART-estate-1SG.POSS). In all three languages the reflexes of *bula are nouns that occur in the direct possessive construction. They are not possessive classifiers.

Tamabo is another language with a four-member classifier system with a reflex of *bula. Tamabo *bula* is used for possessums that are “living things which one owns, and for one’s own use (note that this does not include pigs, which are regarded as ‘money’ and take [the general classifier] *no*” (Jauncey 2002: 615). The association of reflexes of *bula with living things in Tamabo and some other languages can be interpreted as support for Pawley’s suggestion of cognacy of Vanuatu *bula and the Fijian verb *bula* “be alive”.

A four-member classifier system is also found in, for example, Southeast Ambrym (Vanuatu), which besides a food, a drink and a general classifier has a classifier for ‘domestic plants and animals’, *sa-* (Crowley 2002: 664), which is not a reflex of *bula.

Table 3. Mussau possessive classifiers and related nouns (based on Table 17 in Brownie & Brownie 2007: 77)

Classifier	Type of possession	Meaning and form of related noun or verb
<i>ai-</i>	trees	“tree, wood” (noun) <i>ai</i>
<i>ale-</i>	building	“house” (noun) <i>ale</i>
<i>ane-</i>	food	
<i>gholu-</i>	juicy foods	“eat something juicy” (verb)
<i>iema-</i>	cutting instruments	“knife” (noun) <i>iema</i>
<i>ilimo-</i>	water craft	“canoe” (noun) <i>ilimo</i>
<i>kalu-</i>	abstract personal items	
<i>kapu-</i>	personal relationships	“friend, sibling” (noun) <i>kapu</i>
<i>kie-</i>	domestic animals	
<i>kura-</i>	things giving light	“fire” <i>kura</i>
<i>ropi-</i>	things to drink	“drink” (verb) <i>ropi</i>
<i>uma-</i>	pieces of land, gardens	
<i>une-</i>	general things	
<i>usu-</i>	things to suck juice from	“suck juice from” (verb) <i>usu</i>

At least one of the Mussau classifiers can be used as a repeater. (For more on repeaters, see below.) In (35) the ‘tree, tall plant’ classifier *ai-* is used with the possessum noun *niu* ‘coconut’, while in (36) it is used with the noun *ai* ‘tree’:

- (35) *ai-qi* *niu*
 POSS.CLF(TREE)-1SG.POSS coconut
 “my coconut tree” (Ross 2002: 157)
- (36) *ai-qi* *ai*
 POSS.CLF(TREE)-1SG.POSS tree
 “my (tall) tree” (Ross 2002: 157)

Finally, there are also several languages with large numbers of possessive classifiers, the total number not necessarily known. Iaaï (New Caledonia) is one such language. Ozanne-Rivierre (1976) gives over 20 classifiers for Iaaï, but she also says that the set is open and a complete list is difficult if not impossible to determine. She also points out the existence of repeaters: for example, there is a habitat classifier *umwa*, related to the noun *uma* ‘house’:

- (37) *umwə-k* *uma*
 POSS.CLF(HABIT)-1SG.POSS house
 “my house” (Ozanne-Rivierre 1976: 191)

The classifier and the noun ‘house’ reflect POc *Rumaq ‘house’. A similar, independent development of POc *Rumaq into a classifier has taken place in some Micronesian languages, such as Chuukese (Table 5) and Ulithian (see (38)).

Some Iaaï classifiers are related to nouns; some others are derived from verbs. For example, there is a classifier for chewables derived from the verb ‘chew’. However, the formal nature of such verb-to-classifier relations is not specified.

Several Micronesian languages have large numbers of possessive classifiers, and repeaters are by no means uncommon. For Chuukese, Benton (1968) lists over 30 classifiers. He says that repeaters have been excluded, but even so the list may not be complete. The classifiers are listed in Table 4.

Table 4. The semantic categories of Chuukese possessive classifiers (Benton 1968: 133–135)

<i>yaa</i>	general; qualities, activities, emotions, miscellaneous objects
<i>yacawara</i>	loincloth
<i>yana</i>	fish or fruit to be eaten cooked, rice and other staples normally cooked
<i>yanoo, yónóó</i>	speech, word (two different classifiers)
<i>yasama</i>	door opening, entrance
<i>canú</i>	source of liquid
<i>cceñikama, kuusa</i>	bed covering (two different classifiers)
<i>yepita</i>	oil for anointing
<i>yéméti</i>	food for a journey
<i>fataa</i>	planted tree
<i>fénúwa</i>	island
<i>yimwa</i>	overhead shelter
<i>yina</i>	mother
<i>yipwa</i>	footwear
<i>kiya</i>	mat, object for sleeping on
<i>masa</i>	related to the eye
<i>mwari</i>	lei, encircling object
<i>neeniya</i>	place
<i>nánáá</i>	bracelet, ring
<i>nawú</i>	child, property closely associated with person, intimate, subordinate
<i>nikita</i>	leftovers
<i>ñjuta</i>	food for chewing
<i>wocaa</i>	fruit or fish to be eaten raw, meat
<i>piseki</i>	miscellaneous personal property
<i>paa</i>	bait
<i>pwúni</i>	borrowed object
<i>sama</i>	superior, person of father’s generation
<i>sepiya</i>	plate, dish
<i>wúfa</i>	clothing worn on the torso
<i>wúnúma</i>	nonsolid for oral consumption
<i>waa</i>	vehicle
<i>wosa</i>	bundled object
<i>wóku</i>	sticklike object
<i>wúnna</i>	pillow

Many of the classifiers are clearly related to nouns, some of which continue POC nouns. And although Benton (1968: 132) says “excluded [from the list] are possessive repeaters,” some of his examples of the use of the classifiers do include possessive nouns etymologically related to the classifiers. In most of those cases Benton gives a related noun; furthermore, there may be a POC ancestral form. Some of the POC forms continue earlier, pre-POC etyma. In one case, no Chuukese related noun is given, but there is a POC ancestral form. The alphabetization and the use of italics and the slashes in Table 5 are Benton’s.

Table 5. Chuukese possessive classifiers and related nouns (from Benton 1968: 133–135)

Possessive classifier	Related noun
<i>yacawara</i> loincloth	/akkacawar/ “loincloth”
<i>yasama</i> door opening, entrance	/asamwaco/ “window”
<i>canú</i> source of liquid	/caan/ “juice”
<i>ccenikama</i> , <i>kuus</i> bed covering	/ccenikam/, /kuus/ “blanket, sheet”
<i>yepita</i> oil for anointing	/epit/ “fragrant oil”
<i>fénúwa</i> island	POC *panua “land, ground”, “land mass”
<i>yimwa</i> overhead shelter	/iimw/ “house”, POC *Rumaq “house”
<i>yina</i> mother	/iin/ “mother”, POC *tina “mother”
<i>yipwa</i> footwear	/iipw/ “footprint”
<i>masa</i> related to the eye	/maas/ “eye”, POC *mata “eye”
<i>mwari</i> lei, encircling object	/mwaramwar/ “lei”
<i>nawú</i> child, property closely associated with person, intimate, subordinate	/naaw/ “child”
<i>wocaa</i> fruit or fish to be eaten raw, meat	/wuuc/ “banana”
<i>sama</i> superior, person of father’s generation	/saam/ “father”, POC *tama “father”
<i>wúnúma</i> nonsolid for oral consumption	POC *inum “drink”
<i>waa</i> vehicle	/waa/ “canoe”, POC *waga “canoe”
<i>wóku</i> sticklike object	/wóók/ “stick”
<i>wúnna</i> pillow	/wuun/ “pillow”

* The glosses in the table are rudimentary. Pawley (2005: 214–215), from which the glosses have been extracted, gives the following senses for *panua: “1. ‘land, ground, the solid surface of the earth (as opposed to sea, water, sky)’”; “2. ‘a land mass and whatever features are an integral part of it (forests, lakes, rivers, settlements, etc.), any sizeable tract of territory or country’”; “3. (a) ‘territory belonging to a person or community’”; “(b) ‘homeland, home place’”; “4. ‘place, area, district, region’”; “5. in certain multi-word expressions: ‘the world, that which is subject to the day-night cycle, weather and climate’”.

Several aspects of the situation in Chuukese are also found in other Micronesian languages with large numbers of classifiers. The first is the existence of repeaters. The second is the existence of possessive classifiers for kinship terms or related to kinship terms, which is not what one finds in the typical Oceanic pattern, where kinship terms occur in the direct possessive construction. (In some languages not all of the kinship terms occur in the direct construction. Some occur in an indirect

of the possessive systems in a number of Polynesian languages can be found in Fischer (2000), together with an overview by Clark (2000) and a discussion of the semantic correspondences between the Polynesian pattern and the earlier Proto-Central Pacific pattern by Lynch (2000b), with the Proto-Central Pacific pattern essentially continuing the POC pattern.

Let's first look at some formal properties of the Polynesian possessive systems. The semantic/pragmatic aspects are discussed subsequently. Examples (39) and (40) illustrate the A–O possession contrast from Cook Island Maori, where *aa* and *oo* are possessive prepositions:

(39) *te tuatua aa te ariki*
 DET word POSS DET chief
 “the words of the chief” (Clark 2000: 258, originally Buse 1996: 1)

(40) *te 'are oo Tere*
 DET house POSS Tere
 “Tere's house” (Clark 2000: 258, originally Buse 1996: 276–277)

The Polynesian languages have a plethora of possessive forms within the A–O contrast (Clark 2000). In some of them, the possessive prepositions combine with other grammatical elements. In one, they combine with the specific article, which in Proto-Polynesian had the form **te*. (See its determiner reflex *te* in Cook Islands Maori in (39) and (40).) The combination of the article and the possessive prepositions took possessive suffixes. Wilson (1982) gives a partial list of such forms for Proto-Nuclear Polynesian, one of the two primary subgroups of Polynesian; see Table 6, where the article **te* is reduced to **t*.

Table 6. Proto-Nuclear Polynesian forms with the specific article (after Wilson 1982: 61)

	A-possession	O-possession
“my”	* <i>t-a-ku</i>	* <i>t-o-ku</i>
“your (sg)”	* <i>t-a-u</i>	* <i>t-o-u</i>
“his”	* <i>t-a-na</i>	* <i>t-o-na</i>
“our (incl, du)”	* <i>t-a-taa</i>	* <i>t-o-taa</i>

Examples from Cook Island Maori are given in (41) and (42):

(41) *t-aa-ku va'ine*
 DET-POSS-1SG.POSS wife
 “my wife” (Clark 2000: 260, originally Buse 1996: 430)

(42) *t-oo-na manako*
 DET-POSS-3SG.POSS plan
 “his plan” (Clark 2000: 260, originally Buse 1996: 507)

In some Polynesian languages, possessive forms have also incorporated the indefinite article. Both types of forms, with an incorporated specific and indefinite article, existed already in Proto-Polynesian (Wilson 1982; Clark 2000).

As mentioned in §3.1, it is possible that the POc general possessive classifier **na-*, which took possessive suffixes, derives historically from the common article. Similarly, in the history of Polynesian the specific and the indefinite articles were implicated in the development of a set of possessive forms that took possessive suffixes (although possessive prepositions were involved as well).

A different set of possessive forms expresses “irrealis possession”, which serves to express “the intention or anticipation that something will be possessed” (Clark 2000: 262). Here the possessive prepositions combine with an element *m-*, “which can plausibly be identified with an irrealis or optative marker *me* found in various Polynesian languages and Fijian” (Clark 2000: 262). An example of irrealis A-possession is given in (43) from New Zealand Māori:

- (43) *Ka m-aa-ku te whenua.*
 INC IRR-POSS-1SG.POSS DET land
 “The land will become mine.”

(Clark 2000: 262, originally Williams 1971: 171)

On the semantics of the O–A possession contrast, Lynch (2000b) suggests that O-possession continues the semantics of the POc direct construction while A-possession continues the semantics of the indirect constructions (food, drink and general). On the other hand, Wilson (1982) has proposed an Initial Control Theory, the relevant factor being whether the possessor does or does not have control over the initiation of the relation to the possessum. If the possessor does have control, an A-possessive construction is used; if not, an O-possessive construction is used, as illustrated by (44) and (45) from Hawaiian. A person has, in principle, control over having a child (A-possession), but he/she does not have control over having a parent (O-possession).

- (44) *k-ā-na keiki*
 ART-POSS-3SG.POSS child
 “his child”

(Wilson 1982: 19)

- (45) *k-o-na makua*
 ART-POSS-3SG.POSS parent
 “his parent”

(Wilson 1982: 19)

As a referee points out, there may be quite a lot of variation among the Polynesian languages with respect to the semantics/pragmatics of their possessive systems, and it is possible that for some languages it is Lynch’s (2000b) view that provides a better account, while for others it is Wilson’s (1982) view.

In Polynesian languages too one finds fluidity; for example, “my name (that represents me)” (O-possession) and “my name (that I bestow on someone)” (A-possession); that is, passive possession vs. active possession (see §2). See the corresponding examples from Hawaiian in Wilson (1982: 15), (2.1) and (2.2), respectively, there.

3.4 Possessive constructions in Oceanic: A brief summary

In the typical Oceanic pattern there is a basic binary contrast between a direct construction and an indirect construction. Within the indirect construction, there are further subtypes employing different possessive classifiers. By and large, the direct construction expresses inalienable possession, while the indirect constructions express alienable possession. There is a general classifier, used when none of the other constructions is appropriate. There are no subtypes of the direct construction.

There is some fluidity, whereby one and the same noun can occur as the head/possessum in different subtypes of possessive constructions, depending on the semantics/pragmatics of the possessum–possessor relation.

POc had a direct construction, and three subtypes of indirect constructions with food, drink and general classifiers. With the exception of Polynesian, the direct construction continues in the present-day languages. In some languages the number of possessive classifiers has been reduced from the three in POc. In some others there has been an increase in the number of possessive classifiers, which in some of those languages has been considerable.

In the Polynesian languages the direct–indirect formal contrast has been replaced by a different binary contrast, between A-possession and O-possession. There the contrast is carried by two prepositions, which in some constructions fuse with other elements. The A–O contrast too is classificatory.

We now turn our attention to possessive constructions in non-Oceanic Austronesian languages.

4. Multiple possessive constructions in non-Oceanic Austronesian languages

Pawley (1973) presented a number of grammatical phenomena that he argued were innovations attributable to POc (Pawley 1973). One of the construction types reconstructed by Pawley was a system of attributive possessive constructions. In more recent terminology, Pawley posited three types of possessive constructions: a direct construction and two subtypes of indirect constructions with possessive classifiers. One of the classifiers served to express edible and passive/subordinate possession,

and the other one served to express possession that Pawley called ‘dominant’, which corresponds to general possession. Pawley also noted the possibility of a third classifier for drink possession, but did not include it in his list of POc innovations because he did not find the evidence for reconstructing it compelling. Today the existence of three possessive classifiers in POc is generally accepted; see §3.1.

In a 1985 paper on Oceanic possessive constructions I mentioned the existence of multiple possessive construction types in some Austronesian languages that are not Oceanic (Lichtenberk 1985). On the basis of the data available then, I suggested that a formal distinction between inalienable and alienable possession with or without possessive classifiers had emerged before POc times.

Since 1985 considerably more data have become available for non-Oceanic Austronesian languages yielding more evidence of a formal distinction between inalienable and alienable possession and more evidence of a distinction between direct and indirect possessive constructions with the latter type using possessive classifiers. Systems of multiple possessive constructions similar in some respects to the typical Oceanic pattern are found in the South Halmahera–West New Guinea group of languages, which form a sister group of Oceanic (Figure 1 and Map 1 in §1). There is no evidence of such possessive construction systems in Austronesian subgroups more distantly related to Oceanic, such as the Central Malayo-Polynesian languages (but see fn. 8 for Selaru) and the Western Malayo-Polynesian languages.

Van den Berg (2009) surveys the systems of possessive constructions in the South Halmahera–West New Guinea subgroup. He looks in detail at 11 languages and reconstructs a system of possessive constructions for Proto-South Halmahera–West New Guinea (PSHWNG) shown in Table 7. Here too there is no indication of the position of possessor phrases.

Table 7. Proto-South Halmahera–West New Guinea system of possessive constructions (based on van den Berg 2009: 352)

direct possessive construction: *possessum* NOUN-POSS.SUFF

inalienable possession

indirect possessive constructions (with possessive classifiers)

a. **na*-POSS.SUFF POSSESSUM.NOUN

alimentary (food and drink) possession

b. **ri*-POSS.SUFF POSSESSUM.NOUN

general possession

As Table 7 shows, the PSHWNG system had a formal contrast between inalienable and alienable possession. The contrast was in terms of a direct construction and two subtypes of indirect constructions with possessive classifiers. The direct construction was used to express inalienable possession, which included body-part

relations, kinship relations and locative relations. One of the indirect constructions employed a classifier for alimentary possession, which van den Berg terms ‘edible possession’. The category of alimentary possession also included tools, “probably originally only tools used to procure or consume food” (van den Berg 2009: 351). The other classifier was for general possession. (However, van den Berg is not completely confident about the existence and the form of the general classifier.)

Apart from the number of classifiers there are clear parallels between the PSHWNG and the POc systems, compared in Table 8.

Table 8. Systems of possessive constructions in Proto-South Halmahera–West New Guinea and in Proto-Oceanic

PSHWNG (van den Berg 2009)	POc (Lynch et al. 2002)
direct construction: possessum.NOUN-POSS.SUFF inalienable possession	direct construction: possessum.NOUN-POSS.SUFF inalienable possession
indirect constructions (with classifiers): a. *na-POSS.SUFF POSSESSUM.NOUN alimentary (food, drink) possession	indirect constructions (with classifiers): a. *ka-POSS.SUFF POSSESSUM.NOUN food possession ----- b. *m ^(w) a-POSS.SUFF POSSESSUM.NOUN drink possession -----
----- b. *ri-POSS.SUFF POSSESSUM.NOUN general possession	c. *na-POSS.SUFF POSSESSUM.NOUN general possession

With the possible exception of the third person plural, the possessive suffixes in PSHWNG and POc are cognate; see Table 9. The form of the PSHWNG third person plural suffix is uncertain.

Table 9. Possessive suffixes reconstructed for Proto-South Halmahera–West New Guinea (van den Berg 2009: 347) and Proto-Oceanic (Lynch et al. 2002: 67)

	PSHWNG	POc
1SG	*-ku	*-gu
2SG	*-mu	*-mu
3SG	*-na	*-ña
1PL(EXCL)	*-mam	*-ma[m]i
1PL(INCL)	*-ta	*-da
2PL	*-miu	*-m[i]u
3PL	*-ri?	*-dra

On the basis of the similarities between the two systems, van den Berg posits a direct–indirect possessive construction contrast for the ancestor of the two proto-languages, that is for Proto-Eastern Malayo-Polynesian (see Figure 1, §1). The paradigms of the possessive suffixes in PSHWNG and in POc are nearly identical. However, this in itself is not a particularly strong argument for reconstructing the direct–indirect contrast for Proto-Eastern Malayo-Polynesian, because, with the one possible exception of the PSGWNG third person plural form, the suffixes are retentions from earlier stages of Austronesian.

Another thing that is common to POc and PSHWNG is the existence of indirect possessive constructions with possessive classifiers. However, although both systems have a classifier of the form *na-, in POc *na- functioned as a general classifier, while in PSHWNG *na- functioned as an alimentary classifier. If it is indeed the case that the POc classifier *na- was historically linked to the specific article (see §3.1) and if the POc and PSHWNG *na forms are cognate, then the development of the general-classifier function from the article function would a priori seem more plausible than the development of the alimentary-classifier function. That would mean that the alimentary-classifier function in PSHWNG would have been a later reinterpretation of the earlier general-classifier function. Such a shift is not implausible. Shifts in classifier functions have taken place in recent times. According to Alexandre François (personal communication, 26 May 2012), in a few Oceanic languages of northern Vanuatu an erstwhile general classifier has become a drink classifier and a classifier for ‘intimate possession’ (such as items of clothing and other personal items). At this stage, however, it is not possible to conclude that the PSHWNG and the POc *na- forms are indeed cognate. Given the small phonological size of the two classifiers, an accidental identity of form is a relatively strong possibility.

PSHWNG had only two possessive classifiers, while POc had (at least) three. Did the ancestor of PSHWNG and POc have a two-member system or a three-member system of possessive classifiers? If a two-member system, then it might have been like the one in PSHWNG, in which case the three-member POc system would have developed later. However, there would be no straightforward explanation for why the classifier *na- marked alimentary possession if it was related to an earlier article. On the other hand, if the earlier system was a three-member one, then one would have to assume its reduction to a two-member system in PSHWNG, together with the shift in the function of the classifier *na-, from marking general possession to marking alimentary possession. A system with a direct–indirect contrast and with possessive classifiers in the indirect construction existed already in the ancestor of PSHWNG and POc, that is in Proto-Eastern Malayo-Polynesian. What is not clear is the number of possessive classifiers.

Overall, there is less homogeneity in the possessive systems of the South Halmahera–West New Guinea group of languages than in the Oceanic languages, in spite of the great difference in the number of languages in the two groups: 40+ in the former and around 450 in the latter. There is a well-defined typical Oceanic pattern, as discussed in §2. On the other hand, no such typical pattern is evident in the South Halmahera–West New Guinea group, keeping in mind that only a handful of those languages have been described in some detail. Most likely this difference between the two groups is to be attributed to the fact that, unlike speakers of South Halmahera–West New Guinea languages, speakers of pre-POc ceased to be in intensive contact with speakers of Papuan languages quite some time ago. Archaeological evidence suggests that speakers of pre-POc arrived in what was to become the homeland of Oceanic around 3,400 years ago, ultimately from East Nusantara (Pawley 2007). (See §5 on the East Nusantara area.) The homeland area of Oceanic was the Bismarck Archipelago, several hundred kilometres east of East Nusantara. On the other hand, speakers of South Halmahera–West New Guinea languages have remained in close contact with speakers of various Papuan languages, which may have resulted in the great heterogeneity in possessive constructions. A case in point is the affixal expression of the possessor, either on the possessum noun or on a possessive classifier. Early in the history of Austronesian, possessors were indexed by suffixes on the possessum noun (there were no possessive classifiers). Suffixal expression of possessors, on the possessum noun or on a classifier, is, with a few marginal exceptions, also the pattern in Oceanic. (A few languages have possessive prefixes, but the dominant pattern even in those languages is suffixing.) On the other hand, in South Halmahera–West New Guinea there has been a drift away from suffixal indexing of the possessor (still found in a few languages) to prefixal indexing. This has been attributed by Klammer et al. (2008) and by Donohue & Schapper (2008) to contact with Papuan languages, which employ the prefixing strategy (see also fn. 11 below).

5. Multiple possessive constructions in Papuan languages

Klammer et al. (2008) investigate contact between Austronesian and Papuan languages in East Nusantara. They define East Nusantara as the area comprising the islands of eastern Indonesia, East Timor and the Bird's Head Peninsula in the extreme north-west of the island of New Guinea (Map 1 in §1).

Klammer et al. (2008) argue that East Nusantara is a linguistic area, and they identify five features shared by Papuan and Austronesian languages in the area. Two have to do with possessive constructions. One concerns the existence of a grammatical distinction between alienable and inalienable possession, and the other has

to do with the linear order of the expressions of the possessum and the possessor. I restrict my discussion of Klamer et al.'s study to the alienable–inalienable contrast. (See §6 for a list of the five areal features.)

Klamer et al. argue that the presence of the grammatical contrast between alienable and inalienable possession in Austronesian languages, including the Oceanic languages, is due to contact with Papuan languages, rather than being an internal development within Austronesian. They see the alienable–inalienable contrast in the Papuan languages as “a unified phenomenon” (Klamer et al. 2008: 120): nearly all of the Papuan languages of East Nusantara have it. On the other hand, there are Central/Eastern Malayo-Polynesian languages spoken in the area, but only some of them exhibit the contrast. From this Klamer et al. conclude that the development of the contrast is not an innovation of Proto-Central/Eastern Malayo-Polynesian. Rather, its presence in the languages that have it is due to contact with Papuan languages. And they say specifically (2008: 122) that this was also the way the grammatical contrast entered the precursor of POC, referred to as ‘pre-POC’. Pre-POC is assumed to have been spoken in or next to the Bird’s Head area. It was only later that speakers of pre-POC moved east along the north coast of New Guinea into the Bismarck Archipelago area, which was to become the homeland of Oceanic. The Bismarck Archipelago is located off the northeast coast of New Guinea, west of the Solomon Islands (see Map 1 in §1).

Another recent study, Donohue & Schapper (2008), also argues that the existence of the alienable–inalienable distinction in Central/Eastern Malayo-Polynesian languages is due to contact with Papuan languages. Donohue and Schapper suggest that the presence of indirect possessive constructions with possessive classifiers in the Austronesian languages is due to Papuan influence. They see the appearance of possessive classifiers in the Eastern Malayo-Polynesian languages as part of a more general shift to preposed possessors under the influence of Papuan languages (see fn. 11). (Earlier in the history of Austronesian the expression of the possessor followed the expression of the possessum.) What diffused from Papuan to Austronesian, according to Donohue and Schapper, was a construction type, that is an indirect possessive construction with a possessive classifier, not the classifiers themselves. Here, I focus on Klamer et al. (2008), which develops the argument for Papuan influence on Austronesian in considerably more detail than Donohue & Schapper (2008), but some of what is said about the former also applies to the latter.

We do not have direct evidence of Papuan–Austronesian contact in the relevant time period, and so we can only hypothesize about what may have happened. I have no new evidence to offer whether or not it was indeed contact with Papuan languages that gave rise to the systems of possessive constructions in Oceanic and in the non-Oceanic languages of East Nusantara, but the claims of Papuan influence on Austronesian invite a closer look.

First, POc had three possessive classifiers for alienable possession: a food classifier, a drink classifier and a general classifier; see Table 8 in §4. In the non-Oceanic Austronesian languages discussed by van den Berg (2009) there is only a binary contrast in the alienable category between alimentary and general possession; see Table 8 again.⁸ On the other hand, those Papuan languages of East Nusantara that have a formal contrast between alienable and inalienable possession do not make further distinctions within the alienable category of the kinds found in the Austronesian languages. And so, even if the original impetus for the alienable–inalienable distinction did come from Papuan languages, the developments of the contrasts within the alienable category were Austronesian innovations. Furthermore, the tripartite food–drink–general contrast is not found in any of the non-Oceanic Austronesian languages and so was an innovation in POc.

Speakers of what was to become POc reached the Bismarck Archipelago, the homeland of Oceanic, around 3,400 years ago, having left the East Nusantara area some time before that. While the presence of multiple possessive types in non-Oceanic Austronesian languages spoken today in East Nusantara can perhaps be attributed to relatively recent contact with Papuan languages, the stimulus for the development of the POc system would have to have operated BEFORE speakers of pre-POc left the area, before 3,400 or so years ago. The hypothesis of Papuan influence on Proto-Eastern Malayo-Polynesian requires reconstruction of a grammatical system with an alienable–inalienable opposition in a Papuan proto-language that existed earlier than 3,400 years ago. The onus is on proponents of the hypothesis of Papuan influence on Austronesian to show that the alienable–inalienable grammatical contrast did exist in a Papuan proto-language at the time of pre-POc. And the Papuan proto-language would have had to have been spoken in an area where intensive contact with pre-POc was possible.⁹

8. Some, but not all, of the Central Malayo-Polynesian languages, which form a sister group of Eastern Malayo-Polynesian (see Figure 1 in §1), have a formal contrast between alienable and inalienable possession, and at least one of them, Selaru, makes an alimentary–general contrast by means of possessive classifiers (Laidig 1993). However, the Selaru system is most likely the result of a separate development, possibly as a result of contact with one or more South Halmahera–West New Guinea languages. Furthermore, because a formal alienable–inalienable contrast exists in only some Central Malayo-Polynesian languages, reconstructing it for Proto-Central/Eastern Malayo-Polynesian is not warranted (Klamer et al. 2008; Donohue & Schapper 2008).

9. There is another type of change in possessive constructions found in South Halmahera–West New Guinea languages. Early in the history of Austronesian, possessors were indexed by means of suffixes on possessum nouns. Most of the South Halmahera–West New Guinea languages exhibit a shift, partial or complete, from possessive suffixes to possessive prefixes. Klamer et al. (2008) and Donohue & Schapper (2008) attribute this shift to influence from Papuan languages, where only prefixes are found. On the other hand, in POc, possessors were indexed by suffixes

There is another point that needs to be made. As mentioned earlier, in support of their hypothesis of Papuan influence on Austronesian, Klammer et al. (2008) mention the fact that the alienable–inalienable distinction is not found in all Central/Eastern Malayo-Polynesian languages (see also fn. 10 above). On the other hand, they characterize its presence in the Papuan languages of East Nusantara as “a unified phenomenon” (2008: 120). Of course, they know that the term ‘Papuan languages’ does not identify a genetic group but is rather a convenient term for the non-Austronesian languages. In fact, they themselves identify “nine distinct families of Papuan languages in East Nusantara,” and this is, in their own words, “a conservative estimate” (2008: 110). How does one account for the presence of the alienable–inalienable contrast in a number of Papuan language families? One possibility, of course, is independent parallel developments. However, given the number of parallel developments one would have to assume, this scenario is unlikely to have been the case. Another possibility is contact. That is, it is highly likely that the alienable–inalienable distinction was at first present in one Papuan group and then spread, by contact, to the other Papuan groups. Under this scenario, the presence of the alienable–inalienable distinction in the Papuan languages is a unified phenomenon only as a later development, via diffusion, not unlike its presumed appearance in Oceanic and in the non-Oceanic Austronesian languages. And, if the alienable–inalienable contrast was originally present in only one Papuan group, how do we know that it did not acquire it by contact with one or more Austronesian languages? That is, assuming that Austronesian–Papuan contact was a factor in the presence of the contrast in those languages, how do we know what the direction of the spread was? Its presence in a number of Papuan families today does not necessarily tell us anything about the direction of the diffusion of the contrast several millennia ago.

The Papuan languages make only a binary alienable–inalienable distinction, without further distinctions in the alienable category. On the other hand, most Oceanic languages and some of the non-Oceanic Austronesian language do make distinctions in the alienable category (see Table 8 in §4). Some distinctions in the alienable category existed already in the common ancestor of the Oceanic and the South Halmahera–West New Guinea languages. While it is conceivable that the existence of the alienable–inalienable contrast in Proto–Eastern Malayo-Polynesian was due to contact with one or more Papuan language (but see §4), the existence of further contrasts in alienable possession is clearly a development internal to Austronesian. This development took place at least twice, once in Proto–Eastern Malayo-Polynesian and once in Selaru (see fn. 10).

(on possessum nouns and on the possessive classifiers). If prefixal possessor marking is indeed due to Papuan influence, this influence would have operated AFTER speakers of pre-POC left the East Nusantara area.

Some of the South Halmahera-West New Guinea languages have only a binary alienable–inalienable contrast, without any distinction in the alienable category. A referee suggests that this kind of reduction may be due to contact with Papuan languages, which do not make distinctions in alienable possession. While this scenario is not implausible, loss of possessive classifiers has also taken place in the Oceanic group. For example, as discussed in §3.2, Toqabaqita has only a binary contrast, without any possessive classifiers. The language is spoken in southeastern Solomon Islands, and there is no evidence of Papuan languages ever having been spoken there. That is, contact with Papuan languages is not a necessary condition for the reduction of a possessive system to a simpler binary one.

There are two basic ways in which a language can be affected by contact with another language: borrowing and what Thomason & Kaufman (1988) call ‘shift-induced interference’. According to them, structural borrowing only takes place after relatively heavy lexical borrowing. On the other hand, structural change due to shift need not be preceded by heavy lexical borrowing. We now have a sizeable list of lexical reconstructions for POc. Pawley (2007: 26) gives an estimate of the number of lexical reconstructions for POc. The figure he gives is 1500–2000 items, not an inconsiderable number. However, there is no evidence of heavy Papuan lexical influence on POc or on pre-POc. This suggests that if the presence of the alienable–inalienable possessive contrast in POc and further back in Proto-Eastern Malayo-Polynesian was indeed due to Papuan influence, the relevant factor was not borrowing but shift. This in turn means that some speakers of Proto-Eastern Malayo-Polynesian were genetically Papuan rather than Austronesian. And the fact that they were able to structurally affect Proto-Eastern Malayo-Polynesian suggests that their numbers were not negligible. That is, there was a Papuan genetic imprint already in the society or societies speaking Proto-Eastern Malayo-Polynesian. And indeed there is evidence of genetic flow between Papuan-speaking and Austronesian-speaking populations (Lum & Cann 1998; Lum et al. 2002; Hunley et al. 2008).

6. Summary and conclusions

POc had the following system of possessive constructions: a basic binary distinction between direct and indirect constructions, with further subcategories within the latter, distinguished by possessive classifiers, which more closely identified the nature of the possessum–possessor relation: (i) possessum as food for possessor; (ii) possessum as drink for possessor; (iii) and other kinds possessum–possessor relations (general possession). In some Oceanic languages the tripartite indirect system has been reduced, and in some others it has been expanded, in some of those cases considerably.

The etymology of the possessive classifiers is not always transparent. The POc food classifier was apparently related to the transitive verb ‘eat’, possibly via a nominalization (‘my taro, as eating/food’). Whether the (ultimate) source was a noun or a verb, a new grammatical element, a possessive classifier, developed by grammaticalization. Lynch et al. (2002) suggest the co-existence in POc of a construction with a food classifier and a construction with a nominalization of the related verb ‘eat’; see (46a) and (46b), respectively:

- (46) a. *a ka-gu pudu
 ART POSS.CLF(FOOD)-1SG.POSS banana
 “my banana (for eating)”
- b. *a kan-a-gu pudu
 ART eat-NMLZ-1SG.POSS banana
 “my food, a banana” (Lynch et al. 2002: 78)

The POc drink classifier may have been related in a parallel manner to the verb ‘drink’, although here the evidence is weaker. The third POc classifier, the general one, may have been historically connected to the common article and ultimately to a demonstrative. In the languages with highly developed systems of possessive classifiers, specifically some of the Micronesian languages, some classifiers are transparently related to nouns. And some of those nouns existed already in POc as nouns, but not as classifiers. This is evidence that in such cases the emergence of the classifier function was the outcome of grammaticalization processes, from a noun to a classifier.

The indirect constructions, with classifiers, typically express alienable possession, while the direct construction typically expresses inalienable possession. With the exception of some Micronesian languages where there are classifiers for some kinship relations, there are no classifiers for subtypes of inalienable possession (typically kinship relations and part-whole relations). This is due to the fact that in inalienable possession there is normally one strongly dominant kind of possessum–possessor relation and there is, therefore, no need to specify it by means of a classifier. On the other hand, in alienable possession there is often no such strongly dominant possessum–possessor relation (there is more scope for fluidity), and so the development of classifiers is motivated (though, obviously, not necessary) there (Lichtenberk et al. 2011; Lichtenberk 2013).

Systems of possessive constructions with possessive classifiers are also found in some languages that belong in the South Halmahera–West New Guinea group, a sister group of Oceanic, and so a system of possessive classifiers can be assumed to have existed in their common ancestor, Proto-Eastern Malayo-Polynesian.

While it is possible, as Klamer et al. (2008) and Donohue & Schapper (2008) argue, that the existence of the multiple possessive systems in Austronesian is

ultimately due to contact with Papuan languages, the evidence is not conclusive. Various Papuan languages make a distinction between alienable and inalienable possession, but there is considerable heterogeneity in the formal aspects of that distinction. At present it is not known how the alienable–inalienable contrast developed in the Papuan languages that belong in a number of separate groups. Presumably, contact was involved there as well. In any case, to argue for the existence of Papuan effect on Austronesian in the East Nusantara area, one needs to reconstruct a possessive system with an alienable–inalienable contrast to a time before speakers of pre-POc left East Nusantara.

One reviewer suggests that Ross's (2007) notions of grammatical calquing and metatypy are relevant in the case of Papuan–Austronesian case discussed here. By 'metatypy' Ross (2007: 124) means

the diachronic process whereby the morphosyntactic constructions of one of the languages of a bilingual speech community are restructured on the model of the constructions of the speakers' other language, such that the constructions of the replica language come to more closely match those of the model language in both meaning and morphosyntax.

Metatypy involves radical, wholesale restructuring of the target language. It is preceded by lexical and grammatical calquing. I have argued (§5) that if there was a transfer of Papuan possessive patterns into Austronesian, the mechanism involved was shift-induced interference in the sense of Thomason & Kaufman (1988), in which case we are not dealing with metatypy. As Ross (2007: 131) puts it, "[t]he effects of shift are thus very different from those of metatypy." Furthermore, in metatypy, according to Ross, there is radical, wholesale restructuring of one language on the model of another. However, if one accepts Klamer et al.'s (2008) arguments of East Nusantara being a linguistic area (see §5), it is difficult to see how metatypy could be involved. Klamer et al. (2008: 95) identify five areal morphosyntactic features: (i) "possessum–possessor ordering"; (ii) "overt marking of the distinction alienable vs. inalienable possession"; (iii) "clause-final negation"; (iv) "SVO as a primary constituent order"; and (v) "an inclusive/exclusive opposition in the pronominal paradigm." There has been diffusion in both directions: (i), (ii) and (iii) are "Papuan features that have diffused into the Austronesian languages," while (iv) and (v) are "Austronesian features [that] have diffused into the Papuan languages" (Klamer et al. 2008: 95). Such bidirectional diffusion does not fit Ross's notion of metatypy.

The existence of the alienable–inalienable contrast in the Papuan languages and in the Austronesian languages is indeed likely due to contact. However, it was only the semantic/pragmatic aspects of the contrast that spread via contact, not the formal aspects; that is, there was pattern transfer without matter transfer (Sakel 2007).

Acknowledgments

I am grateful to three anonymous referees for insightful and helpful comments on an earlier version of this paper. The usual disclaimers apply.

References

- Benton, Richard A. 1968. "Numeral and Attributive Classifiers in Trukese". *Oceanic Linguistics* 7.104–146. doi:10.2307/3622896
- Blust, Robert A. 1977. "A Rediscovered Austronesian Comparative Paradigm". *Oceanic Linguistics* 16.1–51. doi:10.2307/3622727
- Blust, Robert A. 1983/1984. "More on the Position of the Languages of Eastern Indonesia". *Oceanic Linguistics* 22/23.1–28.
- Blust, Robert A. 2005. "Review of *The Oceanic Languages* by John Lynch, Malcom Ross and Terry Crowley". *Oceanic Linguistics* 44.544–558. doi:10.1353/ol.2005.0030
- Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.
- Brownie, John & Marjo Brownie. 2007. *Mussau Grammar Essentials*. Ukarumpa, Papua New Guinea: SIL-PNG Academic Publications.
- Buse, Jasper (with Raututi Taringa). 1996. *Cook Islands Maori Dictionary with English-Cook Islands Maori Finderlist* ed. by Bruce Biggs & Rangi Moeka'a. Canberra: Pacific Linguistics.
- Clark, Ross. 2000. "Possessive Markers in Polynesian Languages". *Possessive Markers in Central Pacific Languages. Special issue of Sprachtypologie und Universalienforschung, Language Typology and Universals* ed. by Steven R. Fischer, 258–268.
- Crowley, Terry. 1985. "Common Noun Phrase Marking in Proto-Oceanic". *Oceanic Linguistics* 24.135–193. doi:10.2307/3623065
- Crowley, Terry. 2002. "Southeast Ambrym". *The Oceanic Languages* ed. by John Lynch, Malcolm Ross & Terry Crowley, 660–670. Richmond, Surrey: Curzon Press.
- Donohue, Mark. 2002. "Tobati". *The Oceanic Languages* ed. by John Lynch, Malcolm Ross & Terry Crowley, 186–203. Richmond, Surrey: Curzon Press.
- Donohue, Mark & Antoinette Schapper. 2008. "Whence the Austronesian Indirect Possession Construction?" *Oceanic Linguistics* 47.316–327. doi:10.1353/ol.0.0017
- Fischer, Steven R., ed. 2000. *Possessive Markers in Central Pacific Languages*. Special issue of Sprachtypologie und Universalienforschung, Language Typology and Universals.
- François, Alexander. 2002. *Araki: A disappearing language of Vanuatu*. Canberra: Pacific Linguistics.
- Greenberg, Joseph H. 1978. "How Does a Language Acquire Gender Markers?" *Universals of Human Language*, Vol. III: Word structure ed. by Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik, 47–82. Stanford: Stanford University Press.
- Hunley, Keith, Michael Dunn, Eva Lindström, Ger Reesink, Angella Terrill, Meghan E. Healy, George Koki, Françoise R. Friedlaender & Jonathan S. Friedlaender. 2008. "Genetic and Linguistic Coevolution in Northern Island Melanesia". *PLoS Genetics* 4(10).e1000239. doi:10.1371/journal.pgen.1000239.
- Hyslop, Catriona. 2001. *The Lolovoli Dialect of the North-East Ambae Language, Vanuatu*. Canberra: Pacific Linguistics.

- Jauncey, Dorothy. 2002. "Tamabo". *The Oceanic Languages* ed. by John Lynch, Malcolm Ross & Terry Crowley, 608–625. Richmond, Surrey: Curzon Press.
- Klamer, Marian, Ger Reesink & Miriam van Staden. 2008. "East Nusantara as a Linguistic Area". *From Linguistic Areas to Areal Linguistics* ed. by Pieter Muysken, 95–149. Amsterdam: John Benjamins. doi:10.1075/slcs.90.03kla
- La Fontinelle, Jacqueline de. 1976. *La langue de Houailou (Nouvelle-Calédonie): Description phonologique et description syntaxique*. Paris: Société d'Études Linguistiques et Anthropologiques de France.
- Laidig, Wyn D. 1993. "Insights from Larike Possessive Constructions". *Oceanic Linguistics* 32.311–351. doi:10.2307/3623197
- Lee, Kee-dong (with Lyndon Cornelius & Elmer Asher). 1975. *Kusaiean Reference Grammar*. Honolulu: University Press of Hawai'i.
- Lichtenberk, Frantisek. 1983a. *A Grammar of Manam*. Honolulu: University of Hawai'i Press.
- Lichtenberk, Frantisek. 1983b. "Relational Classifiers". *Lingua* 60.147–176. doi:10.1016/0024-3841(83)90072-4
- Lichtenberk, Frantisek. 1985. "Possessive Constructions in Oceanic Languages and in Proto-Oceanic". *Austronesian Linguistics at the 15th Pacific Science Congress* ed. by Andrew Pawley & Lois Carrington, 93–140. Canberra: Pacific Linguistics.
- Lichtenberk, Frantisek. 2008. *A Grammar of Toqabaqita*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110199062
- Lichtenberk, Frantisek. 2009a. "Oceanic Possessive Classifiers". *Oceanic Linguistics* 48.379–402. doi:10.1353/ol.o.0054
- Lichtenberk, Frantisek. 2009b. "Attributive Possessive Constructions in Oceanic". *The Expression of Possession* ed. by William B. McGregor, 249–292. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110213232.249
- Lichtenberk, Frantisek. 2013. "The Rise and Demise of Possessive Classifiers in Austronesian". *Historical Linguistics 2011: Selected papers from the 20th International Conference on Historical Linguistics, Osaka, 25–30 July 2011* ed. by Ritsuko Kikusawa & Lawrence A. Reid, 199–225. Amsterdam: John Benjamins.
- Lichtenberk, Frantisek, Jyotsna Vaid & Hsin-Chin Chen. 2011. "On the Interpretation of Alienable vs. Inalienable Possession: A psycholinguistic investigation". *Cognitive Linguistics* 22.659–689. doi:10.1515/cogl.2011.025
- Lum, J. Koji & Rebecca L. Cann. 1998. "mtDNA and Language Support a Common Origin of Micronesians and Polynesians in Island Southeast Asia". *American Journal of Physical Anthropology* 105.109–119. doi:10.1002/(SICI)1096-8644(199802)105:2<109::AID-AJPA1>3.0.CO;2-Q
- Lum, J. Koji, Lynn B. Jorde & Wulf Schiefenhovel. 2002. "Affinities among Melanesians, Micronesians, and Polynesians: A neutral, biparental genetic perspective". *Human Biology* 74.413–430. doi:10.1353/hub.2002.0031
- Lynch, John. 1973. "Verbal Aspects of Possession in Melanesian Languages". *Oceanic Linguistics* 12.69–102. doi:10.2307/3622853
- Lynch, John. 2000a. *A Grammar of Anejom̃*. Canberra: Pacific Linguistics.
- Lynch, John. 2000b. "Historical Overview of Central Pacific Possessive Markers". *Possessive Markers in Central Pacific Languages*. Special issue of Sprachtypologie und Universalienforschung, Language Typology and Universals ed. by Steven R. Fischer, 233–242.

- Lynch, John. 2001. "Passive and Food Possession in Oceanic Languages". *The Boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* ed. by Andrew Pawley, Malcolm Ross & Darrell Tryon, 193–214. Canberra: Pacific Linguistics.
- Lynch, John, Malcolm Ross & Terry Crowley. 2002. *The Oceanic Languages*. Richmond, Surrey: Curzon Press.
- Milner, George B. 1972. *Fijian Grammar* (3rd edn.). Suva: Government Press.
- Mulder de, Walter & Anne Carlier. 2011. The Grammaticalization of Definite Articles. *The Oxford Handbook of Grammaticalization* ed. by Heiko Narrog & Bernd Heine, 522–534. Oxford: Oxford University Press.
- Osmond, Meredith. 1998. "Horticultural practices". *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society, Vol. I: Material culture* ed. by Malcolm Ross, Andrew Pawley & Meredith Osmond, 115–142. Canberra: Pacific Linguistics.
- Ozanne-Rivierre, Françoise. 1976. *Le iaai: Langue mélanésienne d'Ouvéa (Nouvelle-Calédonie). Phonologie, morphologie, esquisse syntaxique*. Paris: Société d'Études Linguistiques et Anthropologiques de France.
- Palmer, Bill & Dunstan Brown. 2007. "Heads in Oceanic Indirect Possession". *Oceanic Linguistics* 46.199–209. doi:10.1353/ol.2007.0022
- Pawley, Andrew. 1973. "Some Problems in Proto-Oceanic Grammar". *Oceanic Linguistics* 12.103–188. doi:10.2307/3622854
- Pawley, Andrew. 2005. "The Meaning(s) of Proto Oceanic *panua". *A Polymath Anthropologist: Essays in honour of Ann Chowning* ed. by Claudia Gross, Harriet D. Lyons & Dorothy A. Counts, 211–223 (= Research in Anthropology and Linguistics, 6). Department of Anthropology, University of Auckland.
- Pawley, Andrew. 2007. "The Origins of Early Lapita Culture: The testimony of historical linguistics". *Oceanic Explorations: Lapita and western Pacific settlement* ed. by Stuart Bedford, Christophe Sand & Sean P. Connaughton, 17–49. Canberra: ANU ePress.
- Pawley, Andrew & Timoci Sayaba. 1990. "Possessive-Marking in Wayan, a Western Fijian Language: Noun class or relational system?" *Pacific Island Languages: Essays in Honour of G. B. Milner* ed. by Jeremy H. C. S. Davidson, 147–171. London: School of Oriental and African Studies, University of London; and Honolulu: University of Hawai'i Press.
- Rehg, Kenneth L. (with Damian G. Sohl). 1981. *Ponapean Reference Grammar*. Honolulu: University Press of Hawai'i.
- Ross, Malcolm. 2002. "Mussau". *The Oceanic Languages* ed. by John Lynch, Malcolm Ross & Terry Crowley, 148–166. Richmond, Surrey: Curzon Press.
- Ross, Malcolm. 2007. "Calquing and Metatypy". *Journal of Language Contact – THEMATA* 1.116–143. doi:10.1163/000000007792548341
- Ross, Malcolm, Andrew Pawley & Meredith Osmond, eds. 2011. *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society, Vol. IV: Animals*. Canberra: Pacific Linguistics.
- Sakel, Jeanette. 2007. "Types of Loan: Matter and pattern". *Grammatical Borrowing in Cross-linguistic Perspective* ed. by Yaron Matras & Jeanette Sakel, 15–29. Berlin: Mouton de Gruyter.
- Schütz, Albert J. 1985. *The Fijian Language*. Honolulu: University of Hawai'i Press.
- Sohn, Ho-min & Bryon W. Bender. 1973. *A Ulithian Grammar*. Canberra: Pacific Linguistics.
- Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.

- van den Berg, René. 2009. "Possession in South Halmahera–West New Guinea: Typology and reconstruction". *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust* ed. by Alexander Adelaar & Andrew Pawley, 327–357. Canberra: Pacific Linguistics.
- Williams, Herbert W. 1971. *A Dictionary of the Maori Language* (7th edn., revised and augmented by the Advisory Committee on the Teaching of the Maori Language, Department of Education). Wellington: Government Printer.
- Wilson, William H. 1982. *Proto-Polynesian Possessive Marking*. Canberra: Pacific Linguistics.

Development and diffusion of classifier systems in Southwestern Amazonia

Hein van der Voort
Museu Paraense Emílio Goeldi

Comparative research shows that many genealogically unrelated languages in the Guaporé-Mamoré region in the Southwestern Amazon share specific lexical and structural traits, which suggests that the region forms a linguistic area. One of these traits concerns classifier systems. Classifier systems in genealogically diverse Northwestern Amazonian languages display similar structural patterns that involve large sets of bound classifiers with wide morphosyntactic distributions, which represents an areal trait. I show that classifier systems in Southwestern Amazonian languages also share structural and formal characteristics among one another. In addition, important structural properties of these systems are similar to those that characterize Northwestern Amazonian systems. It is likely that certain structures and forms have spread through diffusion, although it may not be possible to determine their precise origin. Apparently, borrowed classifiers and calqued classifying structures have undergone further development and extension in the individual languages.

1. Introduction

The region of the Guaporé and Mamoré river basins, in the Southwestern Amazonian lowlands, which cover parts of Bolivia and Brazil, is home to an exceptional diversity of indigenous languages. The approximately 50 languages represent seven linguistic stocks, some unique to this region, as well as ten linguistic isolates. Most of the languages are highly endangered, but the past two decades have fortunately seen numerous initiatives to document and describe them. Comparative research drawing on such new data shows that many genealogically unrelated languages in the Guaporé-Mamoré region share specific lexical and structural traits, which suggests that the region forms a linguistic area (Crevels & van der Voort 2008). One of these traits concerns classifier systems.

Classifier systems are attested in a many South American languages, and tend to have certain characteristics in common that are not attested in the same manner and constellation as in languages spoken in other parts of the world. Since the mid-1980s, the manifestations of nominal classification have been an issue of debate in Amazonian linguistics. Payne (1986), Derbyshire & Payne (1990), Grinevald (2000), Grinevald & Seifart (2004), Seifart (2005) and Seifart & Payne (2007), among others, have discussed characteristics of especially (north)western Amazonian systems of nominal classification in the light of the differences and similarities with nominal classification systems elsewhere.

Amazonian classifier systems tend to be of a heterogeneous nature. Where African languages are known for closed and strongly grammaticalized inflectional noun class and gender agreement systems, and where Asiatic and Austronesian languages typically have rather open systems of a more lexical nature that involve measure terms and class terms, Amazonian languages may employ intermediate lexico-grammatical systems of nominal classifiers that combine characteristics of both extremes. However, as Seifart & Payne (2007: 384) point out, the Northwestern Amazonian classifier systems should be regarded as “instantiating a coherent system type in their own right” rather than as particularly ‘mixed’ or ‘multiple’ systems. As will become clear, this characterization is not limited to the languages of the Northwestern Amazon, but it also holds for several Southwestern Amazonian languages. Krasnoukhova (2012: 204) calls these ‘multifunctional’ classifier systems.

Classifier systems in Amazonian languages may involve a great number of bound but semantically specific forms (sometimes over a hundred) that express e.g. shape, texture, function, body part and sex. In some languages a generic classifier can also occur in the place of a specific classifier when the discourse context allows, as a kind of grammatical placeholder (Grinevald, personal communication). Conversely, certain classifiers can be highly specific, defining a class with only a few members or even one. The distribution of classifiers is typically not limited to numeral or possessive constructions but may also include attributive, demonstrative, nominal and verbal environments. Apart from producing new compound-like expressions, Amazonian classifiers can also have agreement-like uses and can function as discourse anaphors. It has been suggested (Grinevald 2000, 2007; Seifart & Payne 2007), that Amazonian classifier systems may represent an intermediate stage on a grammaticalization path from compounding and incorporation of lexical forms to the development of strictly grammatical morphemes.

Table 1 is a summary of the typology of nominal classification systems and corresponding terminology, as proposed by Grinevald (2000), organized in a horizontal ‘cline of grammaticalization’.

Table 1. Systems of nominal classification

Lexical	Intermediate	Grammatical
measure terms	numeral classifiers	gender markers
class terms	sortal classifiers	noun class markers
	mensural classifiers	
	noun classifiers	
	verbal classifiers	

In the table, lexico-grammatical nominal classifiers (which in many Amazonian languages can also occur on verbs) are located in an intermediate position between the lexical extreme of class terms and the grammatical extreme of noun classes. The table is not meant to suggest that there is just one intermediate position; rather, there is a whole range of intermediate possibilities between the prototypical extremes of lexical and grammatical systems.

In fact, as Derbyshire & Payne (1990) show, a considerable variety of classifier systems is attested among Amazonian languages. Many have relatively open lexical-like systems. In a few isolating languages, these systems may resemble the Asiatic type, such as the Yanomami languages, e.g. Sanuma (Borgman 1990), Yanomami (Ramirez 1994) and Yanomam (Goodwin Gómez 2000; Perri Ferreira 2009), and certain Tupian languages, e.g. Munduruku (Gonçalves 1987; Picanço 2003). However, in many Amazonian languages such 'open' classifying systems often involve bound morphemes, as in Miraña (Seifart 2005), and several languages apparently have incipient systems, as argued for Hup (Epps 2007).

Seifart & Payne (2007: 384) emphasize that the majority of the classifier systems in Northwestern Amazonia display similar structural patterns involving large sets of bound classifiers with a wide morphosyntactic distribution, and they call attention to the fact that these similarities occur in genealogically unrelated languages, representing a widespread areal trait. Similar statements hold for Southwestern Amazonia, which is suggestive of a possible connection between these areas.

In this article, classifier systems in the Guaporé-Mamoré languages are presented. In Crevels & van der Voort (2008: 167, 171) classifier systems are reported to be relatively rare among the Guaporé-Mamoré languages, and of the 24 languages of their sample only eight are listed as having classifiers (2008: 171). The actual morphosyntactic pattern may be somewhat more widespread in the region than suggested. For present purposes, establishing a hermetic definition of classifiers and classifying systems is not a main concern. Rather, the goal is to identify similar morphosyntactic structures, and their shared semantic properties, across genetic linguistic boundaries, which could indicate areal diffusion. Therefore, phenomena

from certain languages that are regarded by specialists as not having classifiers are nevertheless included in the present discussion.¹

In §2, I present formal and semantic properties of classifier or classifier-like systems in a number of Southwestern Amazonian languages, in particular Kwaza (ISOLATE), Kanoê (ISOLATE), Aikanã (ISOLATE), Movima (ISOLATE), Itonama (ISOLATE), Lakondê (NAMBIKWARAN), Cavineña (TACANAN), Arikapu (MACRO-JÊAN), Baure (ARAWAKAN) and Karo (TUPIAN), in that order. In §3, I present distributional properties of the classifier systems introduced in §2, in the following order of functions and environments: nominal, verbal, nominalization, relationships with nouns, generic function, other constructions. In §4, I present the empty morpheme construction, which represents a striking areal linguistic trait. In §5, I summarize and discuss similarities between the classifier systems, and in §6 I argue that certain similarities emerged through areal diffusion. Since the similarities with classifier systems in the Northwestern Amazon are complemented by other structural and lexical similarities, I also suggest the possibility of an areal linguistic connection between the Northwestern and Southwestern Amazon.

In this article, ‘incorporation’ is used in an informal sense for derivation of a verbal stem by a classifier or noun and does not necessarily imply changes in syntactic relations. Only in Movima does the incorporation of classificatory elements often (but not always) involve changes in valence and case relations, as would be required by narrower definitions of incorporation (e.g. Mithun 1984). Furthermore, in addition to an open class of independent nouns, certain languages have closed classes of bound nouns. In some languages these bound nouns exist in addition to a classifier system and can also be used as classifiers themselves. Some subsets of bound nouns may define single-item classes and can therefore be regarded as ‘repeaters’.² The properties of all these potentially classificatory subsystems differ greatly according to the particular languages, and the descriptive sources differ in their criteria for distinguishing them. The present article therefore also serves the purpose of giving an overview of the facts. A condensed version of some parts of this article was published in Portuguese (van der Voort 2015).

1. The question of what can be regarded as classification systems in Amazonian languages is far from settled (see, e.g., Grinevald & Haude 2013).

2. A ‘repeater’ can be full or partial. It functions as a classifier and is identical to (part of) the noun it classifies, or better, repeats. Unlike classifiers, full repeaters also occur as nouns. Furthermore, a repeater represents only one referent, not a class of referents. Unlike classifiers, repeaters have not undergone semantic bleaching with respect to the noun they relate to. They nevertheless form a category separate from nouns since not all nouns can occur as repeaters (see also notes 5 and 9).

The data from all languages discussed are from recent fieldwork by various colleagues. The origin of each numbered example is indicated, except for those from my own ongoing research on Aikanã, Arikapu and Kwaza. The examples are in a largely phonemic spelling and symbols represent IPA values, with the following exceptions: <ü> = [y] (Aikanã); <y> = [i] ~ [ɨ] (Kwaza, Kanoê) and [j] (Itonama, Movima, Baure, Karo); <j> = [h] (Cavineña, Movima); <x> = [ʂ] (Kwaza) and [ʃ] (elsewhere); <š> = [ʃ] (Baure); <č> = [tʃ] (Baure); <ch> = [tʃ] (Movima).

2. Form and meaning of classifiers

In this section, formal, semantic and etymological aspects of classifier systems in the languages of the Southwestern Amazon are discussed. The languages that have such systems are presented in an order that is determined by their genetic and geographic position, in the following way:

Table 2. Southwestern Amazonian languages with classifier systems

1	Kwaza	ISOLATES	Brazil
2	Kanoê		
3	Aikanã		
4	Movima		Bolivia
5	Itonama		
6	Lakondê	NAMBIKWARAN	Brazil
7	Cavineña	TACANAN	Bolivia
8	Arikapu	JABUTIAN, MACRO-JÊAN	Brazil
9	Baure	SOUTHERN ARAWAKAN	Bolivia
10	Karo	RAMARAMA, TUPIAN	Brazil

2.1 Kwaza (ISOLATE)

Kwaza (also known as Koaiá) is a morphologically complex language isolate spoken in southeastern Rondônia, Brazil, by around 25 remaining speakers. The Kwaza classifier system has been described in van der Voort (2004). The system possesses many properties that are characteristic of Amazonian languages. Kwaza has numerous classifiers (almost 150 suffixes), many of which have a rather specific semantic content that other languages would express by independent lexemes. These suffixes can be called classifiers, because most of them represent or define a semantically motivated set of multiple referents in the real world. To give an example: the classifier *-ri* refers to flat objects or beings, including discs, paddles, plates, fields and

livers. It also occurs in fixed combination with other productive classifiers,³ such as *-tsu* “stick” in *-ritsu* “flat thin wood”, *-xu* “bone” in *-xuri* “chest”, *-lo* “tube” in *-rilo* “jawbone”, or with unidentifiable elements, as in *-rihi* “flat seed”. It is probably fossilized in *kēwēdori*, the disc-shaped freshwater turtle. Note that composite classifiers have a single function and are therefore different from the stacking of multiple classifiers in one word.

In addition to classifiers that are semantically relatively abstract and that are associated with a wide range of referents, there are also classifiers with more restricted semantics, referring to body parts and some metaphorical extensions, such as *-koro* “arm, branch”, *-tōi* “eye, fruit, star”, *-kuty* “head, calabash”. Finally, there are classifiers which were found to be associated with a single referent only, such as *-tehu* “spoon”, *-nāko* “bee” and several body part classifiers like *-tuku* “tongue”, *-kai* “leg”, *-romū* “thigh”, etc.

In terms of etymological transparency, Kwaza classifiers can be divided into three sets: set I classifiers show a certain resemblance to specific nouns and are probably etymologically related to them; set II classifiers are identical with certain specific nouns (or with parts of nouns); set III classifiers are not known to be etymologically related to any specific independent nouns. The classifiers of this latter set share the possibility of attaching to the semantically empty root *e-* to form independent nouns themselves, carrying the meaning of the classifier (see §4.1).

The Kwaza classifiers are not highly grammaticalized noun class markers, because there is no obligatory grammatical relationship between a particular classifier and a particular set of referents. Rather, the classifiers usually describe their referents on the basis of particular physical characteristics singled out for the discourse occasion, as if by way of meronymy (Klein 2000). Then again, they are not highly lexical class terms either, in the first place because of their bound status, in the second place because of their often relatively comprehensive semantics (with the exception of some rather specific classifiers) and, finally, because they fill an obligatory slot in several types of grammatical constructions.

The grammatical distribution of Kwaza classifiers includes nouns, adverbs and verbs, as well as possessive, numeral, attributive and demonstrative constructions. They can be attached as nominalizers to verbs and adverbs, and they may be incorporated into verbs and refer to one of the arguments. Every classifier can in principle occur in every construction. One of the nominalising morphemes of Kwaza, the suffix *-hŷ*, also functions as a generic classifier. It is used whenever a specific classifier is not chosen, filling a morphological slot that is obligatory in possessive and in most attributive constructions. In §3, the distributional characteristics of

3. Around 30% of the classifiers in Kwaza are composite, i.e. consisting of multiple classifiers. As shown below, composite classifiers occur in various languages of the region.

the Kwaza classifier system are discussed, illustrated and compared with those of the other classifier languages.

2.2 Kanoê (ISOLATE)

Kanoê (sometimes also *Kapixana*) is a language isolate spoken in southeastern Rondônia, Brazil, by four remaining speakers. It is a morphologically very complex language, especially with regard to the verbs. There is both morphological and syntactic nominal compounding in Kanoê, although the latter is more productive. There does not seem to be a separate class of adjectives.

The Kanoê classifier system has been described in Bacelar (2004). There are approximately 20 genuine classifiers. They tend to have general descriptive semantics, referring to the shape and texture of things, such as *-ko* “shell, round, shoulder”, *-mu* “liquid”, *-ty* “vine, thread”. Some of the classifiers can be related to existing free nouns: *-kaña* “flour, porridge” ~ *kaña* “flour”; *-tinu* “soup, mud” ~ *tinu* “soup”.

In addition, there is a set of about 80 ‘classifiers’ that can be turned into free nouns by attaching them to the semantically empty noun formative root or prefix *i-*. The vast majority of these elements refer to body or plant parts, such as *-taw* “tongue”, *-ry* “liver”, *-ati* “flower”. About half are composite, such as *-kÿjmu* “tear” < *-kÿj* “eye” plus *-mu* “liquid”, *-kyetæ* “shell” < *-tæ* “concave, round”.

To whichever set they belong, Kanoê classifiers have a relatively wide distribution.

2.3 Aikanã (ISOLATE)

Aikanã (sometimes also *Huari*, *Masaká* or other) is a language isolate spoken in southeastern Rondônia, Brazil, by approximately 250 speakers. It is a morphologically very complex language. The classifier system was touched upon by Vasconcelos (2003) and van der Voort (2005). Here, a more extensive analysis is presented, based on data from recent fieldwork.

The Aikanã classifier system resembles that of other Amazonian languages with respect to its morphosyntactic distribution and the semantic contents of the forms. There are two formal subtypes of classifiers in Aikanã: derivational suffixes and bound body part roots. In addition, free nouns can have a classificatory function in compounds.

There are over 20 bound classifying morphemes, which tend to have general descriptive semantics, referring to the shape and texture of things, such as *-jú* “hollow, concave”, *-mũ* “liquid”. Some classifiers can be related to existing free nouns, such as *-di* ~ *ji* “hair, leaf, feather”.

In addition to these classifiers, there is a somewhat bigger set of elements that often behave as classifiers, and which mostly refer to body part terms, such as *-mūka* “eye”, *-nawā* “nose”, *-nenu* “lip”. They do not seem to be derived from existing free nouns like some from the smaller set of classifiers. Instead they can be turned into free nouns when preceded by the semantically empty element *ka-*, discussed in §4.3. Only very few classifiers in Aikanā are composite.

2.4 Movima (ISOLATE)

Movima is a language isolate spoken in the Beni department, Bolivia, by approximately 1500 speakers. It is a highly morphologically complex language that has been described in detail by Haude (2006). The classificatory system was discussed both by Haude and by Grinevald (2002).

Haude (2006) does not distinguish a dedicated category of classifiers, but she discusses classification as one of the functions of constructions involving bound nominal morphemes. In addition to monomorphemic free nouns, Movima has three different types of bound nominal morphemes. The first type are the approximately 80 bound noun roots (BR) for which no equivalent free nouns exist. The second type are the approximately 40 truncated nominal elements (TRC) that originate as parts of equivalent free nouns. They usually derive from the last syllable of a free noun, such as *-do* < *cha:do* “plate”, *-jel* < *chi:jet* “ant”. The third type are the approximately 20 bound nominal elements (BE), which have no etymological relationship with equivalent free nouns. Bound elements never occur as the only root of a noun, which distinguishes them from bound noun roots. Furthermore, they do not derive from free nouns, which distinguishes them from truncated elements. But in compounds and incorporation they may represent etymologically unrelated free nouns, such as *-as* “sugar cane” vs. *chara:ye* “sugar cane”, *-di* “house, enclosed space” vs. *ro:ya* “house”.

Both nominal compounding and verbal noun incorporation are very productive in Movima and can involve full nouns or bound nominal morphemes. In compounds, the bound morphemes represent the head. In verbs they may represent an incorporated argument and have valence-changing properties.

Movima bound nominal morphemes may have properties that are similar to those of classifiers in other languages of the region, and they may occur in constructions that are highly reminiscent of classifying constructions in other languages.

2.5 Itonama (ISOLATE)

Itonama (or *Sihnipadara*) is a highly morphologically complex language isolate spoken in the Beni department, northeastern Bolivia, by two elderly speakers. It has been studied by Crevels (2002, 2006, 2012).

Itonama has multiple classifier systems: one set of 17 classifiers appearing exclusively on verbs and demonstratives and another set of 8 classifiers on numerals. These systems are strictly separate, even though the different classifiers from both systems may carry identical semantic content, such as ‘sinuous shape’, which is expressed by both the verbal classifier *-lo* and the numeral classifier *ni-*. The fact that the verbal classifiers sometimes function as actual verbs suggests their origin may be ultimately verbal.

There is no nominal compounding, but nouns can be incorporated in verbs. Many nouns are obligatorily possessed and require a possessive prefix when occurring in isolation. When incorporated, nouns appear in their bound, truncated form.

2.6 Lakondê (NORTHERN NAMBIKWARAN)

Lakondê is a Northern Nambikwaran language, spoken in the border region of Rondônia and Mato Grosso, Brazil. The three main branches of the Nambikwaran family are Southern Nambikwaran (approximately 900 speakers), Northern Nambikwaran (approximately 80 speakers) and Sabanê (3 speakers). The languages are phonologically and morphologically relatively complex. They share some areal traits with other languages of the region, including formal and distributional aspects of their classifier systems. The examples here are mainly from Lakondê.

Kroeker (2001) and Lowe (1999) describe the classifier systems of the Southern Nambikwaran dialect complex. Classifiers in Southern Nambikwaran either are attached to nouns or may occur by themselves as anaphors. The classifiers specify nouns semantically with regard to shape, substance and sex, and there is a generic classifier, *te*² ‘person or thing not specified’. The total set of classifiers listed by Kroeker and Lowe combined contains close to 25 forms.

In the Northern Nambikwaran dialect complex classifiers have similar functions and distributions, although there are fewer forms, which are somewhat different from the Southern Nambikwaran ones. Of the classifiers listed by Telles (2002: 187), who described the Lakondê and Latundê dialects (L), and by Eberhard (2009: 349–353), who described Mamaindê (M), many show similarities: (L) *-kah* ‘long, tridimensional’ ~ (M) *-k^hat?* ‘sticklike’, (L) *-ka’loh* ‘flat surface, bidimensional’ ~ (M) *-kalo* ‘flat’, etc. The fact that almost half of the classifiers are probable cognates with classifiers in Southern Nambikwaran suggests that Proto-Nambikwara already had

a classifier system. In addition to classifiers, the Northern Nambikwaran languages also have a limited set of body part nouns that are incorporated in verbs by prefixation (Telles 2002: 238–240; Eberhard 2009: 357, 397–403).

The third branch of the Nambikwaran linguistic family consists of the Sabanê language. Sabanê is shown by Araujo (2004) to have a much more limited classifier system when compared to the other Nambikwaran languages, which nevertheless shows considerable similarities. The distinction between compounding, which is highly productive in Sabanê, and classifier incorporation is not very clear.

2.7 Cavineña (TACANAN)

The Tacanan languages form a family of five languages spoken in the northern Bolivian lowlands. The issue of a genetic connection with the Pano language family has not been fully settled. Cavineña, the best described Tacanan language (Guillaume 2008), has around 600 speakers, and it is morphologically highly complex with prefixes, suffixes and clitics. Although certain constructions are reminiscent of classification, it may not be justified to say that the language has classifiers.

Guillaume (2008: 72) distinguishes three subcategories of nouns in Cavineña: the open set of independent nouns, a set of around 30 nominal kinship terms that require possessive inflection and a set of up to 150 bound noun roots that minimally require a semantically empty prefix *e-* in order to be able to occur as an independent syntactic constituent. This third category of ‘*e*-nouns’ includes 68 body part nouns, 11 plant part nouns, 18 landscape nouns, 12 spatial relationship nouns and 22 miscellaneous nouns, many of which refer to relatively basic concepts such as ‘fire’, ‘water’, ‘egg’. Guillaume (2008: 408) suggests they represent a noun class that can be recognized by its class marker *e-*, which may be a plausible view. This group of noun roots also appears to contain some composite bound noun roots, consisting of fixed combinations of *e*-nouns – or *e*-nouns with free nouns – such as *-wikani* ‘nose’ < *-wi* ‘beak’ and *kani* ‘nose’; *-nakaka* ‘juice, nectar, honey’ < *-na* ‘water’ and *-kaka* ‘fruit’, etc. (see also Guillaume 2008: 463).

Perhaps the Tacanan bound noun roots can be considered as an incipient classifier system. The distribution range of Tacanan bound noun roots is much more limited than that of body-part classifiers in other languages, which have nominalising, anaphoric and other derivational and inflectional properties generally associated with Amazonian classifier systems.

2.8 Arikapu (JABUTIAN, MACRO-JÊAN)

Together, Arikapu (1 speaker) and Djeoromitxi (approximately 35 speakers) form the Jabutian language family, spoken on the headwaters of the Branco River of Rondônia. Jabutian represents the westernmost branch of the Macro-Jêan language family (Ribeiro & van der Voort 2010). Arikapu is not a morphologically complex language.

Apart from common nouns, the language appears to have a number of obligatorily possessed nouns, some of which may behave and function as classifiers. These nouns usually refer to body parts (see §4.7). Furthermore, elements with more general semantic content, such as *-hã* “seed, kernel”, *-ka* “fruit, round”, *kə* “skin, shell, bark” strongly resemble nominal classifiers, and I gloss them as such (except for those that occur independently as well). They are attached to nouns or are preceded by person prefixes. In various respects these classifiers resemble those attested elsewhere in the Guaporé-Mamoré region. Some of the classifying forms even occur in almost identical form in unrelated languages (see §5.1).

Also the neighboring sister language Djeoromitxi has classifier-like elements. Some of these are analyzed by Pires (1992) and de Castro (2012) as the second part of a nominal compound. They were not attested as independent elements but appear obligatorily possessed (with the exception of *kə* “skin, shell” and *nô* “food”). The majority of the classifiers of Arikapu and Djeoromitxi are reconstructable in Proto-Jabuti, as are some in Proto-Jê. Consequently, if areal diffusion of certain classifiers took place, either (Proto-)Jabuti may have been a source or the diffusion took place relatively long ago.

The element *kakə* “shell of a nut, seed or animal” itself seems to represent a composite of two classifiers. It is possible that the classifiers originally represented compounded nouns. They may have been derived from full nouns and represent an intermediate stage in a grammaticalization process.

The classifier-like elements of Arikapu and Djeoromitxi constitute highly similar sets of sometimes rather unproductive morphemes that belong to a small closed category with a rather limited distribution. First person singular and plural marking prefixes, *i-* and *txi-* respectively, are frequently used as dummy elements, expressing an impersonal possessor both on full nouns and classifiers.

2.9 Baure (SOUTHERN ARAWAKAN)

The Baure language of Bolivia (approximately 60 speakers), which has been extensively described by Danielsen (2007), is morphologically complex. Like many other Arawakan languages, Baure has an elaborate classifier system.⁴ It has many of the distributional and semantic properties that are characteristic of Amazonian languages, such as a large set of bound classifying morphemes that can occur in a variety of morphosyntactic environments, e.g. attached to numerals and adjectives, incorporated into verbs and as part of compounds. In Baure, this set of morphemes includes about 40 grammaticalized classifiers, such as *-aro* “liquid”, *-iro* “round”. Furthermore, there are roughly 35 unique classifiers.⁵ As Danielsen (2007: 143) observes, the majority of unique classifiers are derived from bound noun roots, and some of them can be regarded as repeaters. They are usually truncated forms of the corresponding nouns.

In addition, nouns themselves can also have a classifying function. Danielsen (2007) distinguishes two different kinds of nouns in Baure: independent nouns and (approximately 60) obligatorily possessed nouns. Obligatorily possessed nouns are bound nominal roots that must either receive specific possessor inflection or a semantically empty prefix *e-* when the possessor is unknown. This prefix is analyzed by Danielsen (2007: 119) as a marker of unspecified possession. The majority of bound *e-*nouns in Baure refer to part of whole concepts such as body parts, e.g. *-čipi* “roof, back”, *-hi* “horn”, *-tokie* “head/brain”. The distinction between bound noun roots and classifiers in Baure is not complete.

The division between classifiers and bound nouns is semipermeable in the sense that classifiers cannot be turned into independent nouns by attachment to the *e-* prefix, whereas bound nouns can. Furthermore, classifiers tend to have more general semantic characteristics, whereas bound roots have more specific reference. Consequently, when attached to verb roots, classifiers have clearer anaphoric agreement properties and can co-occur with an overt argument noun, which further specifies the interpretation of the classifier. Bound nouns behave rather like incorporated arguments and an additional overt noun phrase is then rare and emphatic, as Danielsen (2007: 208ff.) shows.

4. See, among others, Aikhenvald (1994) for an extensive overview of classifiers in Arawakan languages.

5. Either partial or full ‘repeaters’ (see also notes 2 and 9).

2.10 Karo (RAMARAMA, TUPIAN)

Karo is the only remaining member of the Ramarama branch of the Tupian language family. It is spoken by the majority of the approximately 150 members of the Arara people of eastern Rondônia. It was described by Gabas (1999), who characterizes it typologically as a mildly synthetic and fusional language. Nominalization and compounding are among the few morphological processes.

The classifier system of Karo is small, but nevertheless distinct. The 11 classifiers are particles, most of which refer to shape properties: *ʔaʔ* “round”, *káʔ* “concave or convex”, *peʔ* “flat”. There is no semantically neutral, generic classifier. The classifiers occur only in a limited set of constructions – with nouns, in possessive constructions and in compounds – and they are not obligatory. As Gabas (1999: 228) observes, it has characteristics of both the lexical type (optionality, free morphemes) and the grammatical type (few items, concordance).

The Guaporé-Mamoré region holds the highest internal diversity of the Tupian language family and it is therefore regarded as the earliest center of dispersal of the family. However, Karo is one of possibly only three Tupian languages with a classifier system. The other two languages are Mundurucu and Kuruaya, which make up a separate branch of the Tupian family and which are spoken far to the northeast from Karo. Although they are argued to have a rich classifier system in some sources (Gonçalves 1987; Martines 2007; Picanço 2003), Gomes (2006) presents a different analysis, in which only a small number of nouns are used as classifiers. Other Tupian languages may have constructions involving nouns that are reminiscent of classification, such a certain set of unpossessible noun stems in Gavião (MONDÉ, TUPIAN), described in Moore (1984: 203–205). Certain constructions in the Tupari languages Mekens and Makurap remind one of possessive classification (see §4.3).

3. Distribution and functions of classifiers

The present section treats structural and functional characteristics of classifier systems in the languages of the Southwestern Amazon. The precise range of morphosyntactic constructions where classifiers can appear may differ for each language. Furthermore, classifier systems sometimes consist of different subsets that do not have the same distribution. Comparing languages with regard to the constructions that feature classifiers could give us a better understanding of the systems and of possible areally distributed patterns. The potential classifying constructions are presented in the following order: with nominal roots, with verbal roots, nominalization, interchangeability with nouns, generic function and other constructions.

3.1 Classifiers combined with nominal roots

The attachment of classifiers to nominal roots or stems is attested in all classifier languages of the region, with varying degrees of productivity. They can be used to derive new nouns, as in Kwaza (1), Kanoê (2), Aikanã (3), Lakondê (4) and Baure (5):

- (1) *jo-mũ*
manioc-CLF:liquid
“manioc beer”
- (2) *mutyry-ko-mu*
assai-CLF:shell-CLF:liquid
“chicha of assai” (Bacelar 2004:102)
- (3) *tüwi-nũ*
tobacco-CLF:powder
“crushed tobacco”
- (4) *‘wajh-ki’ñin-te*⁶
assai-CLF:round-REF
“assai palm nut” (Telles 2002: 185)
- (5) *yaki-so*
fire-CLF:stick
“firewood” (Danielsen 2007: 97)

Like classifiers in other languages, bound nominal morphemes in Movima can be combined with existing free nouns and adjectives as in the following:

- (6) *kape:-vas*
coffee-BR:flour
“coffee powder” (Haude 2006: 211)
- (7) *limo:-mi*
lemon-TRC:water
“lemonade” (Haude 2006: 215)
- (8) *tochi’-mo*
small-BE:bird
“chick” (Haude 2006: 220)

Arikapu has a small set of classificatory elements that are rarely attested by themselves. They are usually preceded by person prefixes or by full nouns, as in (9):

6. Independent nouns in isolation, including possessed and derived nouns usually end with the ‘referential’ suffix *-te* or *-tu* in Lakondê and with the ‘final nominal suffix’ *-tu* in Mamaindê.

- (9) *ataw-ka*
 rubber-CLF:fruit
 “rubber ball”

Karo classifiers only occur with nouns. They are optional and not all nouns occur with classifiers. In constructions with nouns, Karo classifiers follow the classifiable noun:

- (10) *naʒyop cíʔ*
 leaf CLF:thin.flat
 “leaf (also: money)” (Gabas 1999: 216–217)

3.2 Classifiers combined with verbal roots

In almost all classifier languages of the region except Karo, classifiers can be attached to verbal roots or stems. In this position, classifiers have different morphosyntactic functions, although these are not always easily distinguishable from one another. They may represent incorporated⁷ arguments and at the same time function as modifiers, creating a derived verb stem or a noun. I discuss incorporation of classifiers and modification by classifiers here and nominalization by classifiers in §3.3.

When incorporated in an intransitive predicate, classifiers can refer anaphorically to the subject of the verb in Kwaza (11), Kanoê (12), Aikanã (13), Lakondê (14) and Baure (15):

- (11) *korejaro nỹ-ro-ki*
 pan fat-CLF:CUP-DECL
 “The pan is big.” (not: “The pan is fat.”)
- (12) *opera voro-n-e-o i-kuta rwa-kuta e-re*
 jaguar black-3-NMLZ-POS \emptyset -CLF:head hurt-CLF:head DECL-AUX
 “The head of the puma is wounded.” (Bacelar 2004: 138)
- (13) *txitxipu hadi-pe-ẽ*
 grasshopper red-CLF:round-DECL
 “The grasshopper is red.”
- (14) *tẽn-‘jãw-‘tãn-ta*
 old-CLF:liquid-IPFV-ANT
 “The (chicha) is old.” (Telles 2002: 192)

7. Unlike the other languages of the region, Movima incorporation of classificatory elements often (but not always) involves changes in valence and case relations.

- (15) *ro=aromo-se-wapa to yašor*
 3SG.M=sink-CLF:oval-COS ART boat
 “The boat has sunk.” (Danielsen 2007: 208)

When incorporated in a transitive predicate they can refer anaphorically to an object of the verb in Kwaza (16), Kanoê (17), Lakondê (18) and Baure (19):

- (16) *awāta-xy-da-ki*
 see-CLF:leaf-1SG-DECL
 “I’m looking at the paper.”
- (17) *aj kometaw uru-taw o-ō-taw ã-e-re*
 I bean new-CLF:bean eat-1-CLF:bean 1-DECL-AUX
 “I eat green beans.” (Bacelar 2004: 134)
- (18) *oh-(ke’jat)-ki’nīn-‘tā:n*
 pound-(maize)-CLF:round-1SBJ.IPFV
 “I pounded (maize) kernels.” (Telles 2002: 244)
- (19) *ro=ni-po-a-po šep*
 3SG.M=eat-CLF:tiny-LK-PRFLX chivé
 “He ate chivé.” (Danielsen 2007: 209)

Classifier incorporation in Kwaza and Kanoê changes neither the valence of the verb nor the case of the argument nouns. This makes it difficult, if not impossible, to distinguish argument incorporation from derivational verb stem modification.

Even though the details of Aikanã grammar are not completely understood, there seems to be some structural distinction between argument incorporation and verb stem modification. On transitive verbs, classifiers may anaphorically refer to the object of the verb. In that case they occur in between the person and mood inflections:

- (20) *zūka-me-resa-na ka-resa-dudu*
 wash-2SG-CLF:foot-INT Ø-CLF:foot-CLF:bark
 “Did you wash your slipper?”

However, Aikanã classifiers can also further specify the semantic content of the verb root, almost like an adverb of manner. In this function, they occur in a position expected for derivational morphemes, before any person and mood inflections:

- (21) *herui-zāw ũ-nũ-ka-re-ka-ẽ*
 jatobá-CLF:fruit suck-CLF:powder-1SG-FUT-1SG-DECL
 “I’m going to ‘eat’ jatobá seeds.”⁸

8. Jatobá seeds are not literally eaten, but the powdery flesh on the outside is sucked off: ‘powder-suck the jatobá seeds’. Note that in Aikanã future tense requires recursion of the first person marker.

Similar to Aikanã, incorporation in Movima can either modify the verb or represent an argument of the verb. Movima differs from the other languages in that argument incorporation often decreases the valence of a verb, and if an independent argument noun phrase is also present, it has to be marked as oblique. Example (22) concerns classificatory modification and (23) represents argument incorporation:

- (22) *dan-a-ba=as os rey dowe*
 chew-BDR-BR:round=N.AB ART.N.PST again clothes
 “It [the cow] chewed my dress into a ball.” (Haude 2006: 385)
- (23) *duk-a:-di n-is aro:so*
 grind-BDR-BR:grain OBL-ART.PL rice
 “[I] ground rice.” (Haude 2006: 284)

Haude (2006: 376, 387) notes that it may be difficult to keep the different kinds of incorporation apart. To a certain extent, Movima truncated nominal elements also resemble classifiers, as in (7). They can be incorporated and function as anaphoric elements. However, they do not define semantic classes, as bound roots and elements do. They refer only to the noun with which they are morphologically related, and thus define single-member formal classes. One could consider them as partial repeaters.⁹

Itonama verbal classifiers always refer to a predicate argument, categorizing it in terms of shape, consistency, size, structure, animacy and position. In Example (24) both a verbal and a numeral classifier occur:

- (24) *oh-ni-yamo'-tyo uwu mi-si-lo-cha'ke*
 3-CLF:sinuous-be.few-PL river REL-EX-CLF:sinuous-MULT
 “How many rivers are there?” (Crevels 2012: 255)

Furthermore, they seem to play a role in participant number marking, as shown in (25), which is a type of verbal number:

- (25) *uh-wamo'-tyo i-wabi mi-si-dili lisiya*
 3-be.few-PL EP-woman REL-EX-CLF:seated.PL church
 “How many women are seated in the church?” (Crevels 2012: 255)

Itonama also has a set of obligatorily possessed nouns. When incorporated, these nouns appear in their truncated form, e.g. *uh-panino* “roof” (*uh* = third person) in:

- (26) *ulune mama'na si-panu-bah-na*
 Monday AUX.FUT 1SG-roof-CLF:flexible-NEUT
 “Monday I’m going to roof (the house).” (Crevels 2012: 248)

9. Seifart (2005: 81) regards truncated elements in Miraña as partial repeaters. They are less grammaticalized than genuine classifiers but more grammaticalized than full repeaters.

Nambikwaran languages are predominantly suffixing. Nevertheless, with verbs of quantification ('one', 'two', 'three', 'many' and 'few') classifiers often occur as prefixes:

- (27) *ka'loh-pan-'tān*
 CLF:flat-two-IPFV
 "There are two pieces of clothing." (Telles 2002: 316)

The classifier system of Arikapu is very restricted. There are only a handful of dedicated classifying elements and their distribution is largely limited to nouns. Only some very rare compounds or incorporations were attested:

- (28) *ihe=nō a-ka=tārī*
 I=1SBJ 2-CLF:fruit=tie.up
 "I'm going to tie you up crooked (in the shape of a ball)."

3.3 Classifiers that function as nominalizers

In Kwaza, Movima and Lakondé, classifiers (or bound roots) can be used to nominalize verbal roots and stems. Kwaza nominalization is especially highly productive in this respect. In the quality of nominalizers, Kwaza classifiers are often encountered in nominal modification constructions, which can have attributive, relative or possessive functions. Kwaza has no adjectives, and attributive modification is realized through noun juxtaposition. Semantically attributive roots are usually intransitive predicates, hence they must be nominalized before they can be used attributively:

- (29) *axy arwa-hỹ*
 house new-NMLZ
 "new house"
- (30) (*axy*) *haka-xy*
 (house) old-CLF:house
 "old house"

As the contrast between these examples shows, a specific classifier can take the place of the nominalizer *-hỹ*. As the next example shows, classifiers are also attached to fully inflected non-attributive verbs, resulting in what one could regard as a relative clause:¹⁰

10. The nominalizing function of classifiers in Kwaza attributive, demonstrative, numeral, relative and possessive constructions is discussed in van der Voort (2004, 2006, 2009).

- (31) *axy haje-a-mũ*
 house smear-1PL-CLF:liquid
 “paint” (lit. “the liquid we smear onto the house”)

Similarly, Movima bound nominal morphemes can have a nominalizing function, although there are only few examples:

- (32) *as am-a:-bun-ra*
 ART.N enter-BDR-BR:mass-BE:neutral
 “the thing in which I put dough” (Haude 2006: 478)

Also, in all Nambikwaran languages nominalization by classifiers is highly productive. The referential suffix, which is attached only to nouns, confirms that the application of the classifier to a verbal root has resulted in a noun in the following examples from Lakondê (33) and Mamaindê (34):

- (33) *laʔ-ni-‘te*
 new-CLF:hemispheric-REF
 “new house” (Telles 2002: 200)

- (34) *mān-kalo-tu*
 hot-CLF:flat-FNS
 “flat hot thing/clothes” (Eberhard 2009: 504)

In Kanoê, Aikanã and Baure there is no clearly distinct class of adjectives, and attributive modification is done by juxtaposition of nouns or nominalized semantically attributive verbs. However, the status of verb roots combined with classifiers is not always perfectly clear. The examples suggest that classifiers cannot nominalize verbs, and that a dedicated nominalizer is required for this purpose in Kanoê (35),¹¹ Aikanã (36) and Baure (37):

- (35) *etsivi-kwa-e*
 urinate-CLF:small.object-NMLZ
 “bladder” (Bacelar 2004: 112)

- (36) *txitxipu hadi-pe-i*
 grasshopper red-CLF:round-NMLZ
 “red grasshopper”

- (37) *p-a-š ka’an moser-a-no*
 one-CLF:animal-one animal red-CLF:animal-NMLZ
 “one red animal” (Danielsen 2007: 105)

11. Although Kanoê classifiers can apparently nominalize attributive verb roots, as Bacelar’s (2004: 101) examples show.

3.4 Interchangeability of classifiers and nouns

In a process of grammaticalization, free lexical nouns may develop into classifiers. Often, but not always, classifiers can be recognized as bound and semantically bleached reflexes of lexical nouns (e.g., Grinevald 2004: 1017, Lehmann 2015: 136–138). Also in the classifier languages of the Southwestern Amazon the nominal origin of classifiers can often be attested. In some languages compounding and noun incorporation may be difficult to distinguish from classification. But even in languages without compounding or incorporation, if there is any evidence at all for the etymological origin of a certain classifier, it is usually a noun.

In Kwaza and Kanôê, no classifiers were found to occur as free morphemes, like in some other languages, and vestiges of noun compounding or incorporation are rare. Nevertheless, the classifiers probably originate from nouns. Both Kwaza and Kanôê have sets of classifiers that are etymologically clearly related to existing nouns, as well as large sets of classifiers that can be turned into independent nouns by an empty prefix.

In some languages discussed here, otherwise free nouns can be used as bound classifiers, and in others, otherwise bound classifiers can be used as free nouns. This is not the general situation, however. In Aikanã, a few free nouns can be used as classifiers. The exact size of this set is unknown; it must be fairly limited, since nominal compounding is rare. They include *ji* ‘hair, leaf, feather’, *jê* ‘flesh’, *zu* ‘bone’, *waru* ‘tongue’.

- (38) *küri=ji*
 babaçu=hair
 ‘babaçu palmtree straw’
- (39) *hane=hane-pa=waru-a-ê*
 hot=hot-APPL=tongue-1SG.OBJ-DECL
 ‘It burns my tongue.’

Also in Baure, there are attestations of lexical nouns used as classifiers (Danielsen 2007: 14, 143), although this is probably quite rare.

Nambikwaran classifiers probably derive from compounded specific nouns. In some Lakondê constructions – but not in Mamaindê – they can even occur as free morphemes:

- (40) *ka’loh ʔa?-tân-ta*
 CLF:flat big-IPFV-ANT
 ‘The board is large.’, ‘The clothes are big.’, etc. (Telles 2002: 193)

In examples like this Telles (2002: 192) considers the classifier as substituting a full noun that was mentioned previously in the discourse, to which it refers

anaphorically (compare Example 57). In Mamaindê, a small number of nouns (three or four) can be used as classifiers (Eberhard 2009: 353). Furthermore, possessor prefixes and demonstrative prefixes in Northern Nambikwaran are usually attached to full nouns, but they can also be attached to classifiers:

- (41) *tq-ni-tu*
 1SG.POS-CLF:hemispheric-REF
 “my house”, “my pan” (Telles 2002: 194)

3.5 Generic classifiers and obligatory grammatical slots

Some classifier languages have an obligatory slot for classifiers. In those languages a generic classifier may function as a grammatical placeholder if a specific classifier is not applied. Kwaza is a prime example of this. Several constructions in Kwaza, such as possessives, numerals, demonstratives, etc., have an obligatory slot for a classifier. If no classifier is used, the nominalizer *-hỹ* occurs, functioning as a generic, or semantically neutral,¹² classifier (see §3.3). A similar situation was found in Movima. The following contrasting examples from Movima show bound nominal morphemes obligatorily incorporated in an adjective:

- (42) *tas-poy is paj-‘i*
 three-BR:animal ART.PL BR:dolphin-ø
 “There are three dolphins.” (Haude 2006: 208)
- (43) *tas-ra is paj-‘i*
 three-BE:neutral ART.PL BR:dolphin-ø
 “There are three dolphins.” (Haude 2006: 208)

Haude (2006, 208) observes that incorporation into verbs is most productive with part-of-whole terms and that not all bound roots (such as *paj* “BR:dolphin”) can be incorporated. Instead, a bound hyperonym is then used, like *poy* “BR:animal”, or the semantically neutral bound element *-ra*, which functions here as a generic classifier.

It appears that Southern Nambikwaran has a nominalizer that can function as a generic classifier (see Kroeker 2001: 27), although its distribution may not be as general as in Kwaza and Movima. In other languages there is no generic classifier and there are no constructions with an obligatory grammatical slot for classifiers. There, the presence of a classifier depends on lexical and pragmatic factors.

12. By a generic classifier I do not mean a classifier with RELATIVELY general semantics, such as ‘liquid’, which can refer hyperonymically to different items, as is common with classifiers.

3.6 Rare constructions

This section sketches classifier constructions particular to one or very few languages of the region. Most classifier systems in the region involve nominal or verbal hosts. Although one could claim that, in e.g. Kwaza, classifiers can occur on demonstrative, adjectival and semantically numeral roots, these roots all are grammatically verbal. However, in some languages, such as Cavineña and Karo, adjectives represent a category that is separate from verbs and nouns. This has no consequences for the classificatory system of Cavineña, but it does for Karo.

In Karo, when a construction is modified by an adjective the classifier must follow both constituents, as in (44):

- (44) *naʒyop cíʔ cú cíʔ*
 leaf CLF:thin.flat big CLF:thin.flat
 “big leaf” (Gabas 1999: 225)

Kwaza is the only language in which classifiers were attested (although very seldom) in combination with adverbials, as the following examples show:

- (45) *a-toto-xy*
 ∅-DR:upward-CLF:house
 “on top of the house”
- (46) *tsiritsa-mũ*
 middle-CLF:liquid
 “the middle of the river”

Itonama keeps its systems of numeral and verbal classifiers strictly separated, both formally and functionally. The numeral classifiers appear exclusively on numeral roots, whereas verbal classifiers also appear on demonstrative roots:

- (47) *noʔ-so o-pi lowoʔ-tya k-aʔ-ki-maku-mu chukʔaʔe*
 DEM:PROX-CLF:flat.PL EP-fish be.rotten-STAT F-2SG-IMP-give-1 other
 “These fish are rotten, give me others!” (Crevels 2012: 252)

4. The empty morpheme construction

The empty morpheme construction, which enables classifiers and bound roots to occur as independent words, is pervasive in the Southwestern Amazon. This construction involves what is either a root or a derivational morpheme that has no semantic content itself but that allows the semantic content of, respectively,

a classifier or a bound nominal root, to be expressed as an independent noun.¹³ Because of cross-linguistic similarities that concern both its formal and functional characteristics, it represents one of the most conspicuous areal linguistic traits of the region. In this section, the construction is elucidated for each language in which it is attested. In §5.2, similarities are discussed.

4.1 Kwaza

One particular subset of classifiers in Kwaza (the set III classifiers) does not seem to be related to existing independent lexemes. What these classifiers furthermore have in common is that they can be attached to the semantically empty root *e-* and form an independent noun that carries the meaning of the classifier:

- (48) *e-mũ*
 ∅-CLF:liquid
 “liquid”
- (49) *e-sĩ*
 ∅-CLF:seed
 “seed”
- (50) *e-kai*
 ∅-CLF:leg
 “(its) lower leg”

An alternative analysis of the dummy root *e-* relates it to the semantically abstract verb root *e-* “to have”, which may have been compounded with nouns in an earlier phase of the language. There is no independent diachronic or synchronic evidence for such hypothetical compounding, however, in Kwaza.

The verb *e-* “to have” is often used to express predicative possession in Kwaza, and, although rarely, as an existential verb. The normal existential verb in Kwaza is *a-* “to exist”, which, as it happens, is homophonous with a semantically empty adverb formative root *a-*. This adverb dummy root is used to create independent directional adverbs out of derivational verbal directional suffixes, as in (51):

13. The phenomenon of the empty root has also been attested in Eskimo languages, such as West Greenlandic, where derivational suffixes tend to carry considerable semantic content, to which the empty root *pi-* can lend syntactically independent status.

- (51) *a-nōwē* *mūi-’ra*
 ∅-DR:upstream submerge-IMP
 “Fetch water upstream!”

It seems possible that a folk-etymological connection was made that led to the creation of the empty adverb formative *a-* construction. The empty root *a-* is probably derived from the semantically abstract verb root *a-* “to exist”, analogous to reanalyzed *e-* “to have”. In addition, certain directional suffixes are difficult to distinguish from classifiers, and in some cases their distributions overlap. Since the two roots *e-* and *a-* both can function as semantically empty word-formative roots and as semantically rather abstract verb roots, their parallel distributions suggest a relationship between their origins.

4.2 Kanoê

Also Kanoê has a large subset of classifiers that are not related to independent nouns but that can turn independent themselves when attached to the empty morpheme *i-*. Although the empty morpheme occurs mainly with ‘inalienable’ items such as body part classifiers (52), it occurs also with classifiers expressing concepts that are not commonly expected to count as inalienable (53), whereas certain expressions for concepts usually regarded as inalienable are never combined with the *i-* morpheme, such as *nū* “breast”.

- (52) *i-kuta*
 ∅-head
 “(its) head” (Bacelar 2004:130)
- (53) *i-ka*
 ∅-CLF:wood
 “wood” (Bacelar 2004: 128)

The behavior of the empty morpheme *i-* in Kanoê possessive constructions and their similarities and differences with those in Kwaza are discussed in detail in van der Voort (2009: 366).

4.3 Aikanã

There is a substantial set of body-part classifiers in Aikanã. They can be turned into independent nouns by attaching them to the semantically empty morpheme *ka-*:

- (54) *ka-pedika*
 ∅-CLF:back
 “back”

They are analyzed here as bound derivational classifiers, and the element *ka-* as a part-whole noun formative root. Not all of these classifiers refer to body parts – *-ta* “outer side of vessel”, *-peka* “rim” – or perhaps even to parts of a whole – *-nape* “edge of forest”, *-rjüa* “sky”. Furthermore, a small set of body-part nouns do not occur with the element *ka-*, e.g. *kuikui* “chin”, *müi* “tooth”, *tinü(pa)* “head”, *waru* “tongue”, *zu* “bone”. Apart from their occurrence with the noun formative root, the body part classifiers have exactly the same distribution as other classifiers.

Some of the classifiers in both groups seem to contain a (fossilized) initial *i-* or *e-* that is reminiscent of the noun formative *e-* root in other languages: *ka-ekawa* “throat, Adam’s apple”, *ka-inedüka* “lower leg”. Perhaps the element *ka-* represents a recent innovation to regularize the expression of body parts. In some cases, however, the (old?) initial *i-* is not fossilized with the classifier but with the regular *ka-* noun formative: *kai-zakape* “top, tree stump”, *kai-depa* “tail”.¹⁴ In one case, the noun formative appears as *ki-*: *ki-nü* “stick, penis”. When the classifier is attached to a lexical root, the entire complex noun formative is absent:

(55) *txā-zakape-a-ē*
 tear-CLF:top-1SG.OBJ-DECL
 “I tore the skin of my skull.”

(56) *jāpuri tjabai-nū-ē*
 manioc big-CLF:stick-DECL
 “It is a big manioc tuber.”

Only (*h*)*e-dudu* “skin” may contain a genuine *e*-like noun formative, which is omitted when the noun is used as a classifier. Some independent nouns may have a fossilized *i-* noun formative, such as *iri* “liver” and *iridai* “nail”, although none of these nouns can be used as classifiers.¹⁵

4.4 Movima

In Movima, there is a small subset of bound noun roots that can be turned into full nouns by a dummy element *-i*. This element is different from noun formative roots in other languages in that it is suffixed rather than prefixed and in that it only serves to add up to the minimally required number of syllables. Bound roots are often monosyllabic in Movima and cannot occur independently, since independent nouns must at least have two syllables. Bound roots occur in compounds and incorporation, but there are strategies to increase the number of their syllables in

14. Note that Aikanā /ai/ is pronounced [əj].

15. In Kwaza and Kanoê, respectively, the classifiers for ‘liver’ are *-ri* and *-ry*, and the Aikanā word is probably etymologically related.

order to lend them independent status: they can be reduplicated; they can receive the absolute state suffix *-kwa* “detached”; or they can be attached to the dummy element *-i*. Haude (2006: 207) lists 12 *-i* nouns, such as *beñ-‘i* “countryside, grassland”, *chaj-‘i* “breast(s) (my)”, *nun-‘i* “bone”, *paj-‘i* “dolphin”. Only three of these roots represent so-called ‘relational’ nouns, which means they are inalienably possessed, e.g. *chaj* “breast(s)”. When unmarked for possession a first person singular possessor is implied (Haude 2006: 230).

4.5 Lakondê

The anaphoric faculty of classifiers makes it possible that the use of a classifier is sufficient if the referent has already been mentioned previously in the discourse. Telles (2002: 192) argues that, if the referent has not yet been mentioned, it can be introduced by a classifier that is attached to the nominal lexical dummy root *ĩ-*.

(57) *ĩ-ka’loh ka’nah-‘tān-ta*
 ∅-CLF:flat one-IPFV-ANT
 “It is a piece of cloth.” (Telles 2002: 194)

(58) *ĩ-‘kah ‘ajh tu-∅*
 ∅-CLF:long go take-IMP
 “Fetch the stick!” (Telles 2002: 194)

The discussion in Telles (2002) is limited, but in the accompanying dictionary CD there are a few additional examples in which the empty element is represented as *in-*, such as *in-kinĩn-* “eye”, and some of these examples may be lexicalized, such as *in’kalaŋ-* “umbilical cord”.

No similar morpheme is attested in Mamaindê or Southern Nambikwaran, but Sabanê classifiers contain vestiges of an empty morpheme. Most of the Sabanê classifiers are *i-* or *a-* initial, and this vowel is present when the classifier is used as an independent noun. When the classifiers are attached to roots or stems, their initial vowel is either truncated or an epenthetic consonant is inserted. The initial vowel of *isu* “bone” is apparently always truncated or assimilated when it is attached to another lexical stem, resulting in *-su*. It is furthermore conspicuous that body part nouns in Sabanê are also fairly consistently *i-* or *a-* initial, which Araujo (2004: 102) seems to regard as possible vestiges of (Proto-Nambikwaran) inalienability marking. In this respect Sabanê is like several unrelated languages of the region, such as Cavineña, Kanoê and Kwaza, where both classifiers and body part nouns are often *i-* or *e-* initial. Also for these languages, inalienability marking is considered as a possible historical explanation for apparently fossilized body-part prefixes, although not more than that. On the other hand, the initial

vowel *i-* or *a-* in Sabanê is not very productive as a semantically empty prefix or dummy root.

4.6 Cavineña

When pronounced in isolation or occurring as independent nominal constituents in the clause, Cavineña bound noun roots require the semantically empty prefix *e-*. When bound noun roots are further derived, as in compounds, incorporation or under adjectivization or verbalization, there is no empty prefix. The following contrasted examples show absence of the empty prefix when the bound noun root is a second member in a noun compound:

- (59) *e-rami*
 ø-flesh
 “flesh, meat”

- (60) *waka rami*
 cow flesh
 “cow meat”

(Guillaume 2008: 411)

However, if a bound noun root is the first member in a noun compound, its empty prefix is maintained:

- (61) *e-tsaka rami*
 ø-leg flesh
 “flesh of the leg”

(Guillaume 2008: 455)

Compounds are often lexicalized, having led to the emergence of composite (*e-*) nouns, such as such as *ebipukaka* “fist”, which is based on *e-bi* “arm” and *e-pukaka* “kneecap” (Guillaume 2008: 450).

Guillaume (2008, 412–415) also describes irregularities in the behavior pattern of various bound roots and the empty prefix. In this respect it is worth mentioning two bound noun roots that cannot occur with *e-* but that require a different empty prefix. The root *-kwija* “thorn” requires an empty prefix *a-*, and the root *-nime* “thought” requires an empty prefix *i-*:

- (62) *a-kwija*
 ø-thorn
 “thorn”

- (63) *i-nime*
 ø-thought
 “thought”

(Guillaume 2008: 412)

Apart from their empty prefix *e-*, bound noun roots are indistinguishable from other nouns with regard to morphosyntactic behavior. Semantically, however, the large majority of bound noun roots refer to body part concepts. This could indicate that the empty prefix originates in a grammatical distinction between alienable and inalienable nouns. Then again, this could at most be a historical remnant of such a distinction, since many Cavineña nouns and concepts do not comply with this pattern. As an alternative explanation, Guillaume (2008: 430–431) suggests that the empty prefix could represent an old indefiniteness marker.

As Vuillermet (2012) shows, the empty prefix construction in Ese Eja is similar to that of Cavineña in almost every detail. In fact, the construction is similar and productive to various degrees in all Tacanan languages, and it may therefore have existed already in Proto-Tacanan.

4.7 Arikapu

Arikapu has a number of obligatorily possessed nouns, usually referring to parts of wholes, resembling the classifiers of other languages of the region. The prefixes for first person singular (*i-*) and first person plural (*txi-*) are also used as third person or impersonal prefixes. These two prefixes function often as dummy possessors with inalienable concepts as well as dummy arguments with verbs. When the possessor or the third person object is expressed by a full noun, the impersonal prefixes are absent. The following examples contrast these constructions:

- (64) *i-txawa*
 1SG-flower
 “flower”, “my flower”
- (65) *bitojo=txawa*
 nettle=flower
 “flower of a nettle (*Fleurya aestuans* L.)”

Arikapu also has a small set of genuine classifiers. In order to form an independent expression of the semantic content of a classifier, the first person singular prefix *i-* can be applied with impersonal meaning, as in *ika* “fruit”, *imrẽ* “porridge, dregs”, *ihã* “seed”, *iko* “hole”, etc. Also the sister language Djeoromitxi probably has bound classifiers. If their semantic content is to be expressed as an independent form, prefixation of third person *i-* or first person plural *hi-*, functioning as dummy possessor, is required, like in *ihõ* “kernel” and *ik#* “hole”.

4.8 Baure

In Baure, obligatorily possessed nouns are bound nominal roots that must either receive specific possessor inflection or a semantically empty prefix *e-* when the possessor is unknown. This prefix *e-* is analyzed by Danielsen (2007: 119) as a marker of unspecified possession.

(66) *ni=ser*
 1SG=tooth
 “my tooth” (Danielsen 2007: 120)

(67) *e-ser*
 \emptyset -tooth
 “a tooth of someone” (Danielsen 2007: 120)

With some roots, however, a construction involving the empty prefix *e-* is not fully transparent and may have an unpredictable specific meaning, e.g. *waki* “palm of the hand” vs. *ewaki* “forked branch of tree” (Danielsen 2007: 120).

5. Discussion

When studying the systems of classification in the languages considered here, a number of structural and formal similarities stand out. These languages tend to have multiple classification systems, usually combining a smaller basic system operating on relatively general semantic principles (dimensions of shape) with larger systems often exhibiting very specific semantics. The multiple distinguishable subsets of classifiers in a language nevertheless tend to be of the same morphosyntactic type (except for Itonama and, to a lesser extent, Baure), sharing the same structural properties and distribution, which is why they are all considered here. In the Table 3 the characteristics of the classificatory systems in the languages discussed are summarized:

In spite of the genetic diversity of the Southwestern Amazonian languages considered here, their classification systems show similarities. Some of these are probably due to universal tendencies, such as a small set of classifiers to express basic notions of shape, consistency and arrangement. It is likely that certain similarities occurring within language families such as Nambikwaran, Tacanan and Jabutian reflect properties of the ancestor language(s), hence are due to a genetic factor. Other similarities, however, cannot be explained by universal tendencies or genetic inheritance; they must be the result of areal diffusion of (sub)systems and forms, the origin of which remains undetermined. Some of these areal similarities

Table 3. Some characteristics of classifying systems

	Kwaza	Kanoê	Aikanã	Movima	Itonama	Lakondê	Cavineña	Baure	Karo	Arikapu
Total nr. of CLs	150	100	45	70	25	20	130	135	10	10
Noun modification	+	+	+	+	+	+	+	+	+	+
Verb argument	+	+	+	+	+	+	+	+	-	(+)
Anaphoric	+	+	+	+		+		+	-	+
Concord/agreement	-	-	+	-	-	+	-	+	+	-
Nominalizing	+	-	-	+	-	+	-	-	-	-
Generic CL	+	-	-	+	-	(-)	-	-	-	-
Multiple sets of CLs	+	+	+	+	+	(+)	-	+	-	-
Stacking of CLs	+	+		+		+			(-)	-
Composite CLs	+	+	(+)				+		-	+
Empty morpheme	+	+	+	+	-	+	+	+	-	(+)
Free use of CL	-	-	-			+	(+)		+	-
Noun used as CL			+					+	-	+

are characteristic of Amazonian classifiers systems in general, such as large sets of bound classifiers with specific semantics and a wide-ranging distribution, confirming general patterns identified by Aikhenvald (2000), Grinevald (2007) and Seifart & Payne (2007). Yet some of the similarities seem to be characteristic of the Southwestern Amazon only.

5.1 Classifying forms

Several classifying morphemes are shared by genealogically unrelated Guaporé-Mamoré languages, which implies that borrowing of bound morphemes must have taken place. Since these classifiers are almost identical across the relevant languages, their diffusion must be of relatively recent date. The origin of the particular classifiers is unclear, as is the direction of borrowing. The languages that share certain classifiers are (with the exception of Tacanan) all conspicuously close neighbors and all are spoken on the Brazilian side of the Guaporé River. Consider the following table:

Table 4. Diffusion of classifiers

	Kwaza	Kanoê	Aikanã	Jabutian languages	Nambikwaran languages	Tacanan languages
'bark'	<i>-kalo</i>				<i>-kalo</i>	
'bone'	<i>-xu</i>		<i>-zu</i>		<i>-su</i>	<i>-tsau</i>
'vessel'	<i>-tay</i>	<i>-tæ</i>	<i>-ta(e)</i>			
'fruit'	<i>-ko</i>	<i>-ko</i>				
'hair'	<i>-xyi</i>		<i>(ji) -di</i>	<i>hi-,txi-</i>		
'liquid'	<i>-mũ</i>	<i>-mu</i>	<i>-mũ</i>	<i>(-mũ)</i>		
'porridge'	<i>-mẽ</i>			<i>-mrẽ</i>		
'powder'	<i>-nũ</i>	<i>-nu</i>	<i>-nũ</i>	<i>nũ</i>	<i>-nũ</i>	
'round'	<i>-tε</i>	<i>-tæ</i>	<i>-zãw</i>			
'thorn'	<i>-nĩ</i>			<i>-nĩ</i>		
'tooth'	<i>-mãi</i>		<i>-mãi</i>			
'liver, flat'	<i>-ri</i>	<i>-ry</i>	<i>(iri)</i>			

The majority of these classifiers are shared by the Rondônian language isolates only, which are spoken by neighboring groups that have been in close contact especially during the past 100 years. Some of the classifiers are shared with Nambikwaran languages, which were spoken by peoples more distantly towards the east with whom relationships were in general not friendly. Two classifiers appear to be shared

exclusively between Kwaza and the Jabutian languages, spoken on the headwaters of the São Pedro and Branco Rivers, respectively, with only a stretch of plains between them.

The classifier *nũ* ‘powder, porridge, granular substance’ is the most widespread. The fact that it is also found in the Tupian language Munduruku is intriguing, since Munduruku is spoken on the lower Tapajós River, far away from the Guaporé-Mamoré region. One could imagine that direct or indirect contact with Nambikwara speakers at the headwaters of the Tapajós led to its diffusion towards the north. The same classifier is also attested in Arikapu, where it functions as a noun, meaning ‘porridge, paste’. It may be cognate with Djeoromitxi *nō* ‘food, dish’, and has been reconstructed for Proto-Jabuti as **nũ* ‘food’, and related to other Macro-Jêan languages such as Karajá *do* ‘food’ and Apinajê (Northern Jê) *ñō* ‘food’ (Ribeiro & van der Voort 2010).¹⁶ Perhaps *nũ* represents one of those elements that are found throughout the entire Amazon, similar to lexical items such as *kumanda* ‘beans’, *manari* ‘sieve’ and others listed in van der Voort (2005: 387).

Of the other classifiers in Jabutian languages only *-mũ* is similar to multiple other languages, but it is only found fossilized in specific river names, such as the *Anamũ* and *Torĩmũ*, that may be borrowed from other Guaporé-Mamoré languages. With respect to Kwaza *-nĩ* ‘thorn, needle’, Proto-Jabuti **-nĩ* ‘thorn’ has likely cognates in several Macro-Jêan languages, such as Apinajê (Northern Jê) *ñi*, Rikbaktsá (*i-*)*ni*. This suggests that, if the similar Kwaza and Jabutian forms are not due to coincidence, the Jabutian languages were probably the source. Alternatively, coincidental similarity with neighboring languages may have been an extra motivation for original Proto-Jabuti traits to be preserved in the Jabutian languages.

The classifier *-su* ‘bone’, which is shared between Nambikwaran, Tacanan and isolate languages, may have been subject to diffusion. Also here, the question of directionality is difficult to assess. The fact that the classifier *-kalo* ‘bark’ is shared between Nambikwaran languages and only a single other language weakens the idea of diffusion. However, the fact that it concerns a polysyllabic element reduces the chance that the similarity is due to coincidence.

5.2 The empty morpheme

As seen in §4, various languages of the region possess a semantically empty morpheme that lends independent nominal status to classifiers or bound nouns, which it usually precedes. This morpheme is described either as an affix or as a root in the relevant works. Its form is usually *e* or *i*. It probably represents an areal feature,

16. Note also the Lakondê / Latundê classifier *-tũ-* and *-to-* ‘food, powder’ and the Djeoromitxi element *tu* ‘powder, flour’.

and its diffusion was touched upon in various publications (van der Voort 2004, 2005, 2009; Crevels & van der Voort 2008). The following examples illustrate the pervasiveness of this feature among the Guaporé-Mamoré languages:

- | | | |
|------|---|---|
| (68) | <i>e-kai</i>
ø-CLF:leg
“lower leg”, “its lower leg” | Kwaza (ISOLATE) |
| (69) | <i>i-kuta</i>
ø-CLF:head
“head”, “its head” | Kanoè (ISOLATE)

(Bacelar 2004: 130) |
| (70) | <i>ka-supā</i>
ø-CLF:face
“face” | Aikanā (ISOLATE) |
| (71) | <i>nun-‘i</i>
bone-ø
“bone” | Movima (ISOLATE)

(Haude 2006: 254) |
| (72) | <i>ī-ka’loh-te</i>
ø-CLF:flat-REF
“board” | Lakondè (NAMBIKWARAN)

(Telles 2002: 96) |
| (73) | <i>e-rami</i>
ø-flesh
“meat” | Cavineña (TACANAN)

(Guillaume 2008: 410) |
| (74) | <i>e-ser</i>
ø-tooth
“a tooth” | Baure (ARAWAKAN)

(Danielsen 2007: 120) |

It is tempting to include the Jabutian languages in the list. Even though the prefix in these languages is a person inflection, which is neither empty nor limited to classifiers, it is frequently used as an impersonal possessor:

- | | | |
|------|--|--------------------------|
| (75) | <i>i-kə</i>
1SG-skin
“paper”, “its paper”, “my skin” | Arikapu (MACRO-JÊAN) |
| (76) | <i>i-rawa</i>
3-flower
“flower”, “its flower” | Djeoromitxi (MACRO-JÊAN) |

The similarity of the Jabutian examples with the other ones is probably not coincidence: one could imagine that the person marker *i-* in the Jabutian languages was reanalyzed as an impersonal possessor under areal influence. The same can perhaps be said of the Sirionó language (TUPI-GUARANIAN) of northeastern Bolivia.

As appears from the work by Schermair (e.g. 1962) and Priest & de Priest (1985), Sirionó has a subclass of obligatorily possessed nouns. If there is no possessor in focus, the third person prefix *e-* is applied.

There are some cross-linguistic differences in the forms and meanings of the empty morpheme construction. In most descriptions, the *e-* or *i-* element is analyzed as a dummy element. Guillaume (2008: 409) defines and glosses it either as a “noun prefix” or as a noun class marker in the Tacanan languages. Danielsen (2007: 119) analyzes and glosses it as an “unspecified possessor prefix” in Baure. In both cases it can be regarded as semantically empty. In the Jabutian languages it is part of the person marking paradigm; in Arikapu it is a first person marker and in Djeoromixti it is a third person marker (Pires 1992; de Castro 2012; Ribeiro & van der Voort 2010). However, in both languages these person markers are attested with an alternative, impersonal interpretation. In Movima, the empty morpheme is a suffix of limited use. In Aikanã an *e-* or *i-* element is attested as a fossil, whereas its productive empty morpheme *ka-* has the same function as the *i-* element in other languages. Kwaza has an additional construction to form independent adverbs out of directional morphemes. It involves the empty element *a-* and is perfectly analogous to the noun-formative construction involving classifiers and the empty element *e-*.

The empty morpheme construction is also attested in other regions. A semantically empty element to turn bound classificatory elements into independent nouns was attested by Hart (1963) in Harakmbut or Amarakaeri, a Harakmbut-Katukinan language of Peru. In this language, the empty “classifier prefix” *wa-* is only attested with a subset of classifiers that often, but not exclusively, refer to body parts. Quite notable is also what Petersen (2007: 399) calls the ‘relational’ morpheme *i-* in Witoto or Muinane, a Witototan language of Colombia with a very rich classifier system. As appears from her remarks, this is a prefix that can be used to turn classifiers into independent nouns, e.g. as citation forms, although it can also be prefixed to other bound nominal morphemes. Vengoechea (2005: 124; 2009: 14) shows that this ‘grammatical’ morpheme is homophonous with the second person plural possessive morpheme and suggests that it may function as a dummy possessor of, especially but not exclusively, body parts. In Miraña (Boran) there are pronominal elements that occur in similar constructions to provide a classifier with the status of independent noun (see Seifart 2005: 125ff.).

The use of a specific person marker to express an impersonal or unspecified possessor is attested in many Amazonian languages. Furthermore, the form *i-* or *e-* as a person prefix is encountered in so many language families of the Americas that it is sometimes referred to as a pan-Americanism (see Campbell 1997: 266; Greenberg 1987: 279). Nevertheless, in most of the Southwestern Amazonian languages discussed here the empty morpheme is not a person prefix.¹⁷ Rather, it is an

17. The majority do not have nominal prefixes for person at all.

from Chiquitano. However, the same construction is found in another language in the Tupari family, Makurap (Braga 2005: 37–39). Furthermore, the domestic animal indirect possessive construction is also found in other Arawakan languages, so a genetic explanation may be involved in the case of Baure. Then again, in the Tupian family the construction is rare outside the Tupi-Guaranian branch, and it is possible that Mekens and Makurap have borrowed the construction, either from a Tupi-Guaranian or an Arawakan language. The construction in Yurakaré (which went unnoticed by van der Voort 2009) may be the result of diffusion.

6. Conclusion

I have dealt mainly with those languages of the Southwestern Amazon region that possess grammaticalized systems of classification. These systems show notable similarities, even though the range of languages across which they appear are genealogically very diverse: five language isolates, several members of three local language families (Jabutian, Nambikwaran and Tacanan) and some members of two big language families (Arawakan and Tupian). Some of the similarities, namely the empty *e-* morpheme construction and certain specific forms, are especially characteristic of the Southwestern Amazon. Although it is difficult to define the exact nature and limits of the linguistic area(s) in the Southwestern Amazon, the region does stand out when contrasted with other parts of South America.

It seems reasonable to assume that classifiers result from the grammaticalization of nouns, possibly, but not necessarily, via a stage of (partial) repeaters. Furthermore, classifiers might develop out of bound nouns. Bound nouns themselves seem to develop from nouns with a relational (inalienable and/or obligatorily possessed) semantic content. Since such nouns tend to occur in a morphological or syntactic construction that expresses possession, their unpossessed occurrence could be felt as unnatural, which may lead to loss of their status as independent parts of speech. Whatever their origin or status vis-à-vis independent nouns, in most languages the different subsets of classificatory morphemes share the same distribution. The existence of a smaller system of classifiers may have functioned as a model for the use and distribution of bound nouns. In some of the Guaporé-Mamoré languages, the original smaller classifier system may have expanded through analogy to become a comprehensive classifier system. It is possible that different languages originally possessed limited systems of classification independently from each other, but they could also have developed such systems under each other's influence. Especially for the isolates and the regional small language families, it is difficult to say whether (limited) classifier systems were inherited or contact-induced. However, it seems quite likely that neighboring languages have

at least influenced one another with respect to the organization of their expanded classifier and bound noun systems. Classifier systems are a general characteristic of Arawakan languages (Aikhenvald 1999: 83). To what extent the Arawakan languages have influenced the Guaporé-Mamoré languages in this respect, or the other way round, is an issue worth pursuing in detail.

Acknowledgments

This article would not have been possible without the hospitality of the Indigenous communities in the Rio Branco, Rio Guaporé and Tubarão-Latundê federal reserves in Rondônia, Brazil. In particular I want to thank my Indigenous teachers, Nazaré Arikapú, Mario Kwaza and Manoel Aikanã for their cooperation and their patience. I am furthermore indebted to Laércio Bacelar, Mily Crevels, Swintha Danielsen, Nilson Gabas Jr., Gale Goodwin Gómez, Colette Grinevald, Antoine Guillaume, Katharina Haude, Olga Krasnoukhova, Bill McGregor, Gessiane Picanço, Jan Rijkhoff, Françoise Rose, Frank Seifart, Stella Telles, Søren Wichmann, Thekla Wiebusch and two anonymous reviewers for comments and suggestions. This article represents one result of the project Language diversity of the Guaporé region, generously funded by the Netherlands Organisation of Scientific Research (NWO), under VIDI grant nr. 276-70-005. This article has also benefitted from feedback at the Jerusalem, Canterbury and Wassenaar meetings on classifiers funded through European Cooperation in Science and Technology (COST) Action 31 *Stability and adaptation of classification systems in a cross-cultural perspective*, chaired by Thekla Wiebusch. Responsibility for the contents of this article is mine only.

References

- Aikhenvald, Alexandra Y. 1994. "Classe nominal e gênero nas línguas Aruák". *Boletim do Museu Paraense Emílio Goeldi, Antropologia* 10(2).137–259.
- Aikhenvald, Alexandra Y. 1999. "The Arawak Language Family". *The Amazonian Languages* ed. by R. M. W. Dixon & Alexandra Y. Aikhenvald, 65–106. Cambridge: Cambridge University Press.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun classification devices*. Oxford: Oxford University Press.
- Araujo, Gabriel Antunes de. 2004. *A Grammar of Sabanê, a Nambikwaran Language*. Utrecht: LOT Publications.
- Bacelar, Laércio N. 2004. *Gramática da língua kanoê*. Ph.D. dissertation, Katholieke Universiteit Nijmegen. <http://repository.ubn.ru.nl/handle/2066/19429>. (Accessed: 27 October 2017.)
- Borgman, Donald M. 1990. "Sanuma". *Handbook of Amazonian Languages*, Vol. II ed. by Desmond C. Derbyshire & Geoffrey K. Pullum, 15–248. Berlin: Mouton de Gruyter.
- Braga, Alzerinda de Oliveira. 2005. *Aspects morphosyntaxiques de la langue makurap*. Ph.D. dissertation, Université Toulouse II.
- Campbell, Lyle. 1997. *American Indian Languages: The historical linguistics of Native America*. New York: Oxford University Press.

- Crevels, Mily. 2002. "Itonama o *sihnpadara*, lengua no clasificada de la Amazonía boliviana". *Estudios de Lingüística Universidad de Alicante* 16.11–41. doi:10.14198/ELUA2002.16.01
- Crevels, Mily. 2006. "Verbal Number in Itonama". *What's in a Verb? Studies in the verbal morphology of the languages of the Americas* ed. by Grażyna J. Rowicka & Eithne B. Carlin, 159–170. Utrecht: LOT Publications.
- Crevels, Mily. 2012. "Itonama". *Lenguas de Bolivia, II: Amazonia* ed. by Mily Crevels & Pieter Muysken, 233–294. La Paz: Plural Editores.
- Crevels, Mily & Hein van der Voort. 2008. "The Guaporé-Mamoré Region as a Linguistic Area". *From Linguistic Areas to Areal Linguistics* ed. by Pieter Muysken, 151–179. Amsterdam: John Benjamins. doi:10.1075/slcs.90.04cre
- Danielsen, Swintha. 2007. *Baure: An Arawak language of Bolivia*. Leiden: CNWS Publications.
- de Castro, Thiago Vital. 2012. *Djeoromixí: Notes on phonology and simple noun phrase structure*. M.A. thesis, The University of Texas at Austin.
- Derbyshire, Desmond C. & Doris L. Payne. 1990. "Noun Classification Systems of Amazonian Languages". *Amazonian Linguistics: Studies in Lowland South American languages* ed. by Doris L. Payne, 243–271. Austin: University of Texas Press.
- Eberhard, David M. 2009. *Mamaindê Grammar: A Northern Nambikwara language and its cultural context*. Utrecht: LOT Publications.
- Epps, Patience. 2007. "Birth of a Noun Classification System: The case of Hup". *Language Endangerment and Endangered Languages* ed. by Leo Wetzels, 107–127. Leiden: CNWS Publications.
- Fabre, Alain. 2007. "Morfosintaxis de los clasificadores posesivos en las lenguas del Gran Chaco (Argentina, Bolivia y Paraguay)". *UniverSOS. Revista de Lenguas Indígenas y Universos Culturales* 4:67–85.
- Gabas, Nilson, Jr. 1999. *A Grammar of Karo, Tupí (Brazil)*. Ph.D. dissertation, University of California at Santa Barbara.
- Galucio, Ana Vilacy. 2001. *The Morphosyntax of Mekens (Tupi)*. Ph.D. dissertation, University of Chicago.
- Gijn, Rik van. 2006. *A Grammar of Yurakaré*. Doctoral dissertation, Radboud Universiteit Nijmegen. <http://www.zora.uzh.ch/id/eprint/84700/1/gramofyu.pdf>. (Accessed: 27 October 2017.)
- Gomes, Dionei Moreira. 2006. *Estudo morfológico e sintático da língua mundurukú*. Ph.D. dissertation. Brasília: Universidade de Brasília.
- Gonçalves, Cristina Helena R. C. 1987. *Concordância em mundurukú*. Campinas: Editora da Unicamp.
- Goodwin Gómez, Gale. 2000. "Noun Classifiers in Ethnobotanical Terminology of a Yanomami Language of Brazil". *Indigenous Languages of Lowland South America* ed. by Hein van der Voort & Simon van de Kerke, 9–24, Leiden: CNWS Publications.
- Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford: Stanford University Press.
- Grinevald, Colette. 2000. "A Morphosyntactic Typology of Classifiers". *Systems of Nominal Classification* ed. by Gunter Senft, 50–92. Cambridge: Cambridge University Press.
- Grinevald, Colette. 2002. "Nominal Classification in Movima". *Current Studies on South American Languages* ed. by Mily Crevels, Simon van de Kerke, Sérgio Meira & Hein van der Voort, 251–239. Leiden: CNWS Publications.
- Grinevald, Colette. 2004. "Classifiers". *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung*, vol. 2, ed. by Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas, with Wolfgang Kesselheim, 1016–1031. Berlin: Walter de Gruyter.

- Grinevald, Colette. 2007. "The Linguistic Categorization of Spatial Entities: Classifiers and other nominal classification systems". *The Categorization of Spatial Entities in Language and Cognition* ed. by Michel Aurnague, Maya Hickmann & Laure Vieu (eds.), 93–121. Amsterdam: John Benjamins. doi:10.1075/hcp.20.07gr
- Grinevald, Colette & Katharina Haude. 2013. "Class Terms, Classifiers or Noun Classes? The case of Movima (Bolivia)". [Hand-out for the 10th annual meeting of the Association of Linguistic Typology, Leipzig].
- Grinevald, Colette & Frank Seifart. 2004. "Noun Classes in African and Amazonian Languages". *Linguistic Typology* 8.243–285. doi:10.1515/lity.2004.007
- Guillaume, Antoine. 2008. *A Grammar of Cavineña*. Berlin: Mouton de Gruyter. doi:10.1515/9783110211771
- Hart, Raymond E. 1963. "Semantic Components of Shape in Amarakaeri Grammar". *Anthropological Linguistics* 5(9).1–7.
- Haude, Katharina. 2006. *A Grammar of Movima*. Ph.D. dissertation, Radboud Universiteit Nijmegen. http://webdoc.uhn.ru.nl/mono/h/haude_k/gramofmo.pdf. (Accessed: 27 October 2017.)
- Klein, Harriet E. Manelis. 2000. "Meronymy or Part-Whole Relations in Indigenous Languages of Lowland South America". *Indigenous Languages of Lowland South America* ed. by Hein van der Voort & Simon van de Kerke, 83–98. Leiden: CNWS Publications.
- Krasnoukhova, Olga. 2012. *The Noun Phrase in the Languages of South America*. Utrecht: LOT Publications.
- Kroeker, Menno H. 2001. "A Descriptive Grammar of Nambikuara". *International Journal of American Linguistics* 67(1).1–87. doi:10.1086/466446
- Lehmann, Christian. 2015. *Thoughts on Grammaticalization*, 3rd edition. Berlin: Language Science Press.
- Lichtenberk, Frantisek. 2009. "Attributive Possessive Constructions in Oceanic". *The Expression of Possession* ed. by William B. McGregor, 249–292. Berlin: Mouton de Gruyter. doi:10.1515/9783110213232.249
- Lowe, Ivan. 1999. "Nambikuara". *The Amazonian Languages* ed. by R. M. W. Dixon & Alexandra Y. Aikhenvald, 268–291. Cambridge: Cambridge University Press.
- Martines, George Vergés. 2007. *Aspectos semânticos dos nomes classificados em mundurukú*. M.A. thesis, Universidade de São Paulo.
- Messineo, Cristina & Ana Gerzenstein. 2007. "La posesión en dos lenguas indígenas del Chaco: toba (guaycurú) y maká (mataguayo)". *Linguas Indígenas Americanas (LIAMES)* 7.61–79.
- Mithun, Marianne. 1984. "The Evolution of Noun Incorporation". *Language* 60.847–894. doi:10.1353/lan.1984.0038
- Moore, Denny. 1984. *Syntax of the Language of the Gavião Indians of Rondônia, Brazil*. Ph.D. dissertation, The City University of New York.
- Payne, Doris L. 1986. "Noun Classification in Yagua". *Noun Classes and Categorization: Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983* ed. by Colette Craig, 113–131. Amsterdam: John Benjamins. doi:10.1075/tsl.7.10pay
- Perri Ferreira, Helder. 2009. *Los clasificadores nominales del yanomama de Papiu (Brasil)*. M.A. thesis, Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS), Mexico, D.F.
- Petersen de Piñeros, Gabriele. 2007. "Nominal Classification in Uitoto". *International Journal of American Linguistics* 73(4).389–409. doi:10.1086/523771

- Picanço, Gessiane Lobato. 2003. "Alienability and Argument Structure". Unpublished manuscript in possession of the author.
- Pires, Nádia Nascimento. 1992. *Estudo da gramática da língua jeoromixi (jabuti) – aspectos sintáticos das cláusulas matrizes*. M.A. thesis, Departamento de Linguística, Universidade Estadual de Campinas.
- Priest, Perry & Anne de Priest. 1985. *Diccionario siriono y castellano*. Cochabamba: Instituto Lingüístico de Verano.
- Ramirez, Henri. 1994. *Le parler yanomami des xamatauteri*. Ph.D. dissertation, Université de Provence, Centre d'Aix.
- Ribeiro, Eduardo & Hein van der Voort. 2010. "Nimuendajú Was Right: The inclusion of the Jabuti language family in the Macro-Jê stock". *International Journal of American Linguistics* 76(4).517–584. doi:10.1086/658056
- Rodrigues, Aryon Dall' Igna. 1997. "Nominal Classification in Kariri". *Opción* 13(22).65–79.
- Rodrigues, Aryon Dall' Igna. 1999. "Macro-Jê". *The Amazonian Languages* ed. by R. M. W. Dixon & Alexandra Y. Aikhenvald, 165–206. Cambridge: Cambridge University Press.
- Schermair, Anselmo E. 1962. *Vocabulario castellano-sirionó*. Innsbruck: Sprachwissenschaftliche Institut der Leopold-Franzens-Universität.
- Seifart, Frank. 2005. *The Structure and Use of Shape-Based Noun Classes in Miraña (North West Amazon)*. Ph.D. dissertation, Radboud Universiteit Nijmegen. http://pubman.mpdl.mpg.de/pubman/item/escidoc:402010:3/component/escidoc:402009/mirana_seifart2005_s.pdf. (Accessed: 27 October 2017.)
- Seifart, Frank & Doris Payne. 2007. "Nominal Classification in the North West Amazon: Issues in areal diffusion and typological characterization". *International Journal of American Linguistics* 73(4).381–387. doi:10.1086/523770
- Telles, Stella. 2002. *Fonologia e gramática Latundê/Lakondê*. Ph.D. dissertation, Vrije Universiteit Amsterdam.
- van der Voort, Hein. 2004. *A Grammar of Kwaza*. Berlin: Mouton de Gruyter. doi:10.1515/9783110197280
- van der Voort, Hein. 2005. "Kwaza in a Comparative Perspective". *International Journal of American Linguistics* 71(4).365–412. doi:10.1086/501245
- van der Voort, Hein. 2006. "Construções atributivas em Kwazá". *Boletim do Museu Paraense Emílio Goeldi* 1(1).87–104. doi:10.1590/S1981-8122006000100007
- van der Voort, Hein. 2009. "Possessive Expressions in the Southwestern Amazon". *The Expression of Possession* ed. by William B. McGregor, 343–388. Berlin: Mouton de Gruyter. doi:10.1515/9783110213232.343
- van der Voort, Hein. 2015. "Sistemas de classificação nominal no Sudoeste amazônico". *Moara – Revista Eletrônica do Programa de Pós-Graduação em Letras* 43(2).5–22. <http://www.periodicos.ufpa.br/index.php/moara/article/view/3836/3861>. (Accessed: 27 October 2017.)
- Vasconcelos, Ione P. 2003. *Aspectos fonológicos e morfofonológicos da língua Aikanã*. Ph.D. dissertation, Universidade Federal de Alagoas.
- Vengoechea, Consuelo. 2005. "Morphosyntax of Muinane: Typological remarks". *Amerindia, Revue d'Ethnolinguistique Amerindienne* 29(30).119–140.
- Vengoechea, Consuelo. 2009. "Categorías léxicas y clasificación nominal en muinane". 53avo Congreso Internacional de Americanistas, México, D.F., 22 July 2009 [handout].
- Vuillermet, Marine. 2012. *A Grammar of Ese Ejja, a Takanan Language of the Bolivian Amazon*. Ph.D. dissertation, Université Lumière Lyon 2.

Nominal and verbal classification

A comparative perspective

Walter Bisang

University of Mainz

The present paper starts from the observation that classification is cross-linguistically very widespread in the domain of the noun and rather rare in the domain of the verb. It argues that this asymmetry is not arbitrary. It is motivated by two conditions: (i) the markers used for classification should not interfere with other grammaticalization processes, and (ii) the use of a classification marker in later processes of grammaticalization should not impair the initial classification system. Since these conditions are violated more easily with verbs than with nouns, classification is more persistent in nominal classification than in verbal classification. The paper looks at compounding, which is the common denominator of many nominal and verbal classification systems and thus provides the starting point from which the differences between the two types of classification can be observed. It then looks at the semantics involved in nominal and verbal classification, and it shows that the two conditions are more easily violated in the domain of the verb than in the domain of the noun.

1. Introduction: Basic idea of the paper

Classification as it is manifested in classifier systems is found with nouns and verbs. While nominal classification is widespread in various types across the world and has been the topic of extensive discussion in the literature (Corbett 1991 on gender; Aikhenvald 2000 on a typology of classification in general; Grinevald 2000; Bisang 1999, 2002, 2012), verbal classification seems to exist in remarkably fewer languages and is discussed less frequently (Schultze-Berndt 2000, 2003; McGregor 1990: 195–203, 557–572; 2002). Starting from morphological compounding as an important historical source of both types of classifier systems, the present paper argues that this asymmetry is functionally motivated rather than the result of a historical accident.

Compounding is not the overall common source of systems of classification but is the common ground for comparing nominal and verbal classification and the processes of grammaticalization that generate them. Thus, it provides a common floor for understanding why classification is more prominent in the domain of the noun than of the verb. In the nominal domain, there are two classifier systems that are based on compounds, i.e., the numeral classifier systems of Southeast Asian languages (Bisang 1999) and the noun classifier systems of Mesoamerican languages (Craig 1986a,b). In the domain of the verb, classification is reflected in various compound verb constructions of Northern Australian languages (Schultze-Berndt 2000, 2003; McGregor 2002). The cross-linguistic range of grammatical functions that can be expressed by classifiers in both domains is impressive. In the nominal domain, classifiers can mark individuation, definiteness, possession, contrastive focus, singulative, size and politeness (cf. §3.1.1). In the verbal domain, classification is associated with categories such as Aktionsart (telicity), direction of movement, degree of activity (dynamic vs. stative), valency and subject/actor control (cf. §3.2).

Since the present paper concentrates on compounding, it excludes systems of event classification frequently found in East and mainland Southeast Asia (Gerner 2009) and the use of nominal classification markers on verbs as described by Sagna (2008) on Gújjolaay Eegimaa (Niger-Congo, Atlantic-BAK, spoken in Casamance, Senegal). Systems of this type are not genuine instances of verbal classification because they depend on nominal classification systems. The occurrence of event classification seems to be limited to languages which also have numeral classifiers. In the case of Gújjolaay Eegimaa, the dependence on the nominal system is even more straightforward since this language uses parts of its noun class system (ten out of its fifteen noun-class prefixes) to form non-finite verb forms (infinitives). These systems are thus excluded not only because they do not belong to compounding but also because they are not genuinely verbal. Interestingly enough, there does not seem to be a system in which verbal classification had an impact on the system of nominal classification. This asymmetry may be another indicator of the comparatively weak status of verbal classification.

If one looks at the functions in which nominal classification is involved, one can see that it is an integrative part of grammatical systems that express the functions mentioned above. This is different for the categories belonging to the verbal domain. From a cross-linguistic perspective, the above verbal functions are rarely associated with classification even though there are languages in Northern Australia and in South America in which this is the case. As is hypothesized in this paper, this asymmetry between nominal and verbal classification is due to two conditions which determine whether classification is maintained in successive processes of grammaticalization:

- i. The individual elements used for classification should not interfere with other processes of grammaticalization that are also based on the same semantics or the same source of grammaticalization.
- ii. The use of a classification marker in later processes of grammaticalization should not impair the initial classification system.

These two conditions do not represent any deep rooted cognitively or universally motivated principles. Still, they are important for processes of grammaticalization. If they are violated, they support the erosion of the initial grammatical system and its independence from other systems. Thus, the initial grammatical system may first be blurred until it gets lost completely at a final stage. As will be shown, the domain of the noun complies very well with both conditions, while the domain of the verb does not. As a consequence, the emergence of verbal classification is not impossible (as is after all shown by data discussed here) but the violation of the two conditions detracts from its stability over time and thus reduces the likelihood of its occurrence.

The paper is organized as follows. Section 2 deals with classification and compounding in the domain of the noun (§2.1) and of the verb (§2.2). It discusses how compound-based processes of grammaticalization create numeral classifiers (Vietnamese [Austro-Asiatic: Mon-Khmer] and Hmong [Hmong-Mien]; §2.1.2), noun classifiers (Hup [Nadahup, NW Amazonian], Jacalteco [Mayan]; §2.1.3) and even noun-class markers (Miraña [Witotoan, Colombian Amazon region]; §2.1.4). Even though most gender or noun-class systems are not based on the grammaticalization of other classifier systems, Miraña (Seifart 2005) shows clear connections between numeral classifier systems and noun-class or gender systems. Compounding in the domain of the verb (§2.2) is illustrated by the example of two North Australian languages (Nyulnyul [Nyulnyulan] and Gooniyandi [Bunuban]) and by a language from Bolivia (Mosetén [Barbacoan, Ecuador]). Section 3 deals with the functions of classifiers in the individual classifier systems presented in §2. A survey of the semantics involved in classifier systems is important for understanding the consequences of conditions (i) and (ii). It provides the information needed for checking whether the semantics of the individual markers used for classification also operate in other processes of grammaticalization (condition (i)) and to what extent the semantic distinctions of a classifier system as a whole remain unimpaired if they are integrated into grammatical categories as different as definiteness/specificity and possession in the nominal domain or telicity and degree of activity in the verbal domain (condition (ii)). Section 3.1 presents the nominal domain in which classifiers mark individuation and identification and the categories that follow from these functions. Similarly, §3.2 deals with the grammatical categories involved in classifiers of the verbal domain. An additional section (§3.3)

highlights some differences between verbal classification, verb serialization and light verbs. On the basis of the data presented in Sections 2 and 3, conditions (i) and (ii) and their consequences for the persistence and stability of classification through successive processes of grammaticalization in the nominal and in the verbal domain are the topic of §4. The paper ends with a short conclusion in §5.

2. Classification and compounding

2.1 Nominal compounds and nominal classification

2.1.1 *General remarks*

The historical relationship between classifiers and nominal compounds is a well-known phenomenon. It has been described for nominal classifiers in Jacaltec (Craig 1986a, 1986b, Grinevald 2000: 59–61) and for numeral classifiers in East and mainland Southeast Asian languages (DeLancey 1986; Bisang 1993, 1999). The compounds from which the process of development starts out consist of a morphological head noun modified by one or more additional lexemes. The head noun, which is called ‘class term’ (Haas 1942; DeLancey 1986; Craig 1986a, 1986b) or ‘class noun’ (Bisang 1993, 1999), is characterized by its semantic generality which is often represented by its reference to higher taxonomic levels. Another semantic relation between the head noun and its modifier is that of meronymy. In this case, the head noun represents the part, while the modifier refers to the whole.

Grinevald (2000: 61–62) describes classifiers as an intermediate lexico-grammatical system situated between the lexicon and grammar. On the lexical side, we find measure terms and class terms (class nouns), while noun-class and gender systems take the grammatical side of the continuum – see Figure 1:



Figure 1. Classifiers between lexical and grammatical function (Grinevald 2000: 61)

The development from class terms (class nouns) to classifiers as presented in the left part of Figure 1 is widely attested cross-linguistically and is the topic of this section. The grammaticalization scenario in the right part of Figure 1, however, is more questionable, at least from the point of view of Indo-European type gender systems.

There does not seem to be a universal cline along which gender or noun-class systems develop out of classifier systems. But there are languages such as Miraña (Seifart 2005) in which such a development looks plausible. From this background, the development of classification systems and their integration into the marking of a wide range of nominal grammatical categories are discussed for two different types of classifier systems and for Miraña. Section 2.1.2 deals with numeral classifiers in mainland Southeast Asian languages, while §2.1.3 discusses noun classifiers in Hup and Jacaltec. Miraña is the topic of §2.1.4.

Grinevald (2000) argues that the emergence of different types of classifier systems is based on particular semantic-cognitive categories (Grinevald 2000). Adopting Denny's (1976) view that classifiers reflect the way in which speakers interact with objects, she claims that numeral classifiers mainly develop from nouns denoting physical properties (shape, size, flexibility, etc.), while noun classifiers emerge from contexts in which material and essence properties prevail – “the substance out of which the objects are made” (Craig 1986a: 272). Even though this scenario can account for various diachronic facts concerning classifier systems, the data discussed in this paper show that the development of classification-based grammatical markers out of nominal compounds does not depend on particular semantic-cognitive categories.

2.1.2 *Nominal compounds and numeral classifiers: mainland Southeast Asia*
Mainland Southeast Asian languages are characterized by a wide variety of compounding structures. The one that is crucial for classifier development consists of class nouns or class terms in the head position, followed by more specific modifiers. In Example (1) from Hmong, the first part of the nominal compound consists of the class nouns *txiv* “fruit”, *ntoo* “tree” and *zaj* “story, song”, which are specified by one or more lexemes in the position(s) following them. As a result of this process, the compound noun stands on a lower taxonomic level than the head noun alone:

- (1) *Class nouns in Hmong*
- a. *txiv* “fruit”: *txiv kab ntxwv* “orange”
txiv ntseej “chestnut”
txiv lws qaum “lemon”
txiv txhais “mango”
 - b. *ntoo* “tree”: *ntoo kab ntxwv* “orange tree”
ntoo ntseej “chestnut tree”
ntoo peeb lab “areca palm”

- c. *zaj* “story, song, etc.” *zaj dab neeg*
 CN spirit human.being
 “legend”
zaj teev ntuj
 CN pray heaven
 “prayer”
zaj tshoob
 CN marriage/wedding
 “wedding song”
zaj hu nkauj
 CN sing song/melody
 “song”

The development of numeral classifiers from class nouns in mainland Southeast Asian languages has been described in detail for taxonomic and meronymic relations by Bisang (1999). For that reason, I provide only a short summary of how nominal compounding and taxonomy are related to numeral classifiers by some examples from Vietnamese and Hmong. These languages are selected because they take two rather divergent positions in the degree to which their class nouns grammaticalized into numeral classifiers. Vietnamese is situated at the low end of the scale of grammaticalization, while Hmong classifiers have reached a high degree of grammaticalization. This becomes evident from a comparison of the following parameters:

- i. obligatoriness of classifier use;
- ii. presence of a special syntactic position (or a special construction) for the classifier;
- iii. degree of control of classifier selection by the lexicon;
- iv. number of classifiers.

As can be seen from the distribution of these parameters (Table 1), the first three parameters are clearly negative or at least less developed in Vietnamese, while they are fully established in Hmong.

Table 1. Classifier systems and their degree of grammaticalization

	Vietnamese	Hmong
i. obligatoriness of classifier use	(+)	+
ii. special syntactic position	–	+
iii. classifier selection by lexicon	±	+
iv. number of classifiers	10 plus many more	8

In the remainder of this section, each of these parameters are discussed individually for both of the two languages.

Vietnamese has eleven core classifiers which represent the defaults, i.e., the semantically most neutral cases (Löbel 1996: 171–172; 2000). As can be seen from (2), the categories of animacy (non-living, tree/plant, non-human, human) and shape (three-dimensional: fruit/round; two-dimensional: flat piece of material, sheet of paper; one-dimensional: hair/thread) are prominently represented:

- (2) *Core classifiers in Vietnamese:* (Löbel 1996: 172)¹
- | | |
|--------------|---|
| <i>cái</i> | “a non-living thing” |
| <i>cây</i> | “a tree or plant, a stick-shaped or plantlike object” |
| <i>chiếc</i> | “individual item of floating object, vehicles” |
| <i>con</i> | “a living thing that is not human” |
| <i>hòn</i> | “a stone or stonelike thing” |
| <i>quả</i> | “a fruit, a round, globular object” |
| <i>quyển</i> | “a volume” |
| <i>sợi</i> | “a hair, thread, cord, etc.” |
| <i>tấm</i> | “a flat piece of material” |
| <i>tờ</i> | “a sheet of paper, document” |
| <i>người</i> | “man [for humans]” |

In addition to the above core classifiers, it is possible for almost any noun to be interpreted as a classifier if the object expressed by the head noun is looked at from a certain, more specific perspective. In the case of nouns denoting humans, the default classifier is *người* “man” as in (3a). In the case of a more specific perspective, there are many more options as illustrated by (3b) to (3d). In (3b), the noun *đồng nghiệp* “colleague” is discussed in terms of friendship. In (3c) and (3d), the contexts in which this noun is described are rank and age, respectively:


- (3) *Vietnamese:* (Löbel 1996: 174)
- | | |
|----|--|
| a. | <i>hai người đồng nghiệp</i> |
| | two CL:MAN colleague |
| | “two colleagues [neutral]” |
| b. | <i>hai bạn đồng nghiệp</i> |
| | two CL:FRIEND colleague |
| | “two colleagues [good friends]” |
| c. | <i>hai vị đồng nghiệp</i> |
| | two CL:RANK colleague |
| | “two colleagues [very respectful]” |
| d. | <i>hai ông đồng nghiệp</i> |
| | two CL:GRANDFATHER colleague |
| | “two colleagues [respectful for two colleagues of higher age]” |

1. Löbel (1996) does not mention the classifier *người* “man” in her list even though she treats it as a core classifier, too.

Example (3) illustrates the basic openness of the number of classifiers (parameter (iv)) in Vietnamese and it shows that the selection is not rigidly determined by the lexicon (parameter (iii)). The following example from Löbel (1996: 171) shows that Vietnamese classifiers lack obligatoriness – there are contexts as the one in (4a), in which the classifier is not allowed. In this example, there is no classifier between the numeral *ba* “three” and the noun *phòng* “room” because there is no need to individuate that concept. What matters is only the overall property of “having three rooms” that determines the head noun *nhà* “house”. This is different in Example (4b), in which the classifier must occur because the three rooms are individuated:

- (4) Vietnamese: (Löbel 1996: 171)
- a. *nhà [ba phòng]*
house [three room]
“a three room house”
 - b. *nhà [với ba cái phòng]*
house [with three CL room]
“a house with three rooms”

The examples discussed so far from the perspective of parameters (i), (iii) and (iv) illustrate the relatively low degree of grammaticalization of Vietnamese classifiers fairly well. An even more clear-cut indicator of low grammaticalization is the lack of a special syntactic position for the classifier (parameter (ii)). In an open classifier system like the one of Vietnamese, basically any head noun N_1 that is taxonomically higher than another noun N_2 following it can be analyzed as a classifier under specific pragmatic conditions in which N_2 is profiled from the perspective of its superordinate concept. Thus, a sequence with two nouns as in (5a) can be reanalyzed as in (5b) as a classifier if it is preceded by a numeral (NUM):

- (5) a. NUM N_1 N_2 ==> b. NUM CL(= N_1) N_2
- 

taxonomic

As illustrated in (6) from Löbel (1996), the process of reanalysis described in (5) can be applied cyclically. The noun *cây* “tree/plant” takes the default classifier *cái* in (6a). In (6b), there is a taxonomic relation between *cây* “plant/tree” and *rau* “vegetable”. Thus, *cây* can be interpreted as a classifier of *rau* “vegetable”. In turn, the noun *rau* “vegetable” is analyzed as a classifier marking *cần* “celery” in (6c):

- (6) Vietnamese: (Truong 1970: 256)
- a. *một cái cây*
one CL tree/plant
“a tree/a plant”

- b. *một cây rau*
 one CL vegetable
 “a vegetable (as a plant)”
- c. *một rau cần*
 one CL celery
 “one celery (a vegetable)”

The lack of an established independent position for classifiers in Vietnamese does not only show that Vietnamese is still in a comparatively initial phase of classifier development, it also illustrates *in vivo*, how a classifier system develops through reanalysis of the head noun of a nominal compound.

In Hmong with its more developed classifier system, the process of ongoing classifier development is no longer active, but it can still be seen synchronically even though it is less visible than in Thai (cf. Bisang 1993, 1999). This is due to the fact that Hmong has only a small set of seven classifiers (parameter (iv)) based on the semantic criteria of animacy, shape, function and part of a pair (7) and that these classifiers cannot be associated with class nouns synchronically. In addition, classifier selection is rigidly determined by the lexicon (parameter (iii)).

- (7) *List of common Hmong classifiers:*
- | | |
|-------------------------------|------------------|
| [+ human], [+ animate]: | <i>leej, tus</i> |
| [- human], [+ animate]: | <i>tus</i> |
| [- animate]: | <i>lub</i> |
| [one-dimensional]: | <i>txoj, tus</i> |
| [two-dimensional]: | <i>daim</i> |
| [three-dimensional]: | <i>lub</i> |
| [function(tool, instrument)]: | <i>rab</i> |
| [one item of a pair]: | <i>txhais</i> |

Hmong classifiers have their own syntactic position as described in (8) and they are obligatory with numerals (9) (parameters (i) and (ii)):

- (8) [NUM CL] N
- (9) *Hmong:*
- a. *plaub tug tsov*
 four CL tiger
 “four tigers”
- b. **plaub tsov*
 four tiger

Evidence for the relationship between class nouns and numeral classifiers comes from a small set of additional marginal classifiers which are used with some specific

semantic domains. One of them is *zaj* “story, song”. It is used as a class noun as illustrated in (1c), and it occurs as a numeral classifier with nouns such as *npau suav* “dream” or *kwv txhiaj* “song, ballad, sung by boys and girls to each other”. If a noun beginning with the class noun *zaj* has to be determined by a classifier, the Hmong language can use the same lexeme, *zaj*, twice over – first in the classifier position and second in the class-noun position:

- (10) *Hmong*: (Bertrais-Charrier 1979: 567)
ob zag² zaj tshoob
 two CL CN weddings
 “two wedding songs”

This offers conclusive evidence that *zaj* performs both functions: that of a class noun and that of a classifier. The language seems, however, to be rather reluctant to exploit this possibility of using the same word twice. Thus, examples such as (11) are much more frequent than (10):

- (11) *Hmong*:
Nws hais ib zag teev ntuj.
 he say one CL/CN pray heaven
 “He is praying once [lit.: one prayer].”

2.1.3 Nominal compounds and noun classifiers

The relation between nominal compounding and noun classifiers is particularly evident in a language like Hup (Epps 2007), whose noun-classifier system is still in its infancy. In this language, nominal compounds consist of a word final head (N_2) and a preceding modifier (N_1) (Epps 2007: 108):

- (12) $N_1[\text{mod}] + N_2[\text{head}]$

This pattern is divided into two subparts which depend on the informational perspective from which the compounds are looked at. In one pattern illustrated by (13), N_1 is the ground and stands for given information, while N_2 represents the figure, i.e., new information. Thus, the different suffixes in the position of N_2 focus on the various parts of N_1 *pihít* “banana” that make up the plant whole. In (14), the figure is in N_1 and the ground is the suffix =*g’æt* “leaf” in the N_2 -position. Thus, the focus is on the properties of different kinds of leaves.

2. Since there are generally no final consonants in Hmong syllables, the syllable final position is used to mark tones by using some consonants. In this example, *zaj* changes its tone into *zag* because of the high level tone of *ob* “two”, which is responsible for tone sandhi. The same role also applies to the classifier *tus* in Example (9a). It appears in the form of *tug* after the tone of *plaub*.

- (13) *Hup*: (Epps 2007: 111)
píhít=g'æt “banana leaf”
píhít=teg “banana tree”
píhít=b'ák “clump of bananas”
píhít=tat “banana fruit”, etc.
- (14) *Hup* (Epps 2007: 111):
cug'ǎt “leaf (generic)”
píhít=g'æt “banana leaf”
b'ab'ǎt=g'æt “embauba leaf”
pehé=g'æt “palm leaf”, etc.

Since the figure always takes a stress accent, the two patterns are phonologically different. Pattern (13) takes the accent on N_2 , pattern (14) on N_1 . While both compound patterns are possible, only the stress pattern in (14) is possible with the N_2 -suffixes in the class-noun function. This pattern seems also to reflect the perspective that is more central to the lives of the *Hup* speakers. Epps (2007: 111) argues that this is due to the cultural relevance of the elements in N_2 :

The identity of a leaf, stick, seed, or other plant part is a frequent topic of discussion on the daily life of the *Hupd'əh* ..., since these are the raw materials that the *Hupd'əh* use to manufacture the things they need, and are also used for foods and medicines.

If the classifier system of *Hup* emerges from the way in which the speakers manipulate objects of daily life, this confirms Denny's (1976) view that the semantics of classifiers is “concerned with objects as they enter into human interactions.” In terms of Epps (2007: 119), this means that “a *Hup* speaker is much more likely to be concerned about the identity of a leaf or stick, because that is crucial to what he or she can do or make with that object.”

The way in which humans interact with their environment is crucial for many more developed noun-classifier systems. One of them is the system of *Jacalteco* as described by Craig (1986a, 1986b). In this language, there are two subsystems. One is concerned with social interactions, the other with physical and functional interactions. Each consists of exactly twelve noun classifiers.

Jacalteco class nouns are free lexemes which occur before the noun they classify. The following example illustrates the use of the classifiers *ix* “girl” for female non-kin (number 5) and *ixim* “corn” for things made corn (number 16) (cf. Table 2):

- (15) *Jacalteco*: (Craig 1986a: 264)
swatx' ix ix ixim b'itx.
 made CL girl CL tamale
 “The girl made the tamales.”

Since it is impossible to provide a detailed account of how the Jacaltec noun-class system reflects the interaction of the speakers with their environment (for that purpose, cf. Craig 1986a, 1986b), I outline briefly how subsystem II works. The assignment of subsystem II classifiers to head nouns is motivated by cultural factors. The classification system “takes into account notions such as the amount of control that Jacaltecs exercise over objects, and the degree of familiarity they experienced with the objects, at the time the classification system was developing” (Craig 1986a: 276).

The twelve classifiers of subsystem II in Table 2 can be divided into two subtypes. There is a basic set of five general classifiers which cover large natural domains of human interaction: *no7* “animal”, *te7* “plant”, *ch'en* “rock”, *ha7* “water” and *k'a7* “fire”. The remaining seven classifiers are specific classifiers. They cover specific objects such as *metx'* “dog”, *ixim* “corn”, *tx'al* “thread”, *tx'an* “twine”, etc. The semantic relation between the head noun and the noun classifier is either taxonomic or what Craig (1986a: 279) calls ‘chaining’, i.e., a relationship characterized by the primary substance of which the head noun is made. In the case of the five general classifiers, the relation is clearly taxonomic inclusion, i.e., the head noun represents a subclass of the class denoted by the classifier. Taxonomy also comes in with certain specific classifiers such as *metx'* “dog” or *ixim* “corn”, although it does so on a lower biological rank. Other specific classifiers such as *tx'al* “thread”, *tx'an* “twine” clearly highlight the material of which the head noun is made of and thus establish a chaining relation between the classifier and its head noun (*tx'an* “twine” refers to a product of the agave plant, *tx'al* “thread” to a product of the cotton plant).

Table 2. Noun classifiers in Jacaltec (Craig 1986a: 266–267)

Subsystem I: Social interaction			Subsystem II: Physical/functional interaction		
1.	cumam	male deity	13.	no7	animal
2.	cumi7	female deity	14.	metx'	dog
3.	ya7	respected human	15.	te7	plant
4.	naj	male non-kin	16.	ixim	corn
5.	ix	female non-kin	17.	tx'al	thread
6.	naj ni7an	young male non-kin	18.	tx'ang	twine
7.	ix ni7an	young female non-kin	19.	k'ap	cloth
8.	ho7	male kin	20.	tx'otx'	soil/dirt
9.	xo7	female kin	21.	ch'en	rock
10.	ho7 ni7an	young male kin	22.	atz'am	salt
11.	xo7 ni7an	young female kin	23.	ha7	water
12.	unin	infant	24.	k'a7	fire

The interactional aspect of classification shows up in both types of classifiers. While the general classifiers refer to “ecologically and culturally important and common

objects manipulated by the Jacaltecs, the specific classifiers identify objects of particular cultural significance to the Jacalteco community of speakers” (Craig 1986a: 279). Of particular importance for the Jacalteco culture are material classifiers. As is shown very convincingly by Craig (1986a), the specific classifiers expressing a chaining relation classify exactly those objects of Jacalteco agricultural life which are essential for the subsistence and the identity of the members of this culture.

2.1.4 *Noun-class markers in Miraña*

In Miraña, there are four different types of nouns. Three of them are associated with the use of class markers in various ways (Seifart 2005: 110–111). Optionally classified nouns can be combined with a class marker, while obligatorily classified nouns must take a class marker. Repeater nouns consist of a single root that is interpreted as singular in its bare form and thus differs from the uncountable non-classified bare nouns. Finally, non-classified nouns are not countable, and they consist of a single noun root with no class marker. Of particular interest for the present paper are optionally classified nouns. They are briefly discussed below to show how class markers are related to compound morphology and derivation. Their grammatical function is the subject of §3.1.4.

Miraña has a large number of specific class markers. Eighteen of them are monosyllabic, and another forty eight are polysyllabic (Seifart 2005: 84–94). Optionally classified nouns express a rather general meaning that is specified by the class marker, which is suffixed to them. A noun such as *úhi* “banana” refers to the concept of ‘banana’ in a very broad sense which can be specified by adding a class marker. The noun *úhi* “banana” occurs with no less than seven class markers, three of them are illustrated in the following example:

- (16) *Miraña*: (Seifart 2005: 114)
- a. *úhi-ʔo*
banana-CL:3DIM
“banana fruit”
 - b. *úhi-ko*
banana-CL:1DIM
“banana plant”
 - c. *úhi-pájhko*
banana-CL:LIQUID
“banana drink”

At least some of the specific class markers are clearly related to nouns (Seifart 2005: 89). Thus, the marker *-ha* “CL:COVER”, which specifies clothes, houses and other types of cover, corresponds to the noun *ha* “house”. The class marker *-huu* “CL:TUBE” is related to the noun *í-ʔhuu* [POSS:3-mouth] “his/her/their mouth, story,

language”. Similarly, *-mi* “CL:TRANSPORT” is derived from *mi:ne* “canoe”, and *-:baj* “CL:CONTAINER” is based on *báne:baj* “pool”.

The correlation between class markers and nouns is particularly evident in the case of repeaters. Repeater nouns do not take classifier suffixes themselves, but they occur in the positions in which class markers occur. In (17), the repeater noun *báhuu* “forest” agrees with the copula verb *ihka* “be” and with the numeral *tsá-* “one” (for more on agreement in Miraña, cf. §3.1.4):

- (17) *Miraña*: (Seifart 2005: 94)
- | | | | |
|--------------|-------------------|------------------|----------------|
| <i>ó-ɔdi</i> | <i>ihka-báhuu</i> | <i>tsá-báhuu</i> | <i>(báhuu)</i> |
| 1.SG-POSS | COP-CL:FOREST | one-CL:FOREST | forest |
- “I have one (stretch of) forest.”

2.2 Verbal compounds and verbal classification

Verbal classification is found in many languages of northern Australia (McGregor 1990, 2002, this volume; Wilson 1999; Schultze-Berndt 2000, 2003) and in some languages of South America as in Mositén (Sakel 2004, 2007) and Tsafiki (a Barbacoan language of Ecuador; Dickinson 2002). This section briefly introduces verbal classification in the two Australian non-Pama-Nyungan languages Nyulnyul (McGregor 2002, this volume) and Gooniyandi (McGregor 1990, 2002) and in Mositén, a language spoken in the foothills of the Bolivian Andes (Sakel 2004, 2007). Notice that verbal classifiers will be represented by capital letters.

Nyulnyul is one of about ten languages in the Nyulnyulan family. In these languages, verb classification is coded in compound constructions that consist of two parts, the predicational verb dubbed ‘preverb’ or ‘uninflecting verb’ by McGregor (this volume) plus the inflecting verb that classifies actions by subsuming them under certain semantic categories. Preverbs obligatorily form compound verb constructions with an inflecting verb, while inflecting verbs can occur alone in simple verb constructions. In some Nyulnyulan languages, the preverb can occur without an inflecting verb in some grammatical contexts as e.g. in non-finite clauses. In general, however, verb classification is obligatory. In the following example from Nyulnyul, the inflecting verb *-j* “say” is used in a single verb construction in (18a). In (18b) we find the same verb in a compound verb construction with the preverb *kurd* “hide”:

- (18) *Nyulnyul* (McGregor, this volume):
- a. *irr-in* *i-li-rr-j-jan.*
- 3AUG.CRD-ERG 3NOM-IRR-AUG-say-1MIN.OBL
- “They might say to me.” Or “They might tell me.”

- b. *bin wamb yil jin kurd i-n-d-in.*
 this man dog 3MIN.OBL hide 3NOM-CM-say-PRS
 “This man’s dog is hiding.”

In the compound verb construction of (18b), the inflecting verb *-j* “say” has the function of assigning the preverb *kurd* “hide” to the category of ‘dynamic event’. The details of the classification functions of inflecting verbs are discussed in §3.2. Here I briefly outline the overall number of inflecting verbs and the relatively small list of the most prominent inflecting verbs (for the data, cf. McGregor, this volume). Eastern Nyulnyulan languages have about one hundred inflecting verbs, while the western languages show up to about two hundred and fifty. About a dozen of these verbs are productive and thus belong to the primary classifying inflecting verbs. In Nyulnyul, McGregor (this volume) mentions the following ten primary inflecting verbs (inflecting verbs are represented in capital letters): *-N* “be”, *-JID* “go”, *-KAL* “wander”, *-K* “carry”, *-R* “pierce”, *-W* “give”, *-NY* “get”, *-M* “put”, *-BARNJ* “exchange” and *-J* “say”, cf. Example (18b).

In Gooniyandi (McGregor 1990: 195–203, 557–572; 2002), the overall structure of the VP can be represented as in (19). The position that is crucial for verbal classification is marked by X, which stands for the classifier. As can be seen from the fact that it is not in brackets it is obligatory together with information on tense and nominative pronominals (NOM:PRO):

- (19) Stem–(Aspect)–(Mood)–TNS + (ACC:PRO)|NOM:PRO +
 + X–(Mood)–(OBL:PRO)–(NO)

The elements taking the position of the classifier in X consist of twelve bound verbal classifiers derived from generic verbs that occur obligatorily with other verbs in the stem-position (the coverbs). They classify the event denoted by the stem by indicating the type of process to which it belongs or by indicating “the category into which the stem is assigned in a particular instance of use, and correspondingly, the category to which the referent event is considered to belong” (McGregor 2002: 44). Since most stems occur with more than one verbal classifier, verbal classifiers do not create disjoint subclasses of verbal lexemes. The choice of a given verbal classifier rather determines the process type under which a given lexical item is subsumed in an utterance.

The following two examples illustrate the use of the verbal classifiers. In (20), the verbal classifier *-ARNI*₂ assigns the action of seeing (*mila-*) to the reflexive and reciprocal process type. The use of the classificatory verb *-A* in (21) subsumes the action of seeing under the category of atelic actions (‘extendible’ actions in terms of McGregor 1990, 2002; the semantics is discussed in §3.2):

- (20) *Gooniyandi: Classifier -ARNI₂, Reflexive/reciprocal:* (McGregor 2002: 63)
nganyi-ngga mila-ng-arni
 I-ERG see-1SG.NOM-ARNI₂
 “I saw myself.”
- (21) *Gooniyandi: Classifier -A, transitive:* (McGregor 2002: 63)
nganyi-ngga wayandi mila-l-ø-a
 I-ERG fire see-1SG.NOM-3SG.ACC-A
 “I saw a fire.”

As pointed out above, the twelve verbal classifiers are derived historically from generic verbs. The following list illustrates the verbal basis of these markers:

Table 3. Gooniyandi classifiers derived from verbs (based on McGregor 1990: 195)

Form of verbal classifier	Meaning of the corresponding verb
-A	“extend”
-ADDI	“put”
-DI	“catch”
-MI	“effect”
-BINI	“hit”
-I	“go, be”
-BINDI	“get, become”
-ANI	“accomplish”
-ARNI ₁	“emerge”
-BIRLI	“consume”
-ARNI ₂	“Extendible: Reflexive/Reciprocal”
-MARNI	“Accomplishment: Reflexive/Reciprocal”

Mosetén (Sakel 2004, 2007) compound verb structures significantly resemble coverb-verb constructions in Australian languages. There is a small set of eight verbs called ‘verbness markers’ by Sakel (2007) which correspond to the inflecting verbs in Australian languages and an open class of verbs which is comparable to the uninflected verbs. As in the case of Example (18) from Nyulnyul, there is a set of elements like *yi* ‘say, DO/BE’ which can either occur independently as finite verbs as in (22a) or as part of the compound verb construction as in (22b). In the latter case, *-yi* in the function of the verbness marker immediately follows the uninflected verb *sak-* ‘leave’ and forms a coherent complex predicate which together takes prefixes and suffixes for expressing various grammatical categories.

- (22) *Mosetén:* (Sakel 2007: 317)
 a. *Yäe yi.*
 1.SG.M say[1]M.SG
 “I say (something).”

- b. *Yäe sak-yi.*
 1.SG.M leave-DO/BE[3]M.SG
 “I leave.”

Six of the eight verbness markers can be associated with independent finite verbs, or ‘simple verbs’ in terms of Sakel (2007), and two verbness markers (*-i-* and *-wi-*) only occur in compound constructions. Finally, two elements which will not be further discussed occur as simple verbs but cannot be employed as verbness markers (*ji-* “pass” and *ban-* “go”). Table 4 lists the eight verbness markers and shows the six correspondences with simple verbs discussed by Sakel (2007). Simple verbs whose connection with a verbness marker is not entirely clear are put in square brackets, while the clear correspondences are not marked. What is remarkable is that the meaning of the simple verbs and the verbness marker they are related to sometimes differ. This is the case with the verb *yi* illustrated in (22). In its function as a simple verb it means “say”. If it is used as a verbness marker, it denotes either dynamic events (“DO”) or stative events (“BE”). The clearest instances of a semantic correspondence between the verbness marker and the simple verb are those with minimal differences in semantic meaning, i.e., the elements (*-*)*tyi-* “PUT, give, put” and (*-*)*jo-* “BECOME, become”.

Table 4. Verbness markers and their correspondences to simple verbs in Mosestén (data from Sakel 2007: 318, 325)

Verbness marker	Simple verb
<i>-tyi-</i> “PUT”	<i>tyi-</i> “give, put”
<i>-jo-</i> “BECOME”	<i>jo-</i> “become”
<i>-yi-</i> “DO/BE”	<i>yi-</i> “say”
<i>-ki-</i> “BE”	[<i>ka-</i>] “bring there”
<i>-ti-</i> “DO”	[<i>ti-</i>] “bring here”
<i>-chhi-</i> “MOVE/HAVE”	[<i>chhi-</i>] “grab”
<i>-i-</i> “BE/HAVE”	–
<i>-wi-</i> “SEE/HEAR”	–

The functional details of the verbness markers are discussed in §3.2. What can be seen quite clearly is that simple verbs and verbness markers are related and that the combination of the two forms a coherent compound, i.e., a compound verb construction.

3. The functional range of verbal and nominal classification

3.1 The functional range of compound-related nominal classification

3.1.1 *General remarks*

Nominal classification can be employed for two pragmatic functions, individuation and identification (Bisang 1999, 2002). In the case of individuation, classification is used to profile the conceptual boundaries of a concept expressed by a noun in order to make it countable. This is accounted for in terms of atomization in Chierchia's (1998) formal approach. The lack of a distinction between count nouns and mass nouns in East and mainland Southeast Asian languages – or, to put it in terms of Chierchia (1998), the absence of a distinction between sets of atoms and sets of sums of atoms – creates the need of an independent level of atomization at which it is possible to produce atoms that can be counted. In the case of identification, classification is employed to narrow down the search space for finding an object in space. A particular sensory perception and its properties are compared to the properties of other sensory perceptions in order to identify that particular perception by subsuming it under a certain concept and thus distinguishing it from other concepts that are equally present within the speech situation in which a given utterance is made. Both functions set the pragmatic stage for processes of grammaticalization which end up with the following grammatical functions and categories (for more details, cf. Bisang 2002):

- Individuation:
 - Counting (numeral classifiers)
 - Singulative
- Identification:
 - Definiteness/specificity
 - Contrastive focus
 - Anaphoricity
 - Possession (identification of the thing possessed)
 - Relative clause (supporting the restrictive function of the relative clause)

As will be seen in the coming three sections, the classifier systems of the languages examined in §2 cover all of these functions.

3.1.2 *The functional range of numeral classifiers in mainland Southeast Asian languages*

Even though most languages of East and mainland Southeast Asia are characterized by the presence of numeral classifiers, there is enormous cross-linguistic variation if one looks at the overall functional range covered by their classifier systems. The minimal function attested in all languages is individuation in the context of

numerals and some other quantificational expressions. In Vietnamese,³ this is the only function of classifiers. In Mandarin Chinese, classifiers are used with numerals and with demonstratives. Table 5 shows that these two functions are only the tip of the iceberg – numeral classifiers can express many more functions in certain languages.

Table 5. Grammatical functions of classifiers (Bisang 2002; on Weining Ahmao, cf. Wang 1957, 1972; Gerner & Bisang 2008)

	Vietnamese	Mandarin	Thai	Hmong	Weining Ahmao
Occurrence with numerals	+	+	+	+	+
Occurrence with demonstratives	–	+	–	–	–
Marking of possession	–	–	–	+	+
Definiteness/specificity	–	–	(+)	+	+
Contrastive focus (with ADJ, DEM)	–	–	+	–	–
Singulative (in [N CL DEM])	–	–	+	–	–
Size (augmentative, diminutive)	–	–	–	–	+
Gender of speaker/politeness	–	–	–	–	+

Since the use of classifiers with numerals and with demonstratives is well known from the literature, these first two functions in Table 5 will not be illustrated. The other functions are briefly discussed in the remainder of this section.

If classifiers are used in possession constructions they usually express a certain conceptual distance between the possessor and the possessed. For that reason, they are obligatory in alienable possession (23) but they can be omitted with inalienable possession (with the exception of body parts) and in instances in which the possessional relation is presupposed and does not need to be established by the classifier (cf. Bisang 1993). In (24), the classifier is omitted because of the closeness of the relation between possessor and possessed. If the classifier *tus* occurs in such a construction, it is to express contrast (e.g. my elder brother in contrast to your elder brother).

- (23) *Hmong*:
kuv rab riam ntaj
 I CL sword
 “my sword”

3. Vietnamese numeral classifiers do not exclusively express definiteness. As was shown by Löbel (1996: 228–229; cf. Bisang 1999), the construction [CL + N] can have either definite or indefinite interpretation, depending on context.

- (24) *Hmong*:
kuv tij
 I elder.brother
 “my elder brother”

Definiteness is expressed by the classifier in various mainland Southeast Asian languages. In many Sinitic languages (cf. Li & Bisang 2012 on Mandarin, Wu and Cantonese), the classifier in the [CL + N] construction can be interpreted as definite or indefinite, depending on the position of the [CL + N] construction relative to the verb. In Thai, definiteness is limited to the adjective construction (for details, cf. Hundius & Kölver 1983; Becker 2005). Finally, Hmong classifiers have developed into fully fledged markers of definiteness in the [CL + N] construction even though there are alternative expressions with the demonstrative construction and the classifier is not obligatory as is the case with English articles. Example (25) illustrates the classifier *tus* in its definite function:

- (25) *Hmong*: (Mottin 1980: 200)
Thaum ub muaj ob tug niam txiv. Tus txiv tuag lawm.
 Long.ago there.are two CL wife husband CL husband die PFV
Tus niam quaj quaj, nrhiav nrhiav tsis tau tus txiv.
 CL wife cry cry look.for look.for NEG get CL husband
 “Long ago there were a wife and a husband. The husband died. The wife was crying a lot but no matter how hard she tried she was not able to find her husband.”

Other functions of the classifier are contrastive focus and singulative. These functions can be expressed in a limited set of constructions in Thai (cf. Hundius & Kölver 1983; Becker 2005). Probably the most interesting classifier system in East and mainland Southeast Asia is found in Weining Ahmao, a Hmong-Mien (Miao-Yao) language spoken in Western Guizhou Province in China. In this language, the classifier is inflected within a twelve-slot paradigm that covers the following functions (Wang 1957, 1972; Gerner & Bisang 2008):

- Number (singular vs. plural)
- Definiteness (definite vs. indefinite)
- Augmentative (= male speaker), medial size (= female speaker), diminutive (child speaker)

Table 6 shows the different forms of the classifier *lu*⁵⁵ in the twelve-slot paradigm for classifier inflection (on the regularities with which these forms are produced, cf. Gerner & Bisang 2008).⁴ Historically, the system developed from two different pathways that took place in the following temporal sequence: (i) the size/gender distinctions are due to the reanalysis of augmentative and diminutive prefixes on the noun as elements of the classifier; (ii) the definiteness/indefiniteness distinction is the result of a fusion of the numeral *i*⁵⁵ “one” with the classifier. This created two forms, the unmarked form for definiteness and the fused form for indefiniteness (Gerner & Bisang 2009).

Table 6. The inflectional paradigm of the classifier *lu*⁵⁵ in Weining Ahmao

Gender/age register	Size	Singular		Plural	
		Definite	Indefinite	Definite	Indefinite
Male	Augmentative	<i>lu</i> ⁵⁵	<i>lu</i> ³³	<i>ti</i> ⁵⁵ <i>a</i> ¹¹ <i>lu</i> ⁵⁵	<i>di</i> ³¹ <i>a</i> ¹¹ <i>lu</i> ⁵⁵
Female	Medial	<i>lai</i> ⁵⁵	<i>lai</i> ²¹³	<i>tiai</i> ⁵⁵ <i>a</i> ¹¹ <i>lu</i> ⁵⁵	<i>diai</i> ²¹³ <i>a</i> ¹¹ <i>lu</i> ⁵⁵
Children	Diminutive	<i>la</i> ⁵³	<i>la</i> ³⁵	<i>tia</i> ⁵⁵ <i>a</i> ¹¹ <i>lu</i> ⁵⁵	<i>dia</i> ⁵⁵ <i>a</i> ¹¹ <i>lu</i> ⁵⁵

In addition to number and definiteness, the classifier system of Weining Ahmao also distinguishes three levels of size, i.e., augmentative, medial size and diminutive. This function, which is taken literally in certain contexts, has further developed into a special type of politeness in which the speaker situates herself/himself as a man, a woman or a child in the speech situation in direct conversation. If the speaker is male, he uses the augmentative form of the classifier by default; if the speaker is female, the default classifier has the form of medium size; and if the speaker is a child, the strongly preferred classifier takes the diminutive form. In narrative texts, it is not necessary to situate the speaker’s position in the speech situation. Thus, the function of the classifier is limited to size (examples on politeness in Ahmao are discussed in Gerner & Bisang 2008: 726–729; 2010: 591–597).

4. The upper-case numbers mark tonality: ⁵ represents the highest pitch, and ¹ stands for the lowest pitch. Level tones are represented by repeating the same figure, while contour tones are expressed by different figures. Thus, ⁵⁵ stands for a high level tone, ⁵³ for a tone falling from high to mid level.

3.1.3 *The functional range of noun classifiers: Mayan languages of Meso-America*

The function of individuation is of central importance in languages of the class-noun type, too. Since class nouns “are crucially found independently of the operation of quantification” (Grinevald 2000: 64) and since languages with class nouns often have a separate system of numeral classifiers, their individuating function is not related to the context of counting but is manifested in the interpretation of nouns occurring with class nouns. Thus, “all nouns occurring in expressions with noun classifiers are interpreted as third-person referential and individuated nouns” (Zavala 2000: 138). This statement concerns Akatek, a language that belongs to the Mayan subgroup of Q’anjob’alan, and it also applies to Jacaltec, which belongs to the same subgroup. The individuating function of classifiers is even observable in a language like Hup, whose classifier system is still at an early stage of development. As Epps (2007: 109–110) points out, the classifier can be omitted in constructions in which nouns are not individuated. This is the case in constructions of plurality, generic use of nouns and negative existence. In other contexts, the classifier, which is morphologically bound to the noun, is strictly required.

In addition to their individuating function, class nouns are used to mark definiteness and anaphoricity and are involved in encoding relative clauses (Aikhenvald 2000: 88). In the incipient classifier system of Hup as described by Epps (2007), class nouns are used for anaphoric reference and they “arguably serve a marginal agreement-marking function by virtue of appearing, optionally, on multiple constituents of the clause” (Epps 2007: 121). Epps (2007) does not report any referential function of noun classifiers in terms of definiteness or specificity. The anaphoric function of noun classifiers is illustrated by (26), while (27) is one of the very few instances of agreement:

(26) *Hup* (Epps 2007: 121):

núp bóda=tat-ǝž, [*núp d’ɔh-yát-æp*] = *tat*
 this ball=CL:FRUIT-PF this rot-lie.ON.ground-DEP=CL:FRUIT
 “This was a ball, this rotting round thing lying there.”

(27) *Hup* (Epps 2007: 121):

núp=(tat) bóda=tat tíh=pǝh=(tat)
 that-CL:FRUIT ball=CL:FRUIT 3SG=big=CL:FRUIT
 “this big ball”

The functional range of Jacaltec class nouns (Craig 1986a) covers all the functions discussed above. In Example (28a), the two noun classifiers *naj* and *noʔ* mark definiteness. In (28b), the same class nouns are used anaphorically:

- (28) *Jacaltec*: (Craig 1986a: 264; Grinevald 2000: 65)
- a. *xil naj xuwan no7 lab'a*.
 saw CL John CL snake
 “John saw the snake.”
- b. *xil naj no7*.
 saw CL CL
 “He saw it.”

The pronominal use of noun classifiers also extends to possession and relative-clause formation. Since pronominally used class nouns cannot be coreferential with nouns to their left, there are different binding conditions between possession constructions with (29a) and without a class noun (29b). While the former excludes coreference of the possessor with the subject of the clause, the latter triggers that interpretation:

- (29) *Jacaltec*: (Craig 1986b: 276)
- a. *xil [naj pel [s-mam naj]]*.
 saw CL Peter POSS:3SG-father CL:MALE
 “Peter_i saw his_j father.”
- b. *xil [naj pel [s-mam ø]]*.
 saw CL Peter POSS:3SG-father
 “Peter_i saw his_i father.”

Finally, class nouns are involved in relative-clause formation. They must be deleted in the relative clause if they are coreferent with the head noun.

3.1.4 *The noun-class markers of Miraña*

As in the case of class nouns, the presence of the class marker is associated with individuation, or ‘unitization’ in terms of Seifart (2005: 223–244). In Miraña, this means that a noun can only take markers of dual and plural if it is combined with a class marker. Since a noun is interpreted as singular if it is marked by the class marker alone, the class marker has singulative function. Examples (30a),(b) illustrate that a non-classified noun like *ko*: “wood/logs” cannot take dual and plural markers. This is only possible if such a noun takes a classifier as in (30d)–(e). If *ko*: is marked only by the class marker, it refers to a single item of wood, which is a log if it is classified by the three-dimensional class marker *-ɪba* (30c).

- (30) *Miraña*: (Seifart 2005: 228–229)
- a. **ko*:-*ku*
 wood-DU
 Intended meaning: “two [pieces of] wood”

- b. *ko:-:nɛ
wood-PL
Intended meaning: “[pieces of] wood”
- c. ko-ʔba
wood-CL:3DIM
“log” (* “logs”)
- d. kó-ʔba:-kuu
wood-CL:3DIM-DU
“two logs” (* “one log”, * “more than two logs”)
- e. ko-ʔbá:-nɛ
wood-CL:3DIM-PL
“logs” (* “one log”, * “two logs”)

Non-classified nouns and optionally classified nouns need a classifier for number marking. In the case of obligatorily classified nouns, number marking is no problem, since the class marker is always present. Bare repeater nouns are countable and have singular reference. As is shown in the following example, they have to be marked for plural if plural is meant. Thus, they have the same status as an optionally classified noun with a class marker.

- (31) *Miraña*: (Seifart 2005: 118)
- a. bahkuu
“bone” (*“bones”)
- b. báhkuu:-nɛ
bone-PL
“bones” (*“bone”)

From the perspective of their individuating function, *Miraña* class markers behave like numeral classifiers and class nouns. At the same time, they are also characterized by their highly developed agreement properties, which are briefly discussed to show what grammatical functions beyond individuation are associated with class markers.

Noun-class markers are suffixed to various pronominal expressions (among them third person pronouns, demonstratives, interrogative pronouns and possessive pronouns), to numerals, to verbs in main clauses and to verbs that function as predicates of relative clauses (Seifart 2005: 123–141). As is typical of many noun-class systems, noun-class markers occur on different constituents. In Example (32) of a more complex noun phrase, the marker for two-dimensional round things is suffixed to the head noun as well as to the demonstrative and the relative clause formed by the intransitive verb *muíhuu* “be big”.

- (32) *Miraña*: (Seifart 2005: 169)
 ε:-*hi* *muíhu-hí* *kuú:muu-hí*
 DEM-CL.2DIM.round be.big.SUB-CL.2DIM.round turtle-CL.2DIM.round
 “that big turtle”

3.2 The functional range of compound-related verbal classification

The categories for which states of affairs are classified in Nyulnyul, Gooniyandi and Mosestén (cf. Section §2.2) cover a broad range of grammatical domains. This is shown by the following list:

- Telicity
 Classification is determined by the classical distinction between telic states of affairs with a terminal boundary and atelic states of affairs with no such boundary. McGregor (1990, 2002) distinguishes between ‘extendible’ (roughly atelic) and ‘accomplishment’ (roughly telic).
- Direction of movement
 The classification of an event depends on its direction with regard to a certain center of orientation (e.g. towards the center or away from it, above it or below it, into it or out of it). McGregor (1990, 2002) uses the term ‘vectorial configuration’.
- Degree of activity
 The classification of an event is determined by the question of whether it is dynamic or stative: [± dynamic].
- Valency
 The classification of an event depends on properties such as monovalent, bivalent and avalent on the one hand and transitive, intransitive, reflexive, reciprocal and middle on the other.
- Control by the subject or the actor
 The classification of an event depends on the extent to which it is caused by the subject or the actor.

The classification system of Nyulnyul is mainly based on telicity. Eight of its ten primary inflecting verbs can be distinguished by the feature of [± telic] with further distinctions such as [± dynamic], movement in space and actor/subject control operating within [+ telic] and [–telic]. Only two primary inflecting verbs are not related to telicity. They mark reflexive/reciprocal (-*BARNJ*) and dynamic activity (-*J* “say”). See McGregor (this volume, Table 7).

Table 7. Verbal classifiers in Gooniyandi (McGregor 2002: 58)

Valency	Extendible	Accomplishment	Clause types
monovalent	-ARNI ₂ “extend-self” -I “be, go”	-MARNI “effect-self”	reflexive/ reciprocal
		-BINDI “get, become” -ANI “fall”	intransitive
avalent		-MI “effect”	transitive, intransitive
		-DI “catch”	middle & transitive
		-BINI “hit”	transitive, impersonal &
		-ARNI ₁ “emerge”	intransitive (rare)
		-BIRLI “consume”	
bivalent	-A “extend”		transitive, middle & intransitive
		-ARRI “throw, put”	transitive

In Gooniyandi, the distinction of [\pm telic] is pervasive, i.e., each of the twelve verbal classifiers either marks [+ telic] or [-telic]. At the same level, valency is the second primary distinction by which verbs are classified in Gooniyandi. The two distinctions form a matrix by which the basic classification system of verbal classifiers can be defined in Gooniyandi. Table 7 from McGregor (2002) reflects this system (as mentioned above, ‘extendible’ roughly stands for [-telic], while ‘accomplishment’ stands for [+ telic]):

The other three criteria – i.e., direction of movement, degree of activity and subject/actor control – also operate in the classification system of Gooniyandi. Even though it is certainly not possible to do justice to the delicate semantic distinctions at work in this system, I will briefly discuss the two criteria of direction of movement and subject/actor control. For the overall richness of the system, I refer the reader to McGregor (1990, 2002).

The two verbal classifiers -*BINDI* and -*ANI* are both [+ telic/accomplishment] and intransitive. From the temporal perspective of telicity, they can be described as follows:

-*ANI* focusses on the point of accomplishment as the initial point, before which there is no evidence of the process. However, after the accomplishment there may be a continuation of the induced state of affairs. -*BINDI* by contrast focusses on the point of accomplishment as the culmination of preceding events, of which there is evidence. After the accomplishment of an -*ANI* process, activity (and change) may still ensue; but the active or dynamic part of a -*BINDI* process is all before the point of accomplishment. (McGregor 1990: 561)

At the same time, the two verbal classifiers also express direction of movement, either horizontal (away from vs. towards a center of interest) or vertical (moving up or down). In the case of horizontal meaning, *-ANI* “refers to leaving, or setting off from a point”, while *-BINDI* “refers to reaching or achieving a point” (McGregor 1990: 562). It is probably needless to say that the spatial function of the two verbal classifiers basically reflects their temporal functions. In both cases, the action is either looked at from its initial point (*-ANI*) or from its terminal point (*-BINDI*). From such a perspective, one might even ask which of the two functions, the spatial function or the temporal function, is primary (cf. e.g. assumptions of grammaticalization from space to time by Heine et al. 1991). The same difference in perspective also operates in contexts of vertical movement. *-ANI* “indicates the direction of motion as downwards from a point” (McGregor 1990: 561). In this function, it occurs with verbs like *gard-* “fall”, *thood-* “descend”, *yood-* “sit down”, *yilij-* “rain to fall”, etc. In each case, the action “may be regarded as accomplished once the point of origin is left” (p. 561). In contrast, actions classified by *-BINDI* are “directed upwards to a point” (p. 562). *-BINDI* occurs with verbs like *doorloog-* “arise”, *bajgi-* “get up and go”, *bar-* “climb upwards”, etc. With each verb, the action is “accomplished once the endpoint has been reached” (p. 562). Finally, the same distinction shows up again metaphorically in cases of change of state. In the case of *-BINDI*, the actions it classifies “typically show signs of occurring before they are actually accomplished” (p. 562), while in actions classified by *-ANI* “the new state is seen as a sudden change of condition of the entity, which cannot differ in degree from any previous condition” (p. 562).

A number of verbs occur either with *-MI* or with *-ANI* or *-BINDI*. As seen in Table 7, *-MI* primarily differs from *-ANI* or *-BINDI* with regard to valency and the clause types it can occur with. While *-ANI* or *-BINDI* seem to be exclusively intransitive, *-MI* also occurs with transitives and middles. But there is an additional contrast in terms of subject/actor control:

The difference invariably lies in the degree of activity of the individual: where the classifier *-MI* is chosen, the state or condition is more strongly under his/her active control, or s/he is more actively involved in it. For example, *thiddi-* “anger”, *ngiddinyjila-* “hungry” (among others) occur with both *-BINDI* and *-MI*, in the latter case typically with the progressive aspect, referring to processes of increased agitation. With *-BINDI* no more than that the state/condition was achieved is indicated. (McGregor 1990: 569)

The above description reveals that in most cases the semantic properties of the classification criteria combine two or more classification categories. Some of these categories are more important than others for the classification system as a whole (telicity in Nyulnyul, telicity and valency in Gooniyandi). Moreover, a look at the

verbal classifiers *-ANI* and *-BINDI* of Gooniyandi reveals that it is sometimes difficult to see clearly which of the functions is primary. Is it the function of telicity or the function of direction of movement that is more basic for *-ANI* and *-BINDI*? In the case of the two Gooniyandi verbal classifiers, it might be impossible to decide. In fact, it might be a property of verbal classifiers in general that they combine different functions in a specific way.

In Mosestén, verbal classification is also characterized by the combination of various properties. In contrast to the Australian languages described above, the role of telicity seems to be of minor importance. The central classificational properties of the verbness markers are activity, valency and control by the subject or the actor, while the property of direction of movement seems to be of low relevance. Table 8 lists the basic functions of the eight verbness markers discussed in §2.2:

Table 8. The basic functions of eight verbness markers in Mosestén (data from Sakel 2007: 318)

Verbness marker	Classificational property
<i>-i-</i> “BE/HAVE”	mostly intransitive, low subject control
<i>-yi-</i> “DO/BE”	general meaning: intransitive and transitive, stative and active
<i>-tyi-</i> “PUT”	transitive, dynamic, high subject control
<i>-jo-</i> “BECOME”	intransitive, mostly stative involuntary, no subject control
<i>-ki-</i> “BE”	intransitive, stative and dynamic
<i>-ti-</i> “DO”	intransitive, dynamic, some subject control
<i>-wi-</i> “SEE/HEAR”	intransitive and transitive, experience or creation of visual or audio stimulus, unproductive
<i>-cchi-</i> “MOVE/HAVE”	intransitive and transitive, unproductive

The verbness marker *-yi-* is the most general verbal classifier. As Sakel (2007: 323) puts it, “[I]t is the most productive VM [= verbness marker; W.B.] and can be added to any non-finite element to turn it into a complex predicate.” The last two markers are unproductive and have special meanings which will not be discussed here. To provide an idea of how the classification system works, I briefly discuss how the selection of the verbness markers *-yi-*, *-tye-*, *-ki-* and *-jo-* classifies actions in terms of subject/actor control:

(33) *Mosestén*: (Sakel 2007: 319)

- a. *Yäe wai'-ye-' öjñi'.*
 1SG.M boil-DO/BE-[1M.SG] > 3F[SG] water[F]
 “I boil the water.”
- b. *Mö' öjñi' wai'-tye-'.*
 3SG.F water[F] boil-PUT-[1M.SG] > 3F[SG]
 “I put the water to boil.”

- c. *Mō' öjñi' wai'-ki-'*
 3SG.F water[F] boil-BE-[3]F[SG]
 “The water boils.”
- d. *Mō' wej wai'-jo-'* *arosh-khan.*
 3SG.F seed[F] boil-BECOME-[3]F[SG] rice.LOAN[F]-IN
 “The seed was (accidentally) boiled in the rice.”

In Example (33), the functions of the verbness markers can easily be compared because the main verb is always the same, i.e., *wai'* “boil”. In (33b), the verbness marker *-tyi-*, realized as *-tye-* because of vowel assimilation, explicitly stands for high subject control, while the marker *-yi-* in (33a), realized as *-ye-* for the same reason, only denotes an action in general and is neutral with regard to subject control. The use of *-ki-* in (33c) focuses on the state property of the verb, while *-jo-* in (33d) points out the involuntary character of the action – the seed fell into the hot water and was boiled accidentally with no subject or actor being around to exert control.

3.3 Light verbs and serial verb constructions in the light of verbal classification

The phenomenon of verbal classification is sometimes compared to serial verb constructions or complex predication. In this context, Butt (2003) treats light verbs of South Asian languages, certain serial verb constructions and compound verb constructions (cf. §2.2 and 3.2) as basically the same phenomenon. As outlined in this section, there are considerable differences between compound verb constructions and the other constructions that express concepts whose functions are similar to the ones expressed by the compound verb constructions. It is argued below that these differences are related to a certain extent to the fact that only the former are used for classification.

Light verbs are an areal phenomenon of South Asia (Masica 1976) which is described in various studies, most of them on Indo-Aryan (Hook 1974, 1991, 1993; Butt 1995, 2003). They take the last position in V-V constructions and are fully inflected finite verbs, while the non-final verbs take non-finite forms which will not be discussed here. Apart from their function as light verbs, they can also occur independently as main predicates. In their light verb function, they “serve to modulate the main predication in a subtle manner” (Butt 2003: 3). Thus, they are involved in the specification of the overall V-V construction for semantic notions such as temporal structure (Aktionsart, telicity in particular), causation, volition, forcefulness, permissiveness, benefaction or suddenness of an event. In Example (34) from Urdu, the light verb *mar-* “hit” marks forcefulness. Analogously, the verbs

di- “give” and *li-* “take” express benefaction (35) and telicity/completion in (36), respectively. In Example (37), the two functions of permissiveness and telicity (completion) are combined:

- (34) Urdu (Butt 2003: 11):
nadya=ne xat=ko lik^h mar-a.
 Nadya.F.SG=ERG letter.M.SG=ACC write hit-PFV.M.SG
 “Nadya dashed off the letter forcefully.”
- (35) Urdu (Butt 2003: 11):
nadya=ne xat lik^h di-ya
 Nadya.F.SG=ERG letter.M.SG.NOM write give-PFV.M.SG
 “Nadya wrote the letter (for somebody else).”
- (36) Urdu (Butt: 2003: 2, 12, 21):
nadya=ne xat lik^h li-ya.
 Nadya.F.SG=ERG letter.M.NOM write take-PFV.M.SG
 “Nadya wrote a letter (completely).”
- (37) Urdu (Butt 2003: 21):
nadya=ne saddaf=ko xat lik^h
 Nadya.F.SG=ERG Saddaf.F.SG=DAT letter.M.NOM write
le-ne di-ya.
 take-INF.OBL give-PFV.M.SG
 “Nadya let Saddaf write a letter (completely).”

Two serial verb constructions look like compound verb constructions, the resultative and the directional construction. Both are generally attested in East and mainland Southeast Asian languages (Bisang 1996, 2010), described here with examples from Chinese. Resultative constructions have been analyzed in many different theoretical approaches (Li Y. 1990; Sybesma 1999; Wu 2004; Her 2007; Li C. 2008; among others). Generally, they consist of a sequence of two adjacent verbs, V_1 and V_2 , in which the second verb “signals some RESULT of the action or process conveyed by the first element” (Li & Thompson 1981: 54). With this function, resultative constructions generally express telicity. In the following example, the second verb *gānjìng* “be clean” denotes a result of the first verb *xǐ* “wash”:

- (38) Chinese:
Tā xǐ-gānjìng yīfu le.
 s/he wash-clean clothes PRF
 “S/he washed the clothes clean.”

Chinese directional serial verb constructions consist of a first verb (V_1), which is usually a verb of movement, plus one or two subsequent verbs (V_2, V_3) which specify

the direction of the main verb. In the following example, the direction of V_1 *diào* “fall” is specified by the directional verbs (vD) *xià* “move down” and *qù* “go [away from center of interest]”. The sequence of the two directional verbs is fixed:

- (39) *Chinese:*
Tā diào xià qù.
 s/he fall vD: move.down vD:go
 “S/he fell down.”

Even though the resultative construction and the directional construction are clearly related historically (Xu 2006; Bisang 2010), they form two different patterns in modern Chinese with different syntactic properties. This does not seem to be the case with light verbs. Whether they express telicity or direction of movement, their V-V structure is basically the same. Thus, light verb constructions in South Asia and the resultative and the directional construction in Chinese and other East and mainland Southeast Asian languages are not quite the same. This is even more apparent if one compares the above light verbs and serial verb constructions with the compound verb constructions described in §2.2 and 3.2. There are the following main differences between compound verb constructions and the light verb/serial verb constructions discussed in this section (and only these – I do not address serial verb constructions in general, for that cf. Bisang 2009):

- i. Serial verbs and light verbs are not obligatory, but inflecting verbs, verbal classifiers and verbness markers in compound verb constructions are.
- ii. Light verbs can be combined (cf. (37)), but compound verbs tend not to be combined.⁵ Directional verbs can be combined within the framework of a fixed pattern in Chinese (cf. (39)) and other East and mainland Southeast Asian languages (Bisang 1996).

Light verbs and serial verbs are not obligatory in the sense that they do not belong to a certain grammatical category consisting of a set of values from which the speaker is forced to select one in a given syntactic context. Thus, a speaker of a language with a binary aspectual system has to select either the value of perfective or imperfective in a finite declarative clause. Similarly, a speaker of Nyulnyul has to use a verbal classifier with all verbs except the inflecting verbs in a finite

5. In Nyulnyul, Gooniyandi and Mosestén, the occurrence of more than one verbal classifier in a single predication is impossible. Since Tsafiki (Dickinson 2002) can combine two layers of verbal classifiers in a limited number of constructions, this condition is not formulated as a rigid condition. As pointed out below, there are functional reasons for why the use of more than one classifier is not very likely.

declarative clause. Light verbs and the verbs used in resultative or directional serial verb constructions are not subject to such an obligatory selection. The speaker may use them for adding information to the meaning of the main verb if s/he decides to provide further specifications. Once a category has become obligatory, it no longer provides additional information – it provides required information. This shift has consequences for the properties of the information in various ways, which certainly need more research. From the present perspective, it implies that the information conveyed by an obligatory category does not necessarily have a modulating function and can also be used for classification. The modulating function can then be seen as a corollary of classification in the sense that certain aspects of a verb (or a noun) which do not belong to its lexical semantics can be added to its meaning through the process of subsuming it under a certain class which has that property (cf. Example (3) on the contribution of the classifier to the semantics of the noun *đồng nghiệp* “colleague”). Thus, classification can be used in a modulating function for the purpose of minimizing the hearer’s search domain for a certain event (or object) in a given context.

Classification may also be one of the reasons why inflecting verbs, verbal classifiers and verbness markers cannot be combined. If an event (or a noun) needs to be classified, there is rarely a discourse context in which classification for more than one particular category is necessary. If a classification system is more grammaticalized, a given event (or a noun) does not even allow the selection of one classifier out of a set of potential classifiers – there is only a single classifier available (cf. Hmong in §2.1.2). In the case of Australian compound verb constructions, there are many verbs which only occur with one particular inflecting verb/verbal classifier.

If, as Butt (2003: 23) says, “light verbs contribute to the structure of the event within the domain of primary predication,” this is a convincing analysis for the grammatical phenomenon she deals with. The phenomenon of compound verb constructions discussed in §3.2 serves a different purpose. It is used to classify states of affairs. If this is the primary function of inflecting verbs, verbal classifiers and verbness markers, the modulating or specifying function is only a corollary of that primary function.

4. Why is classification recurrent in the nominal domain and relatively rare in the verbal domain?

4.1 General remarks

The above discussion shows that compounding is a common source for the development of classification systems in the nominal and in the verbal domain. Thus, classification shares more or less the same initial conditions in both domains. In spite of this, the relevance of classification differs at later stages of grammaticalization – it is preserved with grammatical categories of the noun and tends to lose its importance with verbal categories. As pointed out in §1, this is due to two conditions, repeated here:

- i. The individual elements used for classification should not interfere with other processes of grammaticalization that are also based on the same semantics or the same source of grammaticalization.
- ii. The use of a classification marker in later processes of grammaticalization should not impair the initial classification system.

Both conditions are necessary for the stability of the semantics of a classifier system if its markers develop into various grammatical categories. If individual markers representing a system of classification are co-opted for further processes of grammaticalization, the classification system may disappear if too many of them are employed for expressing other functions. If this happens, condition (i) is violated. Condition (ii) is violated if the semantic distinctions of the classification system cannot be maintained in processes of further grammaticalization. This is the case if the classification system as a whole is not compatible with the new category or if the category into which classifiers develop is not broad enough to include the distinctions made in the initial classification system. Needless to say, the distinctions of a classification system can get lost if their semantics is irrelevant for the new category. As shown below, the grammatical categories of the nominal domain (cf. §4.2) are compatible with classification – classification even enhances their grammaticalization. In the case of the verbal domain, the criteria of classification (cf. §4.3) are not homogeneous enough to maintain the system as a whole, and the different categories into which the initial classifiers develop cannot be motivated by classification. In addition, the new categories consist of a rather small number of oppositions. If a verbal classification system becomes part of an aspectual system that may ultimately move into the basic opposition of perfective vs. imperfective, the number of distinctions for integrating all the members of the classification system is far too large, i.e., there are simply too many classification markers to be

integrated into the system. As a consequence, condition (ii) will sooner or later be violated, and its semantic oppositions eliminated, since the classification system with its comparatively large number of distinctions cannot be squeezed into a system with only a few distinctions or possibly even just a binary distinction. In general, classification can only remain stable if its members act as a single whole without too many ‘stray members’ that opt out of the system and follow other pathways of grammaticalization (condition (i)) and if the semantic properties of that system are compatible with the new grammatical category in terms of semantic motivation and number of oppositions (condition (ii)).

The effects of conditions (i) and (ii) are described here. Section 4.2 deals with the noun, §4.3 with the verb. As will be shown, both conditions are maintained more easily in grammaticalization processes that take place in the nominal domain.

4.2 The role of classification in grammatical categories of the noun

Denny (1976) described the criteria for determining class membership of nouns in terms of three basic kinds of human interactions, physical, functional and social (cf. §2.1). The criteria of physical interaction are roughly identical to the criteria discussed by Allan (1977), i.e., material (animacy, abstracts, inanimates), shape (one-dimensional, two-dimensional, three-dimensional), consistency (flexible, hard or rigid, non-discrete), size (big, small), location, etc. The criteria that matter in functional interaction are man-made object, food, transportation, religious observance, etc. Finally, the criteria concerning social interaction are determined by the social organization of a culture and reflect properties such as honored or respected persons, kin vs. non-kin, as well as deities.

Each of these criteria is based on concrete markers with their source concepts. For the basic numeral classifiers of Vietnamese listed in (2), there are markers referring to concepts of animacy (*người* ‘man [for humans]’, *con* ‘living thing that is not human’, *cây* ‘tree or plant’, *cái* ‘non-living thing’), shape (*sợi* ‘hair, thread, cord [one-dimensional]’, *tấm* ‘flat piece of material [two-dimensional]’, *quả* ‘fruit or round, globular object [three-dimensional]’) and various concrete objects such as *hòn* ‘stone or stonelike thing’ or *chiếc* ‘individual item of floating object, vehicle’, etc. Similarly, the classifier system of Jacaltec summarized in Table 2 consists of a broad variety of concepts which are sometimes highly specific (e.g. *metx’* ‘dog’, *te7* ‘plant’, *tx’ang* ‘twine’, *atz’am* ‘water’, etc.). The vast majority of the concepts represented by these markers are not used for other pathways of grammaticalization. Animacy might appear to be an exception at first glance, but animacy rather stands for different values within the animacy hierarchy as a linguistic universal (Silverstein 1976; Dixon 1979) than for a specific product of grammaticalization. In this function, it motivates various grammatical patterns represented in number

marking, splits in alignment systems, direct/inverse marking and differential object marking (Croft 2003).

A look at Heine & Kuteva (2002: 317–326) reveals that only the source concepts of ‘head’ (2002: 321; cf. the Chinese classifier *tóu*), ‘man’ (2002: 323; cf. Jacaltec *naj* in Table 2), ‘mouth’ (2002: 323, cf. Chinese *kǒu*), ‘thing’ (2002: 326, cf. Vietnamese *cái* in (2)) and ‘woman’ (2002: 326, cf. Jacaltec *ix* in Table 2) are involved in other processes of grammaticalization.⁶ Therefore, the concepts associated with classification generally comply with condition (i), i.e., they do not serve as the source concepts for other grammatical categories and thus do not interfere with other processes of grammaticalization.

The two main functions of classification in the nominal domain are individuation and identification (cf. §3.1.1). Individuation is related to the functions of singulative and numeral classification, while identification is associated with definiteness, contrastive focus, anaphoricity, possession and relative clause marking. In each of these functions, the classification system can not only be maintained in the course of grammaticalization processes, it can even support the use of classifiers for these new functions (on the correlation between classification and individuation/identification, cf. Bisang 2002; for a more formal account of the relation between individuation/atomization and (in)definiteness, cf. Li & Bisang 2012). In the case of individuation, the classification system refers to certain conceptual boundaries that can be used for referring to a lexical concept as a countable unit. Thus, it is not necessary to reduce the distinction of the classification system for expressing individuation. In fact, the system can be used as it is exactly for the purpose of individuating a concept. Similarly, referring to the conceptual properties expressed by the classification system can be used to restrict the search domain for finding a certain object in space. With the use of a classifier, the speaker guides the hearer’s attention towards the object s/he wants her/him to identify. This is a useful property for the expression of definiteness by profiling prominent cognitive properties as they are activated in the speech act situation. It can be used for reference tracking as it is needed in anaphoricity, it supports the creation of contrast between different objects as it is needed in contrastive focus and it helps identifying the object possessed in possessive constructions. Finally, classifiers are used in relative clauses to identify the head noun to make it accessible for further modification. Since there is a relationship of cognitive motivation between classification and individuation/identification, condition (ii) is not violated and, moreover, the semantic oppositions

6. These source concepts can be used for expressing the following functions (Heine & Kuteva 2002: 317–326): ‘head’ > (1) ‘front’, (2) ‘intensive-rel’, (3) ‘middle’, (4) ‘reflexive’, (5) ‘up’; ‘man’ > (1) ‘classifier’, (2) ‘exclamation’, (3) ‘indefinite pronoun’, (4) ‘male’, (5) ‘third-person pronoun’; ‘mouth’ > ‘front’; ‘thing’ > (1) ‘complementizer’, (2) ‘indefinite pronoun’, (3) ‘a-possessive’; ‘woman’ > (1) ‘classifier’, (2) female’.

operating in classification can be integrated into systems which by themselves contain a much smaller number of oppositions. Categories having to do with individuation and identification simply take over the semantic divisions of the classification system they are based on.

4.3 The role of classification in grammatical categories of the verb

The perception of an event covers more than just a single entity. Events consist of event structures with their verbs and their participants, and they can be looked at from different temporal and spatial perspectives. It is for that reason that Lewandowska-Tomaszczyk (2011: 39) states that events are constructed online:

Perception of what is going on in the outside world as events is a matter of the imposition of an event structure i.e. retaining their spatio-temporal characteristics and providing bounding in their construal. In other words, events are constructed online, in an interaction, and this depends on the participants of an action and its wider context.

If events are characterized by their spatio-temporal properties and if these properties are constructed online rather than perceived holistically, categories such as Aktionsart (telicity), direction of movement, degree of activity (stative vs. dynamic), valency and subject/actor control (cf. §3.2) do not have the same degree of stability and reliability for classifying events as the properties employed for the classification of nouns (cf. §3.1). As a consequence, verbal classification criteria have an intrinsic tendency to be used for specifying properties of events. With a predicate like 'walk', the speaker may want to specify that event by adding information concerning its direction and its temporal structure rather than to merely classify it for these categories.

The fact that it is harder to find stable generalizable properties of events that can be used for the classification of verbs may also provide a good explanation for why identification and classification are more strongly associated with reference and nouns than with predication and verbs. Thus, even in languages in which classification is used for counting events (cf. the verbal numeral classifier systems in many East and mainland Southeast Asian languages) verbal classifiers are not used for anaphoric reference to these events. In a context like (40), it would not be possible to use the classifier associated with *see* in the position of *that*. In fact, I know of no language with any system of verbal classification in which this is possible.

- (40) a. *I saw my friend and I was glad about that.*
 b. *I saw my friend and that was great.*

The tendency to co-opt the criteria of verbal classification for specifying properties of events has its impact on condition (i). Markers of Aktionsart (telicity) and direction of movement are also developed under conditions which are unrelated to classification. Light verbs and serial verb constructions are good examples of this (cf. (36) on Aktionsart and (39) on directional verbs). If this is possible in languages without verbal classification, this is also possible in languages with verbal classification. Classification markers associated with one of these functions can thus easily move away from classification into other pathways of grammaticalization that lead to directionality or tense-aspect. Needless to say, this is a violation of condition (i). If one looks at the individual source concepts of verbal classification in Nyulnyul, Gooniyandi and Mosestén, the number of verbs which have the potential to follow alternative pathways of grammaticalization is again considerable – a fact that invites additional violations of condition (i). Table 9 is based on Heine & Kuteva's (2002: 317–326) list of source concepts and presents the core classifiers of each of the three languages whose semantics are cross-linguistically attested to be involved in grammaticalization processes which have nothing to do with classification.⁷ Thus, the source concept of 'put' can also grammaticalize into COMPLETIVE (Heine & Kuteva 2002: 324) and the source concept of 'go' is part of no less than seven grammatical functions, i.e., ANDATIVE, CHANGE-OF-STATE, CONSECUTIVE, CONTINUOUS, DISTAL DEMONSTRATIVE, HABITUAL, HORTATIVE (Heine & Kuteva 2002: 321).

Table 9. List of core verb classifiers with the potential of being used in other pathways of grammaticalization

	Nyulnyul	Gooniyandi	Mosestén
copula ('be')	-N "be"	-I "be, go"	-i- "BE, HAVE" -yi- "DO, BE" -ki- "BE"
'put'	-M "put"	-ARRI "throw, put"	-tyi- "PUT"
'get'	-NY "get"	-BINDI "get, become"	–
'go'	-JID "go"	-I "be, go"	–
'do'	-J "say, do"	–	-yi- "DO, BE" -ti- "DO"
'say'	-J "say, do"	–	–
'give'	-W "give"	–	–
'fall'	–	-ANI "fall"	–
'have'	–	–	-i- "BE, HAVE"

7. The data are taken from Table 7 on Gooniyandi and from Table 8 on Mosestén. The data on Nyulnyul are from Table 7 in McGregor (this volume). The last two classifiers in Table 8 are not integrated into Table 9 because they are not productive.

The fact that the semantics of at least some classification criteria (Aktionsart, direction of movement) and a considerable number of individual verbal classifiers are widely used for various other processes of grammaticalization cross-linguistically weakens the integrity of the classification system as a whole. If too many markers opt out of the domain of classification and thus violate condition (i) classification as a whole loses its homogeneity and is in danger of extinction. This did not happen in Gooniyandi, Nyulnyul and Mosestén, but it might have happened in many other languages in which no trace of a former verbal classification survived.

The example of Gooniyandi in §3.2 has shown that one and the same verbal classifier often represents different functions and that it is sometimes hard to state which of these functions is primary. Thus, the semantic range of the classifiers *-ANI* and *-BINDI* covers the functions of telicity and direction. Such a situation seems to be quite common in systems of verbal classifiers. As soon as the same markers are grammaticalized into other functions, their functional range is streamlined into more consistent semantic properties. If a marker is grammaticalized into an aspect marker, it acquires the semantics of a specific aspectual function. A comparable development takes place with directional markers. Concomitant with this functional specification is the reduction of distinctions within individual grammatical categories. As was already pointed out above, aspectual systems may only consist of a binary distinction between perfective and imperfective. Directional systems usually consist of a few oppositions such as movement to a center of interest or away from it, movement up or down, movement into something or out of it, etc. (cf. (39)). The category of activity with its distinction between stative and dynamic is again of a binary nature, while valency allows for some more distinctions (monovalent, bivalent, trivalent, avalent, (in)transitive, etc.). Finally, the marking of subject/actor control is based on a limited number of positions between deliberate control and no control at all. The reduction of oppositions automatically impairs the number of distinctions made in the classification system and thus leads to the violation of condition (ii) and the substitution of the classifier system by another grammatical system.

The scenario of classifier loss described so far can only be stopped if classification is stabilized as a morphosyntactically autonomous grammatical category and if the alternative grammatical processes create an independent morphosyntactic environment for their own development. This process of bifurcation between classification on the one hand and other processes of grammaticalization on the other must have taken place in Mosestén. In this language, most of the verbness markers listed in Table 8 are subject to other processes of grammaticalization. Thus, the verbness marker *-tyi-* “PUT” also takes the function of an applicative marker. Analogously, the verbness markers *-ki-* “BE” and *-ti-* “DO” are used as antipassive markers (Sakel 2007: 325–329). In spite of these additional functions, the classifier system of Mosestén remains stable because the verbness markers occur in a different

References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith. 1977. "Classifiers". *Language* 53.283–311. doi:10.1353/lan.1977.0043
- Becker, Neele. 2005. *Numeralklassifikatoren im Thai*. Ph.D. dissertation, University of Mainz.
- Bertrais-Charrier, Fr. Yves. 1979 [1964]. *Dictionnaire Hmong (Meo Blanc) – Français*. Vientiane: Mission catholique.
- Bisang, Walter. 1993. "Classifiers, Quantifiers and Class Nouns in Hmong". *Studies in Language* 17:1.1–151. doi:10.1075/sl.17.1.02bis
- Bisang, Walter. 1996. "Areal Typology and Grammaticalization: Processes of grammaticalization based on nouns and verbs in East and Mainland South East Asian languages". *Studies in Language* 20:3.519–598. doi:10.1075/sl.20.3.03bis
- Bisang, Walter. 1999. "Classifiers in East and Southeast Asian Languages: Counting and beyond". *Numeral Types and Changes Worldwide* ed. by Jadranka Gvozdanovic, 113–185. Berlin: Mouton de Gruyter. doi:10.1515/9783110811193.113
- Bisang, Walter. 2002. "Classification and the Evolution of Grammatical Structures: A universal perspective". *Sprachtypologie und Universalienforschung* 55.289–308.
- Bisang, Walter. 2009. "Serial Verb Constructions". *Language and Linguistics Compass* 3:3.792–814. doi: 10.1111/j.1749-818X.2009.00128.x. <http://onlinelibrary.wiley.com/doi/10.1111/j.1749-818X.2009.00128.x/full>. (Accessed: 1 April 2014.)
- Bisang, Walter. 2010. "Grammaticalization in Chinese: A construction based account". *Gradience, Gradualness, and Grammaticalization* ed. by Elisabeth C. Traugott & Graeme Trousdale, 245–277. Amsterdam: John Benjamins. doi:10.1075/tsl.90.13bis
- Bisang, Walter. 2012. "Numeral Classifiers with Plural Marking: A challenge to Greenberg". *Plurality and Classifiers across Languages of China* ed. by Dan Xu, 27–49. Berlin: Mouton de Gruyter. doi:10.1515/9783110293982.23
- Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford, California: CSLI Publications.
- Butt, Miriam. 2003. "The Light Verb Jungle". *Harvard Working Papers in Linguistics* 9.1–49.
- Chierchia, Gennaro. 1998. "Reference to Kinds across Languages". *Natural Language Semantics* 6.339–405. doi:10.1023/A:1008324218506
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166119
- Craig, Colette G. 1986a. "Jaltec Noun Classifiers: A study in grammaticalization". *Lingua* 70.241–284. doi:10.1016/0024-3841(86)90046-X
- Craig, Colette G. 1986b. "Jaltec Noun Classifiers: A study in language and culture". *Noun Classes and Categorization* ed. by Colette G. Craig, 263–293. Amsterdam: John Benjamins. doi:10.1075/tsl.7.17cra
- Croft, William. 2003. *Typology and Universals*. Cambridge: Cambridge University Press.
- DeLancey, Scott. 1986. "Towards a History of Tai Classifier Systems". *Noun Classes and Categorization* ed. by Colette G. Craig, 437–452. Amsterdam: John Benjamins. doi:10.1075/tsl.7.26del
- Denny, J. Peter. 1976. "What Are Noun Classifiers Good For?" Papers from the 12th Regional Meeting, Chicago Linguistic Society, April 23–25, 1976 ed. by Salikoko S. Mufwene, Carol A. Walker & Sanford B. Steever, 122–132. Chicago: Chicago Linguistic Society.
- Dickinson, Connie. 2002. *Complex Predicates in Tsafiki*. Ph.D. dissertation, University of Oregon.

- Dixon, Robert M. W. 1979. "Ergativity". *Language* 55.59–138. doi:10.2307/412519
- Epps, Patience. 2007. "Birth of a Noun Classification System: The case of Hup". *Language Endangerment and Endangered Languages: Linguistic and anthropological studies with special emphasis on the languages and cultures of the Andean-Amazonian border area* ed. by Leo Wetzels, 107–108. Leiden: The Research School of Asian, African and Amerindian Studies (CNWS).
- Gerner, Matthias. 2009. "Instruments as Verb Classifiers in Kam (Dong)". *Linguistics* 47.697–742. doi:10.1515/LING.2009.023
- Gerner, Matthias & Walter Bisang. 2008. "Inflectional Speaker-Role Classifiers in Weining Ahmao". *Journal of Pragmatics* 40.719–731. doi:10.1016/j.pragma.2007.11.004
- Gerner, Matthias & Walter Bisang. 2009. "Inflectional Classifiers in Weining Ahmao: Mirror of the history of a people". *Folia Linguistica Historica* 30.183–218.
- Gerner, Matthias & Walter Bisang. 2010. "Classifier Declinations in an Isolating Language: On a rarity in Weining Ahmao". *Language and Linguistics* 11.579–623.
- Grinevald, Colette. 2000. "A Morphosyntactic Typology of Classifiers". *Systems of Nominal Classification* ed. by Gunter Senft, 50–92. Cambridge: Cambridge University Press.
- Haas, Mary R. 1942. "The Use of Numeral Classifiers in Thai". *Language* 18:3.201–205. doi:10.2307/409553
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511613463
- Her, One-Soon 2007. "Argument-Function Mismatches in Mandarin Resultatives: A lexical mapping account". *Lingua* 117.221–246. doi:10.1016/j.lingua.2006.01.002
- Hook, Peter Edwin. 1974. *The Compound Verb in Hindi*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- Hook, Peter Edwin. 1991. "The Emergence of Perfective Aspect in Indo-Aryan Languages". *Approaches to Grammaticalization* ed. by Elisabeth Traugott & Bernd Heine, 59–89. Amsterdam: John Benjamins. doi:10.1075/tsl.19.2.05hoo
- Hook, Peter Edwin. 1993. "Aspectogenesis and the Compound Verb in Indo-Aryan". *Complex Predicates in South Asian Languages* ed. by Manindra Verma, 97–113. New Delhi: Manohar.
- Hundius, Harald & Ulrike Kölver. 1983. "Syntax and Semantics of Numeral Classifiers in Thai". *Studies in Language* 7:2.165–214. doi:10.1075/sl.7.2.02hun
- Lewandowska-Tomaszczyk, Barbara. 2011. "Events as They Are". *Turning Points in the Philosophy of Language and Linguistics* (= *Łódź Studies in Language*, 21) ed. by Piotr Stalmaszczyk, 35–63. Frankfurt am Main: Peter Lang.
- Li, Charles N. 2008. *Mandarin Resultative Verb Compounds: Where syntax, semantics, and pragmatics meet*. München: Lincom.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Xuping & Walter Bisang. 2012. "Classifiers in Sinitic languages: From individuation to definiteness-marking". *Lingua* 122.335–355. doi:10.1016/j.lingua.2011.12.002
- Li, Yan-Hui Audrey. 1990. *Order and Constituency in Mandarin Chinese*. Dordrecht: Kluwer. doi:10.1007/978-94-009-1898-6
- Löbel, Elisabeth. 1996. *Klassifikatoren. Eine Fallstudie am Beispiel des Vietnamesischen*. Universität Köln: Habilitationsschrift.

- Löbel, Elisabeth. 2000. "Classifiers vs. Genders and Noun Classes: A case study of Vietnamese". *Gender in Grammar and Cognition* ed. by Barbara Unterbeck & Matti Rissanen, 259–319. Berlin: Mouton de Gruyter. doi:10.1515/9783110802603.259
- Masica, Colin. 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- McGregor, William B. 1990. *A Functional Grammar of Gooniyandi*. Amsterdam: John Benjamins. doi:10.1075/slcs.22
- McGregor, William B. 2002. *Verb Classification in Australian Languages*. Berlin: Mouton de Gruyter. doi:10.1515/9783110870879
- Mottin, Jean. 1980. *Contes et légendes Hmong Blanc*. Bangkok: Don Bosco Press.
- Sagna, Serge. 2008. *Formal and Semantic Properties of the Gújjolaay Eegimaa (a.k.a Banjal) Nominal Classification System*. Ph.D. dissertation, University of London.
- Sakel, Jeanette. 2004. *A Grammar of Mosestén*. Berlin: Mouton de Gruyter. doi:10.1515/9783110915280
- Sakel, Jeanette. 2007. "The Verbness Markers of Mosestén from a Typological Perspective". *New Challenges in Typology: Broadening the horizons and redefining the foundations* ed. by Matti Miestamo & Bernhard Wälchli, 315–335. Berlin: Mouton de Gruyter.
- Schultze-Berndt, Eva. 2000. *Simple and Complex Verbs in Jaminjung: A study of event categorisation in an Australian language* (= MPI Series in Psycholinguistics, 14). Wageningen: Ponsen & Looijen.
- Schultze-Berndt, Eva. 2003. "Preverbs as an Open Word Class in Northern Australian Languages: Synchronic and diachronic correlates". *Yearbook of Morphology* ed. by Geert Booij & Jaap van Marle, 145–177. Dordrecht: Kluwer.
- Seifart, Frank. 2005. *The Structure and Use of Shape-Based Noun Classes in Miraña (North West Amazon)* (= MPI Series in Psycholinguistics, 33). Nijmegen: Max Planck Institute for Psycholinguistics. webdoc.ubn.ru.nl/mono/s/seifart_f/struanuso.pdf. (Accessed: 1 March 2014.)
- Silverstein, Michael. 1976. "Hierarchy of Features and Ergativity". *Grammatical Categories in Australian Languages* ed. by Robert M. W. Dixon, 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Sybesma, Rint. 1999. *The Mandarin VP* (= *Studies in Natural Language and Linguistic Theory*, 44). Dordrecht: Kluwer. doi:10.1007/978-94-015-9163-8
- Truong, Van Chinh. 1970. *Structure de la langue vietnamienne*. Paris: Paul Geuthner.
- Wang, Fushi. 1957. "Guizhou Weining Miaoyu Liangci [The Classifier in the Weining Dialect of the Miao Language]". *Yuyan Yanjiu* 1957.75–121.
- Wang, Fushi. 1972 [1957]. "The Classifier in the Weining Dialect of the Miao Language in Kweichow." *Miao and Yao Linguistic Studies: Selected articles in Chinese* ed. by Herbert Purnell, 111–185. Ithaca, NY: Cornell University.
- Wilson, Stephen. 1999. *Coverbs and Complex Predicates in Wagiman*. Stanford, California: CSLI Publications.
- Wu, X.-Z. Zoe. 2004. *Grammaticalization and Language Change in Chinese: A formal view*. London: Curzon Routledge.
- Xu, Dan. 2006. *Typological Change in Chinese Syntax*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199297566.001.0001
- Zavala, Roberto. 2000. "Multiple Classifier Systems in Akatek (Mayan)". *Systems of Nominal Classification* ed. by Gunter Senft, 114–146. Cambridge: Cambridge University Press.

The diachrony of inflectional classes in four Germanic languages

What happens after transparency is lost?

Antje Dammel and Sebastian Kürschner

University of Münster / Catholic University of Eichstätt-Ingolstadt

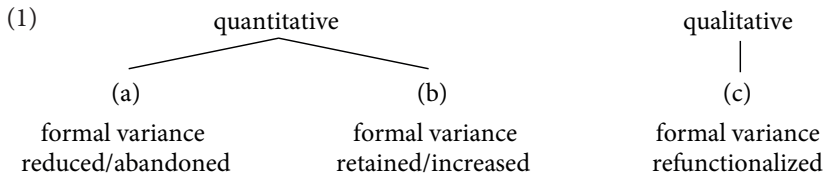
Inflectional classes are classification systems lacking a discernible synchronic function. We ask how former functions can be lost and, more importantly, what happens after this loss. In a comparative analysis contrasting four Germanic languages (Swedish, Danish, Dutch and German) and two lexical categories (noun declensions versus verb conjugations), we describe changes in the number of allomorphs, their relation to grammatical category exponence and the conditioning of class membership. Inflectional classes are retained in all four languages but reorganized in several respects. The comparison reveals common principles (allomorphy is retained if tied to relevant categories) and divergences (different degrees of simplification in the conditioning and number of allomorphs). A general discussion relates the findings to grammaticalization and to fundamental differences of the lexical categories concerned, and considers the question of whether inflectional classes can be termed functional or not.

1. Introduction

Inflectional classes can be canonically¹ defined as inflectional differentiations without a functional equivalent (Corbett 2009). As such, they violate the semiotic principle of ‘one function : one form’ and have no syntactic or referential scope beyond the classified item itself.

1. Corbett’s (2009) definition of inflectional class is a canonical one, based not on language data but on a theoretical principle, the maximal violation against one function : one form. Actual inflectional classes can conform more or less to this canonical ideal. As we will see in the course of this article, inflectional classes of Proto-Germanic had some degree of semantic/derivational motivation and were therefore less canonical in Corbett’s terms than those of the modern Germanic languages.

Diachronically, the inflectional classes of the earliest Germanic languages attested represent an advanced stage of formal and semantic erosion, retaining only formal residues of former (at least to some extent) semantically transparent classification systems. Thus, from a synchronic point of view, inflectional classes seem to be useless classification systems – an observation raising the question of what happens to them in language change towards modern Germanic languages. Theoretically, we can distinguish three possible scenarios:²



Predictions of morphological theorists clearly favor the leftmost option: see Mańczak (1980: 284) in one of his laws of analogy: “The number of morphemes having the same meaning more often diminishes than increases.”³

Nevertheless, inflectional classes are widespread cross-linguistically and often more persistent in language change than functional grammatical categories such as person or case. In the following, we not only focus on the fact THAT classes are retained but also on the question of HOW they are retained and – more importantly – reorganized.

A cursory look at the Germanic languages already reveals that all three scenarios in (1)(a)–(c) are attested. We find a wide spectrum from the highly complex inflectional class systems of Icelandic and Faroese to the complete leveling of inflectional classes in English (declensions) and in Afrikaans (conjugations).

To investigate the scenarios in (1) more thoroughly, we outline the historical development of inflectional classes in four Germanic languages of synchronically medium complexity: Swedish and Danish from the North Germanic branch, and German and Dutch for the West Germanic languages. All retain inflectional classes, but show considerable simplification and reorganization compared to Proto-Germanic.

2. For a more fine-grained typology of inflectional class change see Dammel (2011: Ch. II.2). A reviewer suggests formal renewal of categories as a fourth theoretical scenario. Their example is the formal renewal of reduplicating verbs according to the model of ablauting strong verbs in Germanic. In our account, this would be a case of type (a), as one class is abandoned by shifting its members to another class already in existence. However, formal renewal can also lead to an increase in variance (type b), as with Swedish weak verbs of the *fly*-type developing a new allomorph (§3.1.2). In our view, formal renewal is a criterion cutting across the classification presented in (1).

3. Cf. also Mayerthaler’s Natural Morphology approach sketched in Dressler et al. (1987).

The chapter is structured as follows. Section 2 sketches the prehistoric sources of today's inflectional classes, describing and discussing the semantically transparent and formally overt ancestors of inflectional classes assumed for the reconstructed Proto-Indo-European language (PIE), as well as early reorganization processes in Proto-Germanic and the oldest stages of the four Germanic languages from a grammaticalization perspective. Section 3 presents parallels and divergences between the four languages, focusing on three levels of description:


1. the number of inflectional classes (§3.1);
2. the exponence of inflectional classes related to the exponence of grammatical categories, which we relate to Bybee's (1985) concept of relevance (§3.2);
3. the factors conditioning class membership (§3.3).

We also compare the behavior of the two major lexical categories, nouns and verbs, asking which parallels and differences exist between declension and conjugation class change. Based on these twofold contrastive analyses, we sum up and discuss findings in Section 4.

2. Prehistory of inflectional classes in Germanic languages

According to reconstructions of PIE, the Proto-Germanic inflectional classes can be traced back at least partly to more functional classification systems (e.g. Wurzel 1986). The typical structure of PIE verbs and nouns was tripartite: root + stem formative + case/number inflections, as in (2).

(2) *The tripartite structure of Indo-European nouns and verbs*

* <u>ǵ</u> iHr	-o	-s	“hero/man-NOM.SG”
*nos	-e <u>ǵ</u>	-ō	“rescue-1.SG.PRS.IND”
root	stem	inflectional	
	formative	suffix	
			
stem			

For example, in the paradigm of PIE *ǵiHr-o-s “hero, man”, the stem extension -o figures in the entire case/number paradigm (partly in alternation with its ablauting counterpart -e; see below), case/number expression being fused in the final inflectional suffix (here, -s “NOM/SG”).

PIE had two main classes of nouns and verbs, athematic and thematic. Athematic stem formation is assumed to be older, thematic stem formation to be a later innovation. Thematic nouns and verbs were characterized by the ablauting

stem formative *-e/-o* or stem formatives ending in this vowel. Athematic nouns and verbs, by contrast, lacked this stem extension. Still, they exhibited a number of different stem formatives (see below for examples); also, nouns and verbs lacking stem formatives (root stems) form part of the athematic group. Athematic stem formation is more important for Germanic nouns than for verbs, where traces survive only in small irregular groups (e.g., Old High German *gān* “go”, *stān*, “stand”, *sīn* “be”, *tuon* “do”).

In addition to the characteristics of the tripartite structure, PIE stem formation was also characterized by accent-ablaut patterns which interacted with suffixation/ infixation and as such the form of the stem formatives.⁴

The stem formatives formed a closed class and figured overtly in each paradigm cell. Corresponding to their formal transparency, they are generally assumed to have been semantically transparent/functional in Pre-PIE (cf. Fortson 2010: 119–134; Meier-Brügger 2002: 289–292). While the Proto-Germanic declensions show traces of this functional motivation, stem extensions occurring in the present paradigm of Proto-Germanic strong verbs had become functionally opaque already in PIE. The weak verbs – secondary formations from nouns, adjectives and strong verbs – are an innovation of Proto-Germanic. They also had stem formatives, but these were still unreduced and showed derivational functions in the earliest periods of the attested daughter languages.

Thus, we can identify three different layers of transparency loss regarding stem extensions in Proto-Germanic. Functional erosion is most advanced for the stem formatives of strong verbs, less for declensional stem formatives and least for those of weak verbs. The different diachronic layers of stem formatives can be placed on different points on a grammaticalization cline from derivational morphology to inflectional class markers (Dammel 2011: 6–8). According to this status and their further development in Germanic, we treat them as early class markers here.

From PIE to Proto-Germanic stem extensions are subject to erosion and fusion with grammatical endings. The former tripartite structure of PIE was reduced to a bipartite one, as for example in Proto-Germanic **wer-as* “man-NOM/SG”. Due to this, allomorphy increased considerably, while the semantic basis was blurred even more. Ramat (1981: 62) therefore diagnoses a “crisis” of the inflectional class systems in Proto-Germanic.

4. We cannot deal with the complexity of PIE stem formation in detail here. For our purposes, we focus on the aspect central to our question (degree of semantic transparency), rather than giving a complete picture of the PIE regularities of formation (described for instance in Bammesberger 1990; Fortson 2010; Beekes 2011; Rix 2001; Mailhammer 2007b).

In §2.1 and §2.2, we present the semantically (or functionally) based groups assumed for PIE and Proto-Germanic, respectively, and sketch how they are blurred from Proto-Germanic on. We focus on those morphological phenomena which are continued in Germanic languages and relevant for the further developments.

2.1 Prehistory of declensions

For the further developments in the Germanic languages, it is most important to regard the formation of stems by use of stem extensions. Beginning already in PIE, these stem extensions fused with the inflectional suffixes and formed the basis of the Germanic declension classes.

Table 1 provides an overview of the grammatical functions and semantic distinctions that the authors of two textbooks on PIE (cf. Fortson 2010: 119–134; Meier-Brügger 2002: 289–292) attribute to some of the stem extensions. The table follows the distinction into thematic nouns with stem formatives in *-e/-o* and athematic nouns with stem formatives usually ending in a consonant, *-i* or *-u*, as introduced above.

Table 1. Functions of stem extensions in PIE (selection)

Class	Example suffix	Derivational basis	Function	PIE example
Athematic nouns				
n-stems	*-m _ṅ	verbs	act or result of the action of the verb	*kan-m _ṅ “thing sung” “song”
r-stems	*-ter/*-tor	verbs	agent	*deh ₃ -ter / deh ₃ -tor “giver”
s-stems	*-os	verbs	abstract	*ġéh ₁ -os “birth”
t-stems	*-t	(thematic) nouns / adjectives	abstract	*neuó-teh ₂ -t- “newness”
i-stems	*-ti	verbs	abstract	*mén-ti-s “thought”
u-stems	*-tu	verbs	abstract	*ġéws-tu-s “taste”
Thematic nouns				
o-stems	*-ó	athematic nouns; accent-shifted thematic nouns	possessive	*ġ ^w ih ₂ -ó- “thing having a bowstring” < *ġ ^w ieh ₂ - “bowstring”
	*-o	verbs	abstract	*róudh-o- “moaning”
	*-lo	verbs	various kinds; diminutives	*sed-lo- “seat”
	*-mo	verbs	action/event	*ġ ^w her-mo- “heat”
	*-tlo, *-dhlo, *-tro, *-dhro	verbs	instrument to accomplish an action (tool)	* (s)neh ₁ -tlo- “thing to sew with” *kr(e)i-dhro- “sieve”
	ā-stems	*ā < *-eh ₂	nouns	feminizing

By means of the stem extensions, noun stems dividing the nominal lexicon into functional and semantic groups are formed. These groups are based on fundamental semantic distinctions such as ‘abstract vs. concrete’, ‘human vs. non-human’ (cf. ‘agent vs. abstract nouns’) and ‘masculine vs. feminine’. Also, groups of nouns denoting related objects (such as tools) can share a derivational suffix. Additionally, the suffixes sometimes serve the grammatical function of assigning a lexical category to the stem and are both semantically and functionally based.

In addition to class constitution by means of stem extensions, nouns belong to a gender class in PIE. The history of gender is connected to another basic semantic distinction of nouns, namely that of animates vs. inanimates. A two-gender system based on animacy differences probably existed in the early stages of PIE (cf. Meier-Brügger 2002: 190–192). At later stages, a third gender became part of the system, probably based both on concrete/abstract and on sex differences. In PIE, gender is marked in the behavior of associated words and also in the inflectional paradigms themselves, where neuters inflect differently from masculines and feminines in the nominative and accusative cases. The animate-inanimate distinction is also reflected in the stem extension groupings, because the stem extensions sometimes correlate with one or two genders.

When stem extensions merge with the inflectional markers in Proto-Germanic and form part of the new basis of declension classes, the functional and semantic bases of the stem extensions begin to be blurred. Bammesberger (1990) shows that the semantics of the PIE stem extensions are still mostly preserved in Proto-Germanic (see also Mailhammer 2008), but the picture that emerges when we observe the declension class systems of the oldest Germanic languages is different: a large number of declension classes is no longer motivated by semantic or functional distinctions but mostly appears to be connected to gender (which itself as well has become mostly arbitrarily assigned; cf. §3.3.1). To sum up, declension classes stem from transparent distinctions in the nominal lexicon marked by segmental exponents and develop into a distinction of nominal groups constituted solely by their inflectional behavior.

2.2 Prehistory of conjugations

Proto-Germanic conjugation classes form two main diachronic layers: the strong conjugation, inherited from PIE and reorganized in Proto-Germanic, and the weak conjugation, a Proto-Germanic innovation.⁵ Both conjugations have several subclasses. Proto-Germanic strong verbs go back to PIE thematic stems, and their

5. We do not discuss Proto-Germanic reduplicating verbs here, as these had moved to the other two macro-classes by the earliest stages of the four languages considered.

stem extensions can be traced to PIE primary present formatives. There were about 20 formatives of this kind in PIE (Rix 2001: 14–20), and they did not display any synchronic function; whether they were functional in Pre-PIE is controversial. Rix (2001: 11) assumes that they formed a paradigm of derivative functions expressing Aktionsarten and that this system was obliterated by the change towards the PIE aspect system. Szemerényi (1990: 326) pleads for a case of layering, assuming several diachronic strata of more or less desemantized suffixes (see Mailhammer 2007b: 24–26). Being opaque already in PIE, these formatives are irrelevant for the reorganization of strong verbs.

The system of strong verbs was simplified and systematically restructured from PIE to Proto-Germanic, relying completely on form. Ablaut, which in PIE was an epiphenomenon of stress and affixation, was simplified, redistributed and heavily functionalized for tense/number marking, forming seven sub-patterns based on consonantal templates (similar to Semitic; for an extensive account, see Mailhammer 2007b). In this way purely formally conditioned subclasses arose, characterized by different ablaut alternation patterns and specific stem final consonants. The new class exponents (ablaut patterns) were not independent but were intrinsically tied to grammatical category exponence (tense/number) and were thus at the mercy of changes happening in these categories (see §3.2.2).

In contrast to the strong conjugation, the Proto-Germanic subclasses of weak verbs mainly contained newer derivations. They were clearly semantically motivated in Proto-Germanic through the derivational functions of their stem formatives: secondary suffixes are attested for PIE and became highly productive in Germanic and other Indo-European daughter branches. These subclass-constituting suffixes carried information on valence and situation type (e.g., transitive-causative, intransitive-durative/inchoative). They were most likely shaped by two processes in PIE and Proto-Germanic, namely grammaticalization and reanalysis.⁶

2.2.1 Grammaticalization

Although no complete grammaticalization path from free lexical item towards overt inflectional class marker has been documented, we can reconstruct likely scenarios.

6. A third way of introducing new allomorphy is assumed for the irregular group of preterite present verbs (cf. Birkmann 1987). By the Proto-Germanic stage, the former aspect/tense system had been transformed into a tense system. In the preterite present verbs, former perfect forms had become present forms, for example Old High German *weiz* – *wizzun* “know.3.SG – know.PL.PRS.IND. [preterite present class]” vs. Old High German *reit* – *ritun* “ride.3.SG – ride.PL.PRET.IND [strong class]”. This has been attributed to semantic causes. In Old High German *wizzan* “to know” (cf. the Latin perfect *vidi* “I have seen”), for example, the perfect denotes a new state after fulfilling the action “I have seen → so now I know”. Thus, this class is constituted not by material markers of its own, but by a unique paradigmatic distribution of markers shared with other classes.

The dental suffix of the preterite (e.g., German *-(e)te*, English *-ed*) is the macro class exponent of the weak conjugation. It can be most likely traced back to a Proto-Germanic periphrastic construction consisting of a non-finite lexical verb plus a finite auxiliary verb *do* (cf. Kiparsky 2009 with references; Ringe 2017: 191–194). The dental suffix increases allomorphy in Proto-Germanic, as it constitutes a new class of weak verbs alongside the older strong inflection characterized by ablaut patterns. Because the function of the suffix is to mark tense, it is tied to grammatical category exponence from the beginning of its development.

A scenario similar to the one just sketched might hold for the *j-subclass of the Proto-Germanic weak conjugation. As discussed above, PIE had opaque primary present formatives but also a layer of productive secondary stem formatives. The latter show transparent derivative functions, which, however, differ across the daughter languages. Thus, an original semantic basis cannot be safely reconstructed.⁷ Such formatives may have started out as autonomous lexical items, then developed by way of compounding into productive suffixoids, then into derivative suffixes, until they finally reached the stage between derivational suffix and opaque conjugation class marker documented in the earliest sources of Germanic languages. The Proto-Germanic weak *j-verbs go back to the PIE secondary present formative * $\check{i}e/\check{i}o$, deriving causative-iterative formations (Rix 2001: 22; Fortson 2010: 99). Szemerényi (1990: 299) considers the PIE roots *ei “go” und * $\check{i}o$ “make” as lexical sources. This is speculative for PIE, but it can be supported by the grammaticalization path from periphrastic verb classifier constructions to conjugation classes proposed for Pama-Nyungan languages by McGregor (2002: Ch. 8).

2.2.2 *Reanalysis*

As another morphologically driven path leading to an increase in allomorphy, i.e., to new (sub-)conjugations, overt class markers can evolve through reanalysis of the stem-suffix boundary.

The characteristic subclass-marker of the Germanic weak \bar{o} -verbs (PIE * \bar{a} - $\check{i}é$, cf. Latin *laudāre*, Old High German *lobōn* “praise”) has possibly been reanalyzed from the declension class marker PIE * \bar{a} of feminine noun stems (cf. Szemerényi 1990: 296; Meier-Brügger 2002: 169): Originally, PIE presents in * \bar{a} - $\check{i}é$ were denominal formations from feminine noun stems in PIE * \bar{a} (cf. Latin *planta* → *plantare* “plant”, noun → verb). The PIE feminine \bar{a} -stems were numerous and especially productive in deriving deverbal abstract nouns. When denominal * $\check{i}e/\check{i}o$ -verbs were formed from deverbal abstract nouns, speakers could secondarily assume a direct connection between the original verbal root and the verbal formation in * \bar{a} - $\check{i}é$.

7. The present formative **sk-*, for instance, marked inchoatives in Latin, causatives in Tocharian, and imperfectives in Hittite (Beekes 2011: 257).

This allowed reanalyzing the originally nominal stem extension $-\bar{a}$ as initial part of the verbal derivational suffix. After the reanalysis, $*-\bar{a}-\bar{i}\acute{e}$ became highly productive, deriving verbs from noun classes other than \bar{a} -stems in Germanic and other IE daughter branches.

2.2.3 Semantic conditioning: traces in Old High German

Proto-Germanic had at least three subclasses of weak verbs: the j -verbs (1st conjugation), the \bar{o} -verbs (2nd conjugation) and the \bar{e} -verbs (3rd conjugation). All three are formally preserved in Old High German and functionally still linked to Aktionsart and valence distinctions (see Table 2, following Schwerdt 2008).⁸

Table 2. Morphosemantics of the Old High German weak conjugation⁹

Conjugation (Old High German, infinitive)	State of affairs Semantic role	Valence	Derivational relation, Aktionsart	Example
1st <i>-en</i> Germanic $*-j$ PIE $*-\bar{e}j\acute{e}/-\bar{e}j\bar{o}$	Accomplishment Agent, Causer Goal: affected object	mainly transitive	denominal: factitives ^a ($<$ nouns), factitives ($<$ adjectives)	Old High German <i>weck-en</i> $<$ Germanic $*wak-jan$ “wake up” (transitive)
2nd <i>-ōn</i> Germanic $*-\bar{o}$ PIE $*-\bar{a}(j\acute{e})$	Accomplishment, Position Agent, (Causer), Positioner Goal: (un)affected object		denominal ($<$ nouns): factitives, ornatives (factitives $<$ adjectives) (deverbal)	Old High German <i>wahh-ōn</i> “watch over someone/ something”
3rd <i>-ēn</i> Germanic $*-\bar{e}-$ PIE $*-\bar{e}j\acute{e}/-\bar{e}j\bar{o}*\text{-eh}_1$	Position, Accomplishment, State, Change Agent, Positioner, Processed, Zero	mainly intransitive	denominal ($<$ adjectives) durative, inchoative	Old High German <i>wahh-ēn</i> “be awake”, “wake up” (intransitive)

^a Factitives mark the derivational relation of ‘making an object (= noun) appear’.

8. Gothic, the earliest attested Germanic language, had an additional fourth class of na -verbs (Proto-Germanic $*-n\bar{o}$) thought to be a language specific development although the North Germanic languages show traces of this class as well (Schwerdt 2001). The class was semantically based and contained ingressive verbs, for example *fullnan* “to become full”. These belong to the 3rd class ($-\bar{e}$) in Old High German.

9. For details regarding the PIE reconstructions in the leftmost column, cf. Szemerényi (1990: 295–301), Meier-Brügger (2002: 167, 174), Beekes (2011: 255–257).

Multiple class assignment may occur for one and the same root, as with the Germanic root *wak in Old High German *weck-en* (“wake up-INF”, transitive, 1st conjugation), *wahh-ōn* (“watch over someone/something-INF”, 2nd conjugation) and *wahh-ēn* (“be awake/wake up-INF”, intransitive 3rd conjugation). The semantic distinction is especially salient when we compare deadjectival j-verbs (transitive, causative) with deadjectival ē-verbs (intransitive, durative/inchoative). The 2nd class is semantically less clearly delineable, sharing transitivity with the 1st conjugation and being the most open class in all the early Germanic languages. However, verbs taking arguments with the semantic role of AFFECTED object tend to occur in the 1st conjugation (Schwerdt 2008).

The system was already blurred in Old High German, as more than half of the simplex verbs per class were ambiguous in their actional meaning (Schwerdt 2008: 166).¹⁰ The other early Germanic languages (apart from Gothic) are even more progressive in leveling the old semantically based system, reducing or abandoning the 3rd class: in Old Norse, long stems entered the 1st conjugation, while short stems remained in class 3; in West Germanic languages, all 3rd conjugation verbs changed towards the 2nd class (cf. Schwerdt 2008: 24).

In short, the Proto-Germanic conjugation class system was partly less and partly more semantically transparent than the declension class system. As far as strong verbs are concerned, stem extensions were desemantized as early as PIE, and reorganization in Proto-Germanic yielded a purely formal subclassification. For the weak verbs, however, semantic transparency was still strong: former derivational suffixes developed into inflectional class markers in verbs just as in nouns.

In the following section, we outline what happened to these inflectional class exponents in the four languages investigated.

3. Reorganization of inflectional classes in four Germanic languages

3.1 Number of classes

Generally, the number of inflectional classes increases slightly towards the earliest documented stages of the four languages. This is due to context-induced phonological change, most importantly *i*-umlaut, *a*-umlaut and, in Scandinavian, *u*-umlaut, which led to more complex paradigms and split the classes into subclasses. The

10. The function of differentiating Aktionsart was gradually transferred to verbal prefixation from Old High German on (for example, Modern High German *aufwecken* “wake someone up”, *bewachen* “watch over someone/something”, *erwachen* “wake up [intransitive]”).

main umlauting processes are assumed to have started in the 5th to 6th centuries CE, but written documentation is rare at this early point. Systematic occurrences of *i*-umlaut in German, for instance, are found in Old High German manuscripts from the 8th century on. Later, umlaut lost phonological productivity in these languages.¹¹ In combination with other factors, this development resulted in a decrease in inflectional classes, as discussed in the following section.

3.1.1 *Number of declensions*

The number of declensions decreases mainly due to the separation of case and number expression, which occurred in all four languages. Former case-number markers were reanalyzed as mere plural or case markers. The new plural markers were accordingly deleted from the singular subparadigms (cf. (3) for an example from Middle Dutch):

(3)		Early Middle Dutch		Late Middle Dutch	
		SG	PL	SG	PL
	NOM/ACC	<i>tonge</i>	<i>tongen</i>	<i>tonge</i>	<i>tongen</i>
	GEN/DAT	<i>tongen</i>	<i>tongen</i>	<i>tonge</i>	<i>tongen</i>

The reanalysis of *-n* as a plural marker came to completion when the singular paradigm of *tonge* was leveled: *tongen* > *tonge* in GEN/DAT.SG. The former portmanteau morphs were separated into two additive allomorphs (cf. Modern High German *Kind-er-n* “child-PL-DAT”).¹² In this process, case allomorphy was leveled, and case was even lost as a grammatical category in Dutch, Danish and Swedish (cf. §3.2.1).

Due to these developments, declension classes were increasingly based on number marking. The plural allomorphs primarily mark declension classes. Table 3 provides an overview of the productive and type frequent plural allomorphs in the four present-day languages. We provide orthographic representations for reasons of readability. They represent the same number of allomorphs as a phonological representation would.

11. Umlaut is already unsystematic (and rare) in Old Dutch, Old Danish and Old Swedish, and in German it becomes phonologically unproductive in late Old High German.

12. A possible counterexample is provided by weak masculine nouns in German which still show an (*e*)*n* marker throughout all oblique cases (i.e., in both numbers). However, in cases like Modern High German NOM.SG *der Tag* “the day” – GEN.SG *des Tages* – GEN.PL *der Tage*, there are good reasons to assume that a number suffix is followed by a case suffix: the plural exponent *-e* is found in all four cases (even in DAT.PL *den Tagen*), and there is no such element for all the forms in the singular paradigm, such that an analysis as *des Tag-Ø-es* vs. *der Tag-e-Ø* makes most sense in the genitive forms. There is also variation between the suffix forms *-es* and *-s* for GEN.SG: *des Tages* vs. *des Tags*; *-e* is not mandatory in the GEN.SG, but it is in the whole plural paradigm.

Table 3. Number and shape of plural allomorphs (UL = umlaut)

Dutch	Danish	Swedish	German
2	3	6	7
-s, -(e)n	-(e)r, -e, Ø	-ar, -er, -or, -r, -n, Ø	-(e)n, -e, -e (UL), -er, -s, UL, Ø

While Dutch and Danish retain only 2–3 allomorphs, Swedish and German retain 6–7 type frequent allomorphs. Today’s inventories are not the result of a straightforward reduction, however. In all languages apart from Danish, new plural allomorphs entered the inventory, namely *-s* in German and Dutch and *-n* in Swedish.

3.1.2 Number of conjugations

The two macro-classes, weak and strong verbs, are retained in all four languages. The same holds for the irregular class of preterite present verbs (see fnt. 6). However, the number of subclasses is reduced for weak verbs but diversified for strong verbs.

As sketched in §2.2, Proto-Germanic had at least three semantically conditioned weak conjugations (*j-*, *ō-*, and *ē-*stems), which were more strongly reduced in the two West Germanic than in the two North Germanic languages (see Table 4).

Table 4. Preterite allomorphs of the type frequent weak conjugations

	German	Dutch	Danish	Swedish
Number of weak conjugations	1	1	2	2 + 1 new
Allomorphs	-(e)te	-(t/d)e	-e-de, -te	-a-de, -(t/d)e, -dde

German and Dutch developed a unified weak class with merely phonologically conditioned allomorphy. Danish and Swedish retained the historical, now completely desemantized distinction between Proto-Germanic *j-* and *ō-*verbs, the vowel-initial dental suffixes descend from the latter.

As a notable exception to the general reductive tendency, 15th century Swedish developed a new conjugation for short stems ending in V: (for example INF *fly* “flee” – PRET *flydde* – PP *flytt*), which most likely emerged through reanalysis: a stem final dental stop was reanalyzed as belonging to the dental suffix (e.g., *tyda* “point at” – {tyd}{de} → {ty}{dde}), yielding a new reinforced allomorph *-dde* of the dental suffix, which was subsequently generalized for stems ending in V: (for example *fly* “flee” – {fly}{dde}; cf. Jansson 1966 [1947]; Dammel 2009).

The Proto-Germanic strong conjugation was formally based. The subclasses were constituted by regular formal consonant templates determining the ablaut alternation patterns (cf. Mailhammer 2007b). Later on, the strong verbs split into many more or less irregular subclasses, mainly due to context-induced phonological change. A comparison of the four languages regarding the average number of

verbs per ablaut pattern reveals a higher degree of disintegration (fewer verbs per pattern) in Danish and German (Table 5):

Table 5. Number of strong verbs and extent of their allomorphy (SAG 1999; Allan et al. 1998; ANS 1997; Duden-Grammar 1998)

	Swedish	Danish	Dutch	German
Number of ablaut alternation patterns	23	25	30	39
Number of strong verbs	111	100	176	171
Number of verbs per alternation pattern (average)	4.8	4	5.9	4.3

When we compare the number of weak and strong classes, the two West Germanic languages retain no weak subclasses but a higher number of strong verbs than the two North Germanic languages, which, in contrast, retain weak subclasses.

In all four languages, only one conjugation class is fully productive, namely the weak verbs, in Swedish and Danish specifically the subclass descending from the weak \bar{o} -verbs (Danish *-ede*, Swedish *-ade*). However, in Swedish and especially Dutch, some former ablaut series with relatively large lexical inventories have been semi-productive for weak verbs; cf. for example Dutch *blinken* – *blonk* – *geblonken* “gleam” and *wijzen* – *wees* – *gewezen* “to show” and the Swedish type *dyka* – *dök* – *dykit* “dip”.

3.1.3 *Number of inflectional classes in contrast*

The number of inflectional classes generally decreases, but to different degrees. German and Swedish retain more allomorphs in declension than Dutch and Danish do. In conjugation, weak subclasses are obliterated in German and Dutch but retained in Danish and Swedish. The reverse holds for the number of strong verbs and strong subclasses. They are fewer in Danish and Swedish than in German and Dutch. Comparing the two lexical categories, a higher number of productive classes is found in noun declension than in conjugation. Still, new allomorphs emerged both in declension and conjugation.

3.2 Exponence of inflectional classes

As sketched in §2, the PIE stem extension was placed between root and inflectional suffix in both nouns and verbs; the suffix was an independent morph. In the development towards Proto-Germanic and the Germanic daughter languages, formal transparency was lost. In nouns, the PIE tripartite structure turned into a bipartite one, caused by the fusion of stem extensions and case/number exponents. The process resulted in the formation of declension classes, which are reflected in allomorphic patterns of their host categories, case and number markers.

Regarding conjugations, the *j*-verbs lost part of their phonological integrity early: depending on syllable structure, the stem extension of many of them was deleted in the preterite. As a result, conjugation class exponence became dependent on the expression of the grammatical category tense. By contrast, the suffix of the weak *ō*-class was still formally transparent in the earliest documented stages of the Old Germanic languages, figuring in each cell of the paradigm. The macro-class distinction between strong and weak verbs was tied to the exponence of grammatical categories from the beginning, namely to tense/mood-exponents of different age (ablaut plus nasal suffix vs. dental suffix).

In the development of both lexical categories, we observe an increasing dependency of inflectional class markers on the exponence of grammatical categories.

Bybee's hierarchy of relevance (1985) provides a key to investigate these restructuring processes. The hierarchy is based on the principle of relevance and was tested in a typological survey of verbal inflection. The notion of relevance is defined as follows: "A meaning element is *relevant* to another meaning element if the semantic content of the first directly affects or modifies the semantic content of the second" (Bybee 1985: 13, her emphasis).

According to Bybee, voice, aspect and tense are highly relevant categories of verbs. They affect the verbal action directly, and – in the case of aspect, which is ranked higher than tense – they modify it in its structure. Number and person are less relevant, influencing only the players but not the action itself.

Bybee (1994: 2559) also proposed a hierarchy of relevance for nouns. According to this hierarchy the semantic information of the root is highly affected by number, displaying it as individual or multiple. Case, by contrast, relates the noun to other syntactic constituents but is irrelevant for the actual meaning of the noun. In this line of reasoning, number is more relevant to nouns than case. The relevance hierarchies for nouns and verbs are sketched in Figure 1.

		+ Relevance		–	
		Inherent categories		Contextual categories	
Noun	Number	Case			
Verb	Voice Aspect Tense	Mood	Number	Person	

Figure 1. Hierarchies of relevance for nouns and verbs

Bybee claims that functional relevance and formal fusion obey diagrammatic iconicity: the more relevant some grammatical information is for its lexical base compared to other categories, the closer its formal expression should be placed to

the root and the more it should be able to modify the root internally. Less relevant information, by contrast, should tend to be placed more at the periphery of the word form. This claim and the two hierarchies of relevance allow us to hypothesize about change at the morphology-phonology interface of Germanic languages: as the four languages concerned show a suffixing preference, initial stress, and a diachronic reduction of unstressed syllables, positioning highly relevant information close to the root should protect the markers against reductive phonological change, while the positioning of less relevant information at the periphery of the word form should result in a higher exposure to reductive change.

In German, for example, number is always expressed before case in nominal morphology, as in *den Pferd-e-n* “horse-PL-DAT”, and, by contrast with case, sometimes even fuses with the stem resulting in stem allomorphy, as in *des Hammer-s – den Hämmer-n* “hammer.SG-GEN – hammer.PL-DAT”. The same holds for verb forms: Tense for example is expressed closer to the root than person/number, as in *sie lach-t-en* “laugh-PRET-3.PL”. Only tense frequently fuses with the root in strong verbs, as in *sang-en* “sing.PRET-3.PL”.¹³

Coming back to the change of inflectional classes, this means that more stability is predicted for those class distinctions hosted by relevant grammatical categories than for those tied to less relevant categories. We show below that this is in fact the case. Moreover, the allomorphy hosted by relevant categories is even increased in the four languages (cf. §3.1).

3.2.1 *Exponence of declension classes*

The highly relevant category of number is retained and strengthened in the declension classes of all four languages, as specific suffixes and unlauded roots are reserved exclusively for number expression. In the less relevant category of nominal case, German reduces allomorphy considerably, and in the other three languages the category of nominal case is totally lost.

Fused expression of case and number was inherited from PIE, and it still shapes the paradigms of the oldest Germanic languages. However, in all four languages this portmanteau was divided into two separate morphs later on. In Modern Swedish *skol-a-s* “school-SG-POSS” vs. *skol-or-s* “school-PL-POSS”, the number markers are found before the possessive marker, which is the remainder of a former genitive case marker.¹⁴

13. The categories of aspect and diathesis are not considered here, as their expressions in the four languages are not inflectional (aspect) or not directly comparable (passive: in Swedish and Danish, suffix/periphrasis; in German and Dutch, different type of periphrasis).

14. The *s*-marker turned from a case marker into a phrasal clitic in the history of Swedish (cf. Norde 1997). Comparable developments happened in Danish and Dutch, just as in English.

With respect to number marking, the number of suffixes is reduced by Middle High German, Middle Dutch and Old Danish, since all unstressed vowels are reduced to schwa. Compared with Middle High German, Middle Dutch and Old Danish lose even more suffixes, because umlaut is not adopted for number marking. Additionally, in Middle Dutch Ø-plurals are totally lost. Material number exponence is strengthened in Middle High German and Middle Dutch by heavily reducing the number of Ø-plurals, a high number of which had been inherited from Proto-Germanic (cf. for example Middle High German/Middle Dutch: *wort* – *wort* “word.SG – word.PL” > late Middle High German plural *wort* – *wörter/worte* “word.SG – word.PL”; Middle Dutch *wort* – *worde* “word.SG – word.PL”). In Modern Dutch, Ø-plurals are no longer found with nouns. By contrast, Danish and Swedish retain Ø-plurals with many neuter nouns.

In the Old Germanic languages, context induced phonological change, most prominently umlaut, caused root vowel alternations. In German, umlaut was restricted to back vowels which palatalized before *i* in a following syllable (*i*-umlaut; cf. Old High German *kalb* – *kelbir* “calf.SG – calf.PL”). The Scandinavian languages additionally underwent *u*-umlaut (for example Danish *barn* – *børn* “child.SG – child.PL”). In its phonologically conditioned distribution, umlaut was found in the paradigms of both numbers (cf. for example Old High German *anst* “grace.NOM.SG” – *ensti* “grace.GEN/DAT.SG” and “grace.NOM/ACC.PL”), but after becoming phonologically unproductive, umlaut alternations were redistributed on a morphological basis, with the new function of marking the number feature: umlaut became restricted to plural forms. This happened regularly in German, but only exceptionally in Danish and Swedish. Only in German did plural allomorphs including umlaut become morphologically productive. Dutch lost almost all umlaut alternations, the sole relic being the alternation *stad* – *steden* “town.SG – town.PL”.

It fits well with the concept of relevance that umlaut is retained and morphologized in number but not in case marking: the highly relevant number information, on the one hand, is expressed in a secure position next to the root or even root-internally. Case, on the other hand, is always expressed at the right margin of the suffix chain. While number is retained as a grammatical category in all four languages, German is the only one which still has nominal case today. The other three languages lost it as a nominal category.

3.2.2 *Exponence of conjugation classes*

The influence of relevance is also manifest in the development of conjugation classes. Just as in Proto-Germanic, person and number are still fused in a portmanteau suffix in Modern German and Dutch. However, the Old Germanic languages had considerable allomorphy for person/number; this category bundle participated duly in the early conjugation class distinctions. However, this allomorphy was leveled

in all four modern languages investigated, and class neutral uniform person/number suffixes developed, excluding the least relevant categories person and number from determining conjugation class.¹⁵ Figure 2 exemplifies this process for the 2nd person singular in German. Although in this case tense is indirectly affected as well (the former person/number-exponents were tense specific), we show below that allomorphy hosted by tense alone is retained in all four languages.

	Middle High German	Modern High German
Preterite-present verbs (present)	-t <i>du darf-t</i>	<i>du darf-st</i> “you may”
Strong verbs (preterite)	UL+ -e <i>du fűer-e</i>	<i>du fuhr-st</i> “you drove”
Elsewhere	-s(t) <i>du lobe-st</i>	-st <i>du lob-st</i> “you praise”

Figure 2. Reduction of allomorphy hosted by person (exemplified for 2 sg.; UL = umlaut)

Morphological change in the two Scandinavian languages shows that the development of uniform symbolization of a grammatical category can be the first step towards its full deflection: in Danish and Swedish, the development of generalized person/number markers was followed by the complete leveling of these two categories (cf. today’s uniform present in -(V)r, for example Danish *spis-er* “eat.PRESENT”; Dammel 2011: 103–107).

The mirror image of developing uniform symbolization is observable for the highly relevant category tense. The ablaut alternations of strong verbs were formerly not only used for tense marking, but also for number marking. All four languages have retained them as a tense marking device. By contrast, the alternations marking number in the preterite were leveled in all four languages in the late Middle Ages, for instance Dutch *fand* (“find.PRET.SG”) – *fonden* (“find.PRET.PL”) → *fond, fonden* (“find.PRET.SG, PL”), German *fand* (“find.PRET.SG”) – *funden* (“find.PRET.PL”) → *fand, fanden* (“find.PRET.SG, PL”). As a result, only the relevant category tense is marked by ablaut today.

15. The only exceptions are preterite present verbs in German and Dutch, which have -Ø in 1/3SG.PRS (e.g., German *ich/sie weiß* “I/she know(s)”) opposed to person marking in all other classes (e.g., German *ich schweiß-el/sie schweiß-t* “I/she weld(s)”). Furthermore, stem alternations for the bundle person/number/tense/mood occur in the present forms of about 55 German strong verbs (for example umlaut in *ich fahre, du fährst, sie fährt* “to drive.1–3SG.PRS.IND”).

As sketched in §3.1.2, tense allomorphy was not only retained but increased; in strong verbs by splitting up the ablaut series, in the Swedish weak verbs by a new dental suffix allomorph *-dde* (cf. §3.1.2). The new allomorph not only increases tense allomorphy, but its development was presumably also morphologically motivated by avoiding tense homonymy (Jansson 1966 [1947]; Dammel 2009).

Stem alternations arose through *i*-umlaut and Northwest Germanic raising of *e > i* (triggered by following *i*, *j*, and *u*). The four languages deal with these new alternations in quite different ways, by and large in parallel with the situation sketched for nouns above: only German retained the alternations and even morphologized umlaut for expressing subjunctive mood in strong verbs.¹⁶ In the other three languages, mood has become a marginalized morphological category.

In sum, conjugational allomorphy is best retained – and even increased – in the category tense. This holds consistently for the DERIVED features of tense: preterite and past participle. In the basic feature present, allomorphy is reduced. For most verbs of contemporary German, Dutch and Danish, conjugation class is therefore only distinguishable in the preterite/past participle.¹⁷ This means that tense expression dominates inflectional class expression.

3.2.3 *Exponence of inflectional classes: comparison*

In all four languages and in nouns as well as in verbs, inflectional class change is subject to the hierarchy of relevance (Bybee 1985). In relevant categories, allomorphy is retained and even increased, whereas allomorphy is reduced to uniform expression or even completely lost (cf. case and person/number in Danish and Swedish) in less relevant categories (Dammel & Nübling 2006).

In both lexical categories and all four languages, inflectional class exponents stemming from autonomous morphemes gradually fused with markers of grammatical categories and became dependent on the expression of their host category. After that, class distinctions were reduced to those subparadigms comprising the derived features of the relevant categories number (noun) and tense (verb). Therefore, inflectional class expression is now dependent on grammatical category expression, as the choice of a grammatical category feature (tense: preterite; number: plural) determines whether inflectional class is overtly expressed or not.¹⁸

16. Subjunctive forms were based on the former PRET.PL stem (*fund-* “fnd.PRET.PL.IND → *fünd-* “fnd.SBJV”). This stem was leveled in several ablaut series in Early Modern High German, and new subjunctive forms developed based on the generalized PRET.SG stem *fand-*, applying umlaut morphologically: *fänd-* “fnd.SBJV”.

17. In Swedish, class distinctions in the present have been better retained. The stem extension *-a* plus accent II are distinct class features of the 2nd conjugation (former *ö*-verbs).

18. For other types of dependencies between grammatical subsystems see Aikhenvald & Dixon (1998).

The distribution of stem alternation throughout the nominal and verbal lexicon is controlled by the external factor of token frequency in all four languages. Only token frequent nouns and verbs retain or adopt stem alternation (e.g., Wegener 2003 for German nouns; Nübling 2000 for verbs in 10 Germanic languages). As a divergence in language contrast, we found that German alone has morphologized umlaut and applies it quite regularly in nouns (number) as well as in verbs (mood).

Irregularities such as stem alternation and suppletion are more frequent in verbs than in nouns. One reason for this difference is observable in any random frequency dictionary (e.g., Allén 1971): more verbs than nouns occur with highest token frequency values. Stem alternations in verbs are more often used and therefore more easily acquired and better entrenched in the mental lexicon.

3.3 Conditioning of class membership

The assignment of lexemes to inflectional classes (conditioning) can be controlled by several principles (cf. Neef 2000a, 2000b). Figure 3 presents a scale of assignment principles.

<i>signifiant</i> -based		<i>signifié</i> -based	
final sound	prosody	semantics	gender idiosyncratic/ lexical

Figure 3. Conditioning of class membership

As a basic distinction, we differentiate between *signifié*-based assignment principles (often called lexical conditioning) and *signifiant*-based assignment principles (phonological conditioning). Phonological conditioning can be segmental or suprasegmental. A textbook example of segmental phonological conditioning is the English plural allomorphy with /əz/ after sibilants (*pages*), /z/ after voiced non-sibilants, and /s/ after voiceless non-sibilants (*chapters*, *books*). In supra-segmental conditioning, the number of syllables or an accent pattern is decisive.

Signifié-based assignment principles can be transparent, based on semantic features (such as concrete/abstract, mass/count or animacy in nouns and Aktionsart, valence, or modal function in verbs), or non-transparent, based on lexical gender, which is arbitrarily assigned in parts of the lexicon. If an allomorph is not predictable at all by general conditioning factors, its conditioning is idiosyncratic/lexical; in this case, memorization is necessarily item-specific. The scale holds for both lexical categories (with gender lacking in verbs).

In §3.3.1, we sketch the major developments regarding declensions. The languages differ considerably in the degree to which gender is retained as a conditioning

factor or superseded by semantic and formal conditioning. Conjugations (see §3.3.2) only exceptionally developed semantic conditioning, drifting to lexical conditioning and heading for formal conditioning factors instead.

3.3.1 *Conditioning of declension classes*

Declension classes were reorganized according to gender in the Old Germanic languages. In this way, two categories with dubious functional status reinforced each other (though gender with its syntagmatic scope is functional at least in reference tracking). The four languages take quite different directions after that:¹⁹ German and Swedish retain GENDER AS A CONDITIONING FACTOR, while Danish and Dutch abandon it. German is the only one of the four languages keeping up a three gender system. In the other languages, only two genders remain after a merger of masculine and feminine into common gender.

Although German retains three genders, two mostly form coalitions regarding the choice of allomorphs. The nature of this coalition changes from Old High German to Modern High German. In Old High German, masculine and feminine declensions grouped together versus the neuter. In the development from Old High German to Modern German, this configuration broke up. Phonological reduction affected many feminine nouns, resulting in Ø-plurals, for instance Old High German *geba* – *gebā* > *geba* – *geba* > Middle High German *gebe* – *gebe* “gift.SG – gift.PL”. To avoid Ø-plurals, such feminines changed their declension class in Early Modern High German, increasing the group of feminine *n*-plural nouns (for example *Gabe* – *Gaben* “gift.SG – gift.PL”). In this way, the *n*-plural has become strongly linked to feminine gender today, paving the way for a new opposition in plural marking between feminine and non-feminine nouns (cf. Nübling 2008).²⁰

Swedish is the only one of the three languages reducing genders from three to two that retains gender as a conditioning factor of declensions. A distribution stemming from Old Swedish is maintained towards Modern Swedish: Common gender nouns mainly take the plural allomorphs *-ar*, *-er*, *-or* and *-r*, neuter nouns take Ø. Additionally, a new neuter suffix, *-n*, entered the system for nouns ending in

19. Cf. Kürschner & Nübling (2011) for a broader comparison of Germanic languages and dialects of German.

20. Only two groups of masculine nouns retain *n*-plural as well: the so-called weak and mixed masculines. Comprising together about 200 native nouns, these classes are rather insignificant in comparison to the feminines of the *n*-plural class. The weak masculines do not only take the *n*-plural as the feminines do, but in fact form a declension class of their own with *n*-markers in the singular oblique cases (cf. NOM.SG *der Junge* “the boy” – GEN.SG *des Jungen* – NOM.PL *die Jungen*). Additionally, there are seven (formerly weak) neuter nouns with *n*-plural (numbers according to Augst 1975 and Kürschner 2008: 135–136).

V: (*bi* – *bin* “bee.SG – bee.PL”). It developed during the 14th–15th century through reanalysis of a definiteness marker (Kågerman 1985) and reinforced gender as a conditioning principle.

In Danish, gender was lost as a conditioning factor with the Old Danish reduction of unstressed vowels (which Swedish did not undergo). The number of plural allomorphs was reduced to three, neuters remaining distinct by their Ø-plural. On the way to Modern Danish, more and more neuter nouns adopted the former common gender allomorphs (*-e*, *-er*) with material number marking, for instance Old Danish *land* – *land* > *land* – *landæ* “land.SG – land.PL”, *knæ* – *knæ* > *knæ* – *kneer* “knee.SG – knee.PL”. Recently, Ø-plural has been opened for common gender nouns, obliterating the last gender based distinction in the conditioning of Danish plural allomorphy, for instance *laks* – *laks* “salmon.SG – salmon.PL”.

Dutch lost the inherited gender conditioning by the Middle Dutch period, just as in Danish mainly due to reduction of unstressed vowels. The only remaining type frequent allomorphs *-(e)n* and *-s* developed phonological conditioning (see below).

SEMANTIC (RE)CONDITIONING is quite common in nouns. Animacy is already part of the Proto-Germanic system in that many of the weak masculine and feminine nouns share the feature ‘animate’ (e.g. Froschauer 2003: 490). Still, at this point, semantic conditioning is subordinated to gender conditioning, as it only applies to certain genders in one and the same declension class (for example, weak).

In later stages, semantic conditioning independent from gender could be established. The most straightforward case is the Danish *e*-plural. In monosyllabic simplex nouns and some derivatives, the *e*-plural attracts nouns with the feature ‘animate’ from other classes (for example *smed* – *smeder* > *smede* “smith.SG – smith.PL”) as well as neologisms denoting animate entities, for instance *laps* – *lapse* “dandy.SG – dandy.PL”. This is especially clear with derivatives in *-ing* denoting humans, which prefer the *e*-plural, while those denoting non-humans take an *er*-plural, for example *islænding-e* “Icelanders” vs. *melding-er* “news” (cf. Kürschner 2008: 220–223 for a more detailed account).

In Dutch, the *s*-plural, which entered the declension system in late Middle Dutch (Philippa 1982), started out semantically conditioned. It was first restricted to agent nouns in *-er*, e.g., Middle Dutch *richt-er(e)-s* “judge-PL”, *vercop-er(e)-s* “clerc-PL”. The conditioning changed, first extending the *s*-plural to nouns denoting human beings, for instance *vader-s* “father-PL”, *kok-s* “cook-PL”, and then towards a form-based conditioning: all stems ending in schwa + sonorant take *-s* today, regardless of their semantics, for example *vogel-s* “bird-PL”, *bliksem-s* “lightning-PL”, *teken-s* “sign-PL”.

Semantic conditioning dominated by gender developed in German and Swedish. The German weak masculine declension is characterized by *(e)n*-plural and *-(e)n* in all oblique singular forms. For this class, membership became

restricted to nouns denoting higher animates (cf. Köpcke 2000): From Middle High German on, masculine nouns denoting inanimate entities have systematically left this class.²¹ Some masculine nouns denoting animate entities even took the opposite direction and entered the weak declension, for instance *Hirte* “herdsman”, *Christ* “Christian”.²²

Concerning FORMAL CONDITIONING, Dutch is the prime example. The two allomorphs of the Dutch plural, non-syllabic *-s* and syllabic *-en*, are subject to a general prosodic well-formedness condition yielding plural forms ending in a trochaic foot (Booij 2002: 24; van der Hulst & Kooij 1998: 188; Kürschner 2009). The suffix *-en* is used with stressed final syllables (for example *tas* – *tassen* “bag.SG – bag.PL”), while *-s* is used with unstressed final syllables (for example *spiegel* – *spiegels* “mirror.SG – mirror.PL”).

Prosodic conditioning fundamentally shapes the German declension system as well (Neef 2000b): Most German plural forms are subject to the output condition ‘end in a reduced syllable’, that is, a syllable containing a reduced vowel (schwa).²³ Masculine and neuter nouns with singular forms already fulfilling this condition do not take a material plural marker, for instance masculine *Löffel* – *Löffel* “spoon.SG – spoon.PL”, neuter *Messer* – *Messer* “knife.SG – knife.PL”. Feminine nouns with singular forms in a reduced syllable take the non-syllabic variant of the *(e)n*-suffix, for example *Gabel* – *Gabel-n* “fork.SG – fork.PL”. All allomorphs apart from *-s* are in complementary distribution according to the reduced syllable condition:

- *(e)*: *Hund-e* “dog-PL” vs. *Messer-Ø* “knife-PL”
- *(e)-UL*: *Gründ-e* “reasons-PL” vs. *Gärten-Ø* “garden.PL”
- *(e)n*: *Bär-en* “bear-PL” vs. *Kunde-n* “customer-PL”
- *er* (only with stressed syllables): *Kind-er* “child-PL”

The conditioning of German plural allomorphy is thus controlled both by gender (see above) and this prosodic principle in combination.

To sum up, different conditioning principles prevail in different languages. Most saliently, gender conditioning is strengthened in German and Swedish but abandoned in Danish and Dutch. While German combines gender with the prosodic

21. These nouns either changed their declension towards strong declension (by adopting *-n* in nominative singular, for example *haufe* > *Haufen* “heap”) or changed their gender to feminine (for example *der Fahne* > *die Fahne* “the.M > F flag”).

22. Swedish weak masculine nouns underwent a process similar to German, conditioning class membership according to animacy (for example Wurzel 2001: 104, 147). The weak masculines, characterized by stem inflection today (e.g., *pojke* – *pojkar* “boy”), have also been open for animate loan words.

23. In consonant-final words, schwa is in fact often left unpronounced in favor of reducing it to a syllabic consonant (cf. variation between [ˈlœfəl] and [ˈlœf̩] for *Löffel* “spoon”).

principle of the reduced syllable plural (formal conditioning), Swedish declensions are primarily conditioned by gender.

Dutch reduces *signifié*-based conditioning completely for the sake of one consistent formal principle (trochee plural). The conditioning system of Danish is less clearly dominated by only one or two principles. It has been open to formal conditioning (polysyllabic nouns tend to take the *er*-plural), but by now, no consistent formal principle is observable. For the *e*-plural, semantic conditioning prevails, and with the \emptyset -plural, remainders of the former conditioning by means of neuter gender are still recognizable.

3.3.2 *Conditioning of conjugation classes*

While in nouns, semantic conditioning commonly occurs at least as a subregularity in the four languages, in verbs we only find one small conjugation class to be semantically conditioned, but consistently so across Germanic languages. These are the preterite-present verbs (for example Modern High German *müssen* ‘have to’, *dürfen* ‘be allowed to’, *können* ‘be able to’, etc.), which evolved into the class of modals (Birkmann 1987). All other conjugation classes developed formal or idiosyncratic conditioning.

Strong verbs started out in Proto-Germanic with relatively few, clearly phonologically conditioned subclasses. Through context-induced sound change and different directions in morphological levelling, this neat formal assignment turned more and more into idiosyncratic conditioning. Nevertheless, ‘islands of reliability’ remain (Albright & Hayes 2003; Mailhammer 2007a). E.g., the sequence *-i* + nasal + consonant (infinitive, present), as in *sing*, is still a reliable indicator of strong inflection in several Germanic languages (e.g., Bybee & Moder 1983 for English). Some former ablaut series with relatively many members of Swedish and especially Dutch have been semi-productive, integrating weak verbs based on their formal shape (see §3.1.2). In most other cases, the information of strong inflection has to be memorized item-specifically for each verb. This effort is easily managed under conditions of high token frequency, which turns out to be the system-external common denominator of most strong verbs (that is, those not supported by a type frequent phonological schema). However, we do not regard token frequency as a conditioning factor in the narrow sense, because conditioning is defined as system internal and compulsory. The latter is not the case, as token frequent weak verbs do not become strong.

The conditioning of the weak subclasses was still in part semantically transparent at the earliest stage of at least German, organized in Aktionsart and valence distinctions (see §2.2). In the late medieval varieties, formal conditioning increased for the weak subclasses, but to different degrees in the North versus West Germanic languages (Table 6):

Table 6. Weak subclasses in comparison (exemplified for 1/3 SG. PRET. IND.; bold lines depict lexically conditioned class distinctions)

Proto-Germanic	*-(i)da	*-ō-da	*-ē-da
Swedish	-(t/d)e	-a-de [adə]	-dde (< 15th ct.)
Danish	-te	-e-de [əðə]	
German	Middle High German	-te	-e-t-e
	Modern German	-(e)te	
Dutch		-(t/d)e	

The systems of Swedish and Danish retain a lexical class distinction marked by the thematic vowel *-a/-e* (< former *ō*-verbs). In Swedish, this vowel occurs independently throughout the whole paradigm, while in Danish, it is neutralized in the infinitive and present and in this way has become linked to tense expression (see (4)).

- (4) present preterite translation
Swedish *kastar, köper* *kastade, köpte* “throw”, “buy”
Danish *kaster, køber* *kastede, købte* “throw”, “buy”

Subordinated to this primary lexically conditioned distinction ‘thematic vowel: yes/no’, the shape of the Swedish dental suffix is phonologically conditioned by the final sound of the root. Roots ending in a voiced final sound get *-de* (cf. *kasta* – *kastade* “throw”, *ärva* – *ärvde* “inherit”), and roots ending in a voiceless final sound get *-te* (cf. *köpa* – *köpte* “buy”). The new 3rd weak conjugation developed in Middle Swedish also seems to be phonologically conditioned at first glance: it contains only short roots ending in a long stressed vowel, such as *tro* “believe”. They take the suffixes *-dde* (preterite) und *-tt* (past participle). Nevertheless, minimal pairs such as *roa* “amuse” (2nd = former *ō*-conjugation) vs. *ro* “row” (new 3rd conjugation) show that the conditioning is phonological-lexical.

Dutch and German weak subclasses changed completely from lexical to phonological conditioning, but the two languages set opposite priorities: Dutch lost the variant of the dental suffix with an initial vowel. The only remaining suffix is PRET *-(t/d)e*, which – just as in Swedish – assimilates to the stem final sound: *-te* with voiceless, *-de* with voiced final sound, for example [x]: *poch-te* “boast-1/3.SG.PRET”, [ʏ]: *veeg-de* “sweep-1/3.SG.PRET” (Booij 2002: 57–58). German retains both formal variants *-ete* und *-te*, but they are now totally unrelated to the former lexical *ō*- vs. *j*-class distinction. This is a fundamental difference compared to the

two North Germanic languages. The variation was completely reorganized, the ratio behind the new phonological conditioning being formal distinctness of stem and grammatical information: stems ending in a dental plosive and lacking vowel alternation take the variant with initial schwa, e.g., *breit-ete* “spread-PRET” vs. *leg-te* “lay-PRET”. This rule keeps the lexical root and the grammatical information (dental suffix) formally distinct. It ‘bites’ exactly when the morphological information carried by the suffix is not marked by vowel alternation. Consider for example the two preterite-variants of *senden*: *sand-te* [zantə] and *send-ete* [zɛndətə], both “send-PRET.1/3.SG”, but not **sendte*, **sandete*. With vowel alternation present, the stem final dental and the suffix initial dental fuse. If the dental suffix is the only tense exponent, it is always kept distinct by schwa insertion (Harnisch 1999). In conclusion, although in both German and Dutch the conditioning is purely phonological, it follows the speaker-friendly principle of assimilation in Dutch (for example *praten* [pra:tə] – *praatte* [pra:tə] “speak.INF – speak.PRET”), whereas in German the hearer-friendly principle of distinctness works.

3.3.3 Conditioning of inflectional classes: comparison

We found interesting areal convergences and clear contrasts between declension and weak conjugation, schematized in Figure 4.

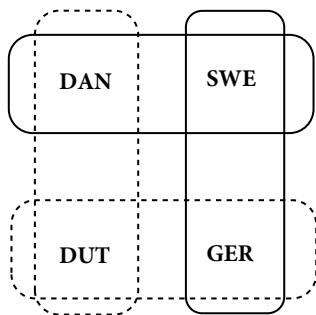


Figure 4. Comparing change in inflectional class conditioning (horizontal boxes: weak conjugations; vertical boxes: gender in declensions)

Danish and Dutch (dotted vertical box) group together in abandoning gender as a conditioning factor for declensions. Swedish and German (solid vertical box), by contrast, retained and even strengthened conditioning by gender. Note that the change in conditioning declensions crosses subfamily boundaries. In contrast to nouns, the conditioning of weak verbs is family business: the conditioning is completely form-based in the two West Germanic languages (dotted horizontal box), while the North Germanic languages continue the lexically conditioned Proto Germanic *j*- and *ō*-verbs (solid horizontal box).

4. General discussion

4.1 Declensions versus conjugations

When we compare the behavior of the two lexical categories, it is obvious that nouns, in contrast to verbs, tend to reorganize allomorphy in *signifié*-based conditioning, mainly based on gender and animacy. One reason for this is trivial: verbs do not have any gender-like second classification system to link conjugation classes with. Looking at semantic re-conditioning, it also seems to be more common in nouns. The small class of preterite-present verbs forms a counter-example having specialized for modal function across the four (and more) Germanic languages. However, modality structures only a minimal part of the conjugational system compared to more comprehensive nominal conditioning by gender (and to a lesser extent animacy), which structures the entire nominal lexicon.

Furthermore, nouns retain more productive declension classes compared to only one fully productive conjugation class per language. However, this comparison is not entirely sound because the number of verbs is much smaller than the number of nouns. And how should one evaluate for instance the semi-productivity of strong verb ablaut observed in Dutch and Swedish (see §3.1.2)?

In search of possible explanations for the observed noun-verb-differences, one could argue that verbs normally bear more grammatical information (inflectional categories) than nouns and – due to this higher functional load – do not tolerate as much additional complexity of inflectional classes as nouns do. However, the four languages show no substantial difference in the number of grammatical categories between nouns and verbs.

A second argument seems more promising: there might be differences in the cognitive categorization of verbs and nouns, with nouns providing more accessible starting points for classification. As nouns prototypically denote concrete entities, relations of contiguity and similarity (the basis of the metonymic and metaphoric chains observed in noun classification) are more salient for them than for verbs (cf. McGregor 2002: 34; for a discussion from a psycholinguistic perspective, see Mätzig et al. 2009).

While nouns have their functional domain in reference, the function of verbs is predication (cf. Bisang, this volume). This explains why a common diachronic source of declension classes lies in (semantically more or less transparent) classification systems with syntagmatic scope (agreement) which are central in reference tracking (genders, noun classes). Syntagmatic scope can be lost as grammaticalization advances, leading to declension classes (cf. Wurzel 1986; Grinevald 2002).

Conjugations may evolve from periphrastic constructions specifying the kind of action involved, for instance by marking Aktionsart or valence distinctions (cf.

McGregor 2002: Ch. 8 and discussion in §2.2). Classification in constructions such as these could be regarded as a mere side effect of grammaticalization. However, in our view it would be premature to conclude that classifying verbs cannot be an end in itself, based only on the argument that verbs do not function in the same way as nouns do and verb classification can carry grammatical information.

4.2 Refunctionalization

Going back to the scenarios in §1, a first general observation from our diachronic outline is that inflectional classes – declensions as well as conjugations – are not simply abandoned but retained over thousands of years. The variation at first increases rather than decreases as phonological processes introduce new stem alternations (mainly umlaut). Later changes lead to a moderate reduction of classes in both lexical categories, but in none of the four languages are inflectional classes abandoned completely. The wider Germanic family includes counterexamples with English for declension classes (generalized *s*-plural) and Afrikaans for conjugations.²⁴ Nevertheless, none of the Germanic languages has lost inflectional classes in both lexical categories at the same time.

As a parallel in language contrast, we find that classes are not simply retained, but reorganized. In these four languages, we find parallel factors governing this reorganization:

Change in inflectional class exponence is governed by the HIERARCHY OF RELEVANCE in all four languages and for both lexical categories. Formerly autonomous class exponents become dependent on grammatical category expression, thereby sharing the fate of the category. In the exponence of relevant categories, allomorphy is retained due to its safe place close to or in the root. The exponence of less relevant categories reduces allomorphy approaching uniform expression. Nouns have retained allomorphy for number, while the less relevant category case has been lost except in German, where it is now close to uniform expression. Number exponence in German and Swedish even gained new allomorphy, with German morphologizing umlaut and adopting the new allomorph *-s* and Swedish developing the marker *-n*. Verb allomorphy hosted by the relevant tense category is retained in all four languages and even increases, while the less relevant categories of person and (verbal) number became uniform or – in the case of the Scandinavian languages – underwent complete deflection.

24. If we look at the chronological order in which deflection hit the verbal categories of Afrikaans, it is remarkable that conjugation classes are not the first but the last verbal category to be leveled, long after the functional categories person, number and finiteness, and along with tense, the host category of the conjugation classes (Dammel 2011: 107–113).

Change in the CONDITIONING of inflectional classes reveals differences between both languages and lexical categories. In declension class change, German groups together with Swedish, retaining and partly strengthening gender as a conditioning factor, while Danish and Dutch abandon it. The role of conditioning by gender is quite interesting: Linking gender and declension slows down or even inhibits the reduction of both classification systems. The linkage seems to be profitable, with both classification systems backing one another up in their memorability (cf. Kürschner & Nübling 2011; Wurzel 2001).

Furthermore, conditioning in German and Dutch follows prosodic principles (German reduced syllable plural, Dutch trochee plural). Both structures match the typical phonological patterns for inflected word forms, which facilitates morphological interpretation. This means that both in gender based and in prosodic conditioning, the reorganization of inflectional classes strengthens other linguistic subsystems and stabilizes the formal variation at the same time.

Alongside the two conditioning factors, gender and form, one of the most salient semantic features, animacy (including humanness), shapes the conditioning of declension classes in all four languages. This type of restructuring can be viewed as a back-to-the-roots movement, as the basis of the nominal classification system in Indo-European was semantic. However, semantic conditioning in the four languages mostly developed as a minor principle dominated by gender or form.

The only case of semantic reconditioning in the domain of conjugations is the preterite-present inflection reserved for modal verbs. For other conjugations, reconditioning followed formal principles, with different degrees of formalization in the four languages. Dutch has the most regular strong inflection with semi-productive ablaut series based on formal features. Swedish and Danish retain several lexical subclasses of weak verbs and continue the Proto-Germanic system, intertwined with a newly developed formal conditioning. Dutch and German lost the historical weak subclass distinctions completely and reconditioned the allomorphic variation according to formal principles. They developed one single weak conjugation, the allomorphy of which serves different phonological needs: it facilitates production in Dutch and guarantees distinctness of stem and suffix in German. In this way, the formal conditioning of conjugations serves the needs of another grammatical subsystem, similar to what we observed for nouns above.

5. Conclusions

Two main factors shape the reorganization of inflectional class systems here: (a) the functional principle of relevance predicting the order and consequently the robustness of allomorphs; (b) the interaction of different conditioning factors

serving functions in other linguistic subsystems such as grammar (gender, modality), semantics (animacy) and phonology (trochee output, voicing assimilation). Some of the languages tend to exploit one conditioning factor exclusively (Swedish nouns: gender; Dutch nouns: prosody), while others combine several conditioning factors (German nouns: gender and prosody; Swedish weak verbs: lexical and phonological conditioning). Divergences between lexical categories are common (for example Swedish with gender in nouns and lexical-phonological conditioning in verbs).

Although the present study did not go into detail with respect to performance-based factors, these should not be neglected. Token frequency is the main factor shaping the distribution of regular and irregular allomorphs throughout the nominal and verbal lexicon.

Finally, one question remains to be discussed, namely whether reorganizing inflectional class systems in the way described in our case studies implies an increase in functionality of the allomorphic variation. In our case studies, functionalization is always based on several interacting factors. We cannot identify one global function, but rather several minor functions in subsystems. The new functions hold for single inflectional classes or subclasses but not for the system as a whole. Restructuring is not innovative: it always reinforces principles already in the system. If conditioning of declensions is linked to gender distinctions, it strengthens gender. If phonological factors reshape inflectional class conditioning, this can facilitate perception (cf. German and Dutch trochee structures as prosodic clues for morphologically complex forms, or schwa-insertion securing morphological transparency of the stem-suffix border), or it can ease production (cf. the voicing assimilation of the weak dental suffix in Dutch and Swedish). The complexity of the inflectional class system decreases if formerly unpredictable formal variation is exploited for reinforcing already established patterns.

Termining this the function of inflectional classes would be a telic line of reasoning not adequate to language change as an evolutionary phenomenon. From an evolutionary perspective, semantically opaque inflectional classes provide a pool of allomorphic variation, they have no synchronic use, but they can be and are used – ‘exapted’ in Lass’s (1990) terms – for different optimizations in a grammatical system. Rather than a goal in itself, the reorganization of inflectional classes is more adequately regarded as an economic side effect of language change and the pattern-seeking nature of human cognition (Blevins & Blevins 2009: 1). The most clear-cut case of self-contained functionalization seems to be the establishment of semantic conditioning. If the formal variance is restructured to map salient cognitive distinctions such as ‘animate’ vs. ‘inanimate’ nouns or ‘modal’ vs. ‘non-modal’ verbs, this will optimize lexical organization.

References

- Aikhenvald, Alexandra Y. & R. M. W. Dixon. 1998. "Dependencies between Grammatical Systems". *Language* 74.56–80. doi:10.1353/lan.1998.0165
- Albright, Adam & Bruce Hayes. 2003. "Rules vs. Analogy in English Past Tenses: A computational/experimental study". *Cognition* 90.119–161. doi:10.1016/S0010-0277(03)00146-X
- Allan, Robin, Philip Holmes & Tom Lundskaer-Nielsen. 1998. *Danish: A comprehensive grammar*. 2nd edn. London: Routledge.
- Allén, Sture. 1971. *Frequency Dictionary of Present Day Swedish Based on Newspaper Material, Vol. II: Lemmas*. Stockholm: Almqvist & Wiksell.
- ANS =Haeseryn, W., K. Romijn, G. Geerts, J. de Rooij & M. C. van den Toorn. 1997. *Algemene Nederlandse Spraakkunst*. 2nd edn. Groningen: Martinus Nijhoff uitgevers/Wolters Plantyn. <http://ans.ruhosting.nl/>. (Accessed: 21 January 2015.)
- Augst, Gerhard. 1975 "Zum Pluralsystem". *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache* (Forschungsberichte des Instituts für deutsche Sprache, Mannheim, 25), 5–70. Tübingen: Narr.
- Bammesberger, Alfred. 1990. *Die Morphologie des urgermanischen Nomens*. Heidelberg: Winter.
- Beekes, Robert S. P. 2011. *Comparative Indo-European Linguistics: An introduction*. 2nd edn. Revised and corrected by de Vaan Michiel. Amsterdam: John Benjamin. doi:10.1075/z.172
- Birkmann, Thomas. 1987. *Präteritopräsentia. Morphologische Entwicklungen einer Sonderklasse in den altgermanischen Sprachen* (= Linguistische Arbeiten, 188). Tübingen: Niemeyer.
- Blevins, James P. & Juliette Blevins 2009. *Analogy in Grammar: Form and acquisition*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199547548.001.0001
- Booij, Geert. 2002. *The Morphology of Dutch*. Oxford: Oxford University Press.
- Bybee, Joan L. 1985. *Morphology: A study on the relation between meaning and form*. Amsterdam: John Benjamins. doi:10.1075/tsl.9
- Bybee, Joan L. 1994. "Morphological Universals and Language Change". *The Encyclopedia of Language and Linguistics* ed. by R. E. Asher, 2557–2562. Oxford: Pergamon Press.
- Bybee, Joan L. & Carol L. Moder 1983. "Morphological Classes as Natural Categories". *Language* 59.251–270. doi:10.2307/413574
- Corbett, Greville G. 2009. "Canonical Inflectional Classes". *Selected Proceedings of the 6th Décebrettes: Morphology in Bordeaux* ed. by Montermini abio, Gille Boyé & esse Tseng, 1–11. Sommerville, MA: Cascadilla Proceedings Project. <http://www.lingref.com/cpp/decemb/6/paper2231.pdf>. (Accessed: 25 April 2017.)
- Dammel, Antje. 2009. "How – and Why – Do Inflectional Classes Arise? A case study on Swedish and Norwegian conjugation". *Selected Proceedings of the 6th Décebrettes: Morphology in Bordeaux* ed. by Montermini Fabi, Gille Boyé & Jess Tseng, 12–21. Sommerville, MA: Cascadilla Proceedings Project. <http://www.lingref.com/cpp/decemb/6/paper2232.pdf>. (Accessed: 25 April 2017.)
- Dammel, Antje. 2011. *Konjugationsklassenwandel. Prinzipien des Ab-, Um- und Ausbaus verbalflexivischer Allomorphie in germanischen Sprachen* (= Studia Linguistica Germanica, 103). Berlin: Walter de Gruyter.
- Dammel, Antje & Damaris Nübling. 2006. "The Superstable Marker as an Indicator of Categorical Weakness?" *Folia Linguistica* 40:1–2. 97–113. doi:10.1515/flin.40.1part2.97
- Dressler, Wolfgang U., Willi Mayerthaler, Oswald Panagl & Wolfgang U. Wurzel. 1987. *Leitmotifs in Natural Morphology* (= Studies in Language Companion Series, 10). Amsterdam: John Benjamins. doi:10.1075/slcs.10

- Duden-Grammar = *Duden-Grammatik*. 1998. Ed. by the Dudenredaktion. [Duden 4]. Mannheim: Dudenverlag.
- Fortson, Benjamin W. 2010. *Indo-European Language and Culture: An introduction*. 2nd edn. Malden: Wiley-Blackwell.
- Froschauer, Regine. 2003. *Genus im Althochdeutschen: Eine funktionale Analyse des Mehrfachgenus althochdeutscher Substantive* (= Germanistische Bibliothek, 16). Heidelberg: Winter.
- Grinevald, Colette. 2002. "Making Sense of Nominal Classification Systems: Noun classifiers and the grammaticalization variable". *New Reflections on Grammaticalization* ed. by Ilse Wischer & Gabriele Diewald, 259–275. Amsterdam: John Benjamins. doi:10.1075/tsl.49.17gri
- Harnisch, Rüdiger. 1999. "Zum Zusammenspiel von Auslautkonsonanz und Stammvokalismus beim Verb. Südhüringisch/Hochdeutsch kontrastiv". *Variation in der Wortstruktur* ed. by Matthias Butt & Nanna Fuhrhop (= Germanistische Linguistik, 41–42), 61–76. Hildesheim: Olms.
- Jansson, Valter. 1966 [1947]. "Uppkomsten av tredje konjugationen". *Tre uppsatser i språkvetenskap* ed. by Otto von Friesen, Gösta Franzén & Valter Jansson, 68–87. Stockholm: Almqvist & Wiksell.
- Kiparsky, Paul. 2009. "The Old High German Weak Preterite". *On Inflection* ed. by Patrick O. Steinkrüger & Manfred Krifka (= Trends in Linguistics, 184), 107–124. Berlin: Walter de Gruyter.
- Köpcke, Klaus-Michael. 2000. "Chaos und Ordnung – Zur semantischen Remotivierung einer Deklinationsklasse im Übergang vom Mittelhochdeutschen zum Neuhochdeutschen". *Angemessene Strukturen: Systemorganisation in Phonologie, Morphologie und Syntax* ed. by Andreas Bittner, Dagmar Bittner & Klaus-Michael Köpcke, 107–122. Hildesheim: Olms.
- Kürschner, Sebastian. 2008. *Deklinationenklassen-Wandel: Eine diachron-kontrastive Untersuchung zum Wandel der Pluralallomorphie im Deutschen, Niederländischen, Schwedischen und Dänischen* (= Studia Linguistica Germanica, 92). Berlin: Walter de Gruyter.
- Kürschner, Sebastian. 2009. "Morphological Non-Blocking in Dutch Plural Allomorphy: A contrastive approach". *Language Typology and Universals* 62:4.285–306. doi:10.1524/stuf.2009.0022
- Kürschner, Sebastian & Damaris Nübling. 2011. "The Interaction of Gender and Declension in Germanic Languages". *Folia Linguistica* 45:2.355–388. doi:10.1515/flin.2011.014
- Kågerman, Elisabet. 1985. *Obestånd plural hos neutra med avledningsmorfermet -e mellan omkring 1300 och 1750* (= Lundastudier i Nordisk Språkvetenskap, A 36). Malmö: Liber.
- Lass, Roger. 1990. "How to Do Things with Junk: Exaptation in language evolution". *Journal of Linguistics* 26.79–102. doi:10.1017/S0022226700014432
- Mätzig, Simone, Judit Druks, Jackie Masterson & Gabriella Vigliocco. 2009. "Noun and Verb Differences in Picture Naming: Past studies and new evidence". *Cortex* 45:6.738–758. doi:10.1016/j.cortex.2008.10.003
- Mailhammer, Robert. 2007a. "Islands of Resilience: The history of the German strong verbs from a systemic point of view". *Morphology* 17:1.77–1108. doi:10.1007/s11525-007-9113-y
- Mailhammer, Robert. 2007b. *The Germanic Strong Verbs: Foundations and developments of a new system* (= Trends in Linguistics, 183). Berlin: de Gruyter Mouton. doi:10.1515/9783110198782
- Mailhammer, Robert. 2008. "Ablaut Variation in the Proto-Germanic Noun: The long arm of the strong verbs". *Sprachwissenschaft* 33.279–300.
- Mańczak, Witold. 1980. "Laws of Analogy". *Historical Morphology* ed. by Jacek Fisiak, 283–288. The Hague: Mouton. doi:10.1515/9783110823127.283
- McGregor, William B. 2002. *Verb Classification in Australian languages* (= Empirical Approaches to Language Typology, 25). Berlin: Mouton de Gruyter. doi:10.1515/9783110870879

- Meier-Brügger, Michael. 2002. *Indogermanische Sprachwissenschaft*. 8th edn. Berlin: Walter de Gruyter.
- Neef, Martin. 2000a. "Phonologische Konditionierung". *Morphology: An international handbook on inflection and word-formation*, Vol. I (= Handbooks of Linguistics and Communication Science, 17.1) Ed. by Geert Booij, Christian Lehmann & Joachim Mugdan, 463–473. Berlin: de Gruyter Mouton.
- Neef, Martin. 2000b. "Morphologische und syntaktische Konditionierung". *Morphology: An international handbook on inflection and word-formation*, Vol. I (= Handbooks of Linguistics and Communication Science, 17.1) Ed. by Geert Booij, Christian Lehmann & Joachim Mugdan, 473–484. Berlin: de Gruyter Mouton.
- Norde, Muriel. 1997. *The History of the Genitive in Swedish: A case study in degrammaticalization*. Amsterdam: Universiteit van Amsterdam.
- Nübling, Damaris. 2000. *Prinzipien der Irregularisierung* (= Linguistische Arbeiten, 415). Tübingen: Niemeyer. doi:10.1515/9783110915082
- Nübling, Damaris. 2008. "Was tun mit Flexionsklassen? Deklinationsklassen und ihr Wandel im Deutschen und seinen Dialekten". *Zeitschrift für Dialektologie und Linguistik* 75:3.282–330.
- Philippa, Marlies. 1982. "Problematiek rond het *s*-meervoud; een diachroon overzicht". *De Nieuwe Taalgids* 75.407–417.
- Ramat, Paolo. 1981. *Einführung in das Germanische*. Tübingen: Niemeyer. doi:10.1515/9783111346113
- Ringe, Donald A. 2017. *A Linguistic History of English, Vol. I: From Proto-Indo-European to Proto-Germanic*. 2nd edn. Oxford: Oxford University Press.
- Rix, Helmut. 2001. *Lexikon der indogermanischen Verben. Die Wurzeln und ihre Primärstamm-bildungen*. 2nd edn. Wiesbaden: Ludwig Reichert.
- SAG =Teleman, Ulf, Erik Andersson & Staffan Hellberg. 1999. *Svenska Akademiens Grammatik, Vol. II: Ord*. Stockholm: Norstedts Ordbok.
- Schwerdt, Judith. 2001. "Zur Bedeutung des *-nan*-Suffixes der gotischen schwachen Verben". *Beiträge zur Geschichte der deutschen Sprache und Literatur* (PBB) 123:175–210.
- Schwerdt, Judith. 2008. *Morphosemantik der schwachen Verben im Ostgermanischen und Kontinentalwestgermanischen*. Frankfurt a.M.: Lang.
- Szemerényi, Oswald. 1990. *Einführung in die vergleichende Sprachwissenschaft*. 4th edn. Darmstadt: Wissenschaftliche Buchgesellschaft.
- van der Hulst, Harry & Jan G. Kooij. 1998. "Prosodic Choices and the Dutch Nominal Plural". *Phonology and Morphology of the Germanic Languages* ed. by Wolfgang Kehrein & Richard Wiese (= Linguistische Arbeiten, 386), 187–197. Tübingen: Niemeyer. doi:10.1515/9783110919769.187
- Wegener, Heide. 2003. "Normprobleme bei der Pluralbildung fremder und nativer Substantive". *Linguistik online* 16:4. http://www.linguistik-online.de/16_03/wegener.html. (Accessed: 8 June 2014.)
- Wurzel, Wolfgang U. 1986. "Die wiederholte Klassifikation von Substantiven. Zur Entstehung von Deklinationsklassen". *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39:1.76–96.
- Wurzel, Wolfgang U. 2001. *Flexionsmorphologie und Natürlichkeit: Ein Beitrag zur morphologischen Theoriebildung* (= Studia Grammatica, 21). 2nd edn. Berlin: Akademie-Verlag

The history of verb classification in Nyulnyulan languages

William B. McGregor

Aarhus University

Most languages of northern Australia show a compound verb construction in which an uninflecting verbal element collocates with one of a small set of inflecting verbs. McGregor (2002) argues that this construction typically represents a system of verb classification in which the inflecting verbal root categorizes the uninflecting element. He proposes further that such systems of verb classification diffused areally across the northern part of the continent. For the languages of the Kimberley region, there is evidence that this diffusion predated the break-up of the proto-languages, and that systems of verb classification (or their precursors) can be reconstructed for the proto-languages. I argue that such a system can be reconstructed for Proto-Nyulnyulan, and that subsequent historical changes to this system gave rise to the modern systems of the daughter languages.

1. Introduction

1.1 Aims and purposes

Many, perhaps even the majority, of languages of northern Australia show systems of verb classification (McGregor 2002). In most languages the system is formally coded in compound verb constructions (CVCs) which consist of an uninflecting element in collocation with an inflecting verb, which usually follows it. The uninflecting and inflecting items belong to two distinct parts of speech, both of which correspond to verbs of standard average European languages (McGregor 2002). They are henceforth dubbed preverbs (PVs) and inflecting verbs (IVs), respectively. PVs usually form an open class of items which readily admit new members, while IVs typically constitute a closed class, usually with between ten and two hundred members. PVs are generally found in CVCs, in which constructions around a dozen or so IVs habitually occur. Some IVs in the typical language have the potential of

occurring alone, independently of PVs, in simple verb constructions (SVCs). The following Nyulnyul examples illustrate the two lexical types and the two corresponding construction types: (1) shows the IV -J “say” in a SVC, while (2) shows the same IV in a CVC involving the PV *kurd* “hide”.¹

- (1) Nyulnyul
irr-in i-li-rr-j-jan
 3AUG.CRD-ERG 3NOM-IRR-AUG-say-1MIN.OBL
 “They might say to me.” or “They might tell me.”
- (2) Nyulnyul
bin wamb yiil jin kurd i-n-d-in
 this man dog 3MIN.OBL hide 3NOM-CM-say-PRS
 “This man’s dog is hiding.”

As argued in McGregor (2002), in CVCs the PV root is classified by the IV root. Thus in (2) -J “say” assigns the PV to one of about a dozen categories, here a generic category of activities, i.e. dynamic events (McGregor 2002, 2012).

It seems likely that the CVC verb classification system diffused areally in the Kimberley and adjacent regions in northern Australia, most probably prior to the splitting up of the proto-languages. One piece of evidence in favor of this is the fact that it appears that both PVs and IVs can be reconstructed for many of the proto-languages (e.g., for Proto-Nyulnyulan, Stokes & McGregor 2003; for Proto-Worrorran, McGregor & Rumsey 2009).

The main purpose of this paper is to strengthen the argument by showing that it is possible to reconstruct something of the CVC system in Proto-Nyulnyulan. This is contra the negative stance of Bowern (2008), who argues that there are difficulties in reconstructing the verb classification system of Proto-Nyulnyulan because: (a) the classification system differs among the languages; (b) PV-IV pairs are not readily reconstructible for Proto-Nyulnyulan; (c) the combinations are subject to reanalysis; (d) the syntax of the PV-IV collocations differs markedly across Nyulnyulan languages; and (e) the number of cognate PVs is quite small, partly due to the fact that many PVs are recruited from other word classes and are highly borrowable.

Whilst there is a degree of truth to these observations, I argue in this paper that things are not as bad as Bowern (2008) presumes. First, I propose that – given some plausible presumptions – it is possible to reconstruct a number of characteristics of the CVC in Proto-Nyulnyulan. This may or may not have been a classification system; the evidence is inconclusive, though not inconsistent with the possibility

1. Inflecting verbs (IVs) are cited in capital letters, and person categories are indicated by the first three numerals.

that it was, and indeed some evidence makes it not unlikely that this was so. Second, it is demonstrated that the systems of modern Nyulnyulan languages are more similar amongst themselves than they are to systems of other Kimberley languages. Third, I show that there are ways of getting a handle on the PV-IV collocations of Proto-Nyulnyulan. And fourth, there is a good deal of commonality among Nyulnyulan languages as regards the IVs that collocate with any given PV, and unlike the sets of IVs that collocate with corresponding PVs in other Kimberley languages. Moreover, across Nyulnyulan languages the classification systems themselves share a good deal in common: the semantic bases of the systems are remarkably similar, and there is a good degree of similarity amongst the PVs that collocate with particular IVs, and far less similarity outside of Nyulnyulan.

The paper is organized as follows. I begin in the next section by outlining some necessary background information on the Nyulnyulan languages. This is followed up in §2 by an overview of the general characteristics of the verbal systems in Nyulnyulan languages. This section also comments on likely features of the verbal construction of Proto-Nyulnyulan. Following this, in §3 we compare the sets of IVs that collocate with about a score of frequent PVs that are reflexes of Proto-Nyulnyulan PVs; these sets are then compared and contrasted with the sets of IVs collocating with corresponding PVs in other languages of the region. It is shown that the sets of collocating IVs in Nyulnyulan languages are quite similar, but very different from the collocating IVs in the other languages. Section 4 then attempts – building on an approach developed in McGregor (2002) – a system-level comparison of verb classification in two fairly distantly related Nyulnyulan languages, Nyulnyul and Warrwa. This reveals that the systems in the two languages are closer than either is to the system in unrelated Gooniyandi. The origins and evolution of verb classification in Nyulnyulan are overviewed in §5, and §6 winds up the paper with some conclusions and suggestions for future research.

1.2 Nyulnyulan family

The Nyulnyulan family is a small group of about ten genetically related language varieties spoken on Dampier Land and nearby parts of the Kimberley mainland (Stokes & McGregor 2003). No relation has been demonstrated to any other languages of the continent (Bowern 2004a; Stokes & McGregor 2003), and thus the languages are non-Pama-Nyungan. There are two primary subgroups of the family, Western Nyulnyulan (Bardi, Jawi, Nyulnyul, Jabirrabirri, Nimanburru, Ngumbarl) and Eastern Nyulnyulan (Warrwa, Nyikina, Yawuru, Jukun) (Stokes & McGregor 2003).

The languages are fairly closely related genetically and typologically. All are either highly endangered or extinct. The strongest, Bardi, has no more than about

fifty speakers, and is not being passed on to children as their main language of everyday communication (Bowern 2004b).

The main sources of information on the languages are: my own audio recordings and fieldnotes for Nyulnyul and Warrwa; Stokes (1982) for Nyikina; Hosokawa (1991) for Yawuru; Metcalfe (1975); Aklif (1991a, 1991b, 1999); Bowern (2004b) for Bardi; and Nekes & Worms (1953, 2006) for Bardi, Jabirrabirri, Nimanburru, Yawuru, and Nyikina. Daisy Bates' collection held in the National Library of Australia (MS 365 *Papers of Daisy Bates*) holds information on a number of Nyulnyulan languages, though the bulk of it is lexical.

The limitations of these sources are manifest. With the possible exception of Bardi, none of the languages can be said to be well documented, and – with the exception of the two languages I have worked on myself – what is available to me for the present analysis is secondary rather than primary data. No comprehensive electronic documentation exists for any Nyulnyulan language.

2. Features of verbal lexemes and constructions in Nyulnyulan languages

Nyulnyulan languages are fairly typical of northern Australian languages in terms of their verbal elements and constructions. They possess between about one hundred and two hundred and fifty or so IVs for the Eastern and Western Nyulnyulan languages respectively. These form (fairly) closed classes that do not readily admit borrowings. Each language has an open class of PVs that numbers several hundred items; this class readily admits borrowings.

PVs usually occur in collocation with IVs in CVCs. Exceptions exist in some languages (e.g. Nyulnyul) where the PV may occur independently of an IV, for instance in a non-finite clause (McGregor 2012). Regardless of the size of the IV class, about a dozen IVs in each language are productive in the sense that they occur in collocation with a fair number of different PVs, and with some degree of frequency in CVCs. In addition, a further ten or so IVs occur irregularly in a handful of distinct CVCs.

The productive IVs are mostly basic ones, with a high frequency of use. Thus, for instance, they account for about 50% of IV tokens in Bardi (my statistics on a small corpus), for over 60% of IV tokens in my Nyulnyul corpus, and for around 75% of the tokens in Nyikina and Warrwa. (These statistics include all instances of the IVs, regardless of the type of the verbal construction.)

The most productive IVs in CVCs are recurrent across Nyulnyulan languages, as revealed by Table 1, which provides data on five of the best studied languages, three Eastern and two Western Nyulnyulan languages. It is significant that most

Table 1. Most common inflecting verbs in four Nyulnyulan languages

Nyikina		Warrwa		Yawuru		Bardi		Nyulnyul	
-I “do, say”	56%	-JI “do, say”	30%	-JU “say”	33%	-JOO “say, do”	31%	-JU “do, say”	33%
-NI “sit”	3%	-NI ~ -NGA “be”	12%	-NI “be”	20%	-NI “sit, be”	2%	-N “be”	3%
-MA “make, put”	9%	-MA “put”	11%	-MA “put”	13%	-MA “put”	20%	-M “put”	13%
-		-∅ ~ -WA “give”	8%	-		-∅ ~ -WA “give”	9%	-W “give”	11%
-MA “go”	7%	-ARNDA “go”	8%	-RNDIRA “go”	6%	-JIIDI “go”	4%	-JID “go”	4%
-NGARA “become”	4%	-NGARA “become”	6%	-NGARA “become”	11%	-		-	
-BANJI “share”	3%	-BANYJI “exchange”	6%	-		-BANJI “exchange”	1%	-BANY “exchange”	4%
-ANDI “pick up”	4%	-ANDI “get, fetch”	5%	-NYA “catch”	4%	-(I)NYA “pick up”	10%	-NYU “get, catch”	13%
-A “carry”	9%	-KA “carry”	3%	-KA “carry”	4%	-KA “carry”	4%	-K “carry”	6%
-		-NKA “hit”	2%	-		-BI “hit”	1%	-	
-		-		-RA “spear”	3%	-BOO “poke”	2%	-R “poke”	0.5%
-		-		-		-KAL “wander”	3%	-KAL “wander”	2%

Note. The percentages in the columns do not add up to 100% because only the most common inflecting verbs are listed.

IVs regularly occurring in CVCs are evidently cognate. Thus the “say”, “sit”, “put”, “exchange” and “carry” IVs show clearly related forms in all or most languages, and can be reconstructed to Proto-Nyulnyulan (Stokes & McGregor 2003):

*-JU “say”

*-NI “sit, be”

*-MA “put”

*-BANJI “exchange”

*-KA “carry”

In addition, *-WA “give” and *-NYA “get” can also probably be reconstructed in Proto-Nyulnyulan (Stokes & McGregor 2003). *-NGARA “become” and *-ARND “go” are reconstructible in Proto-Eastern Nyulnyulan, and *-KAL “wander” and *-JID “go” in Proto-Western Nyulnyulan.

It is striking not just that seven IVs can be reconstructed with some degree of certainty in Proto-Nyulnyulan, but also that there is a high level of agreement among the languages in terms of the proportion of the attested PV class that collocate with the modern reflexes of the proto-IVs.² Thus the reflexes of *-JU “say” occur in a third of all CVCs in all languages except Nyikina, where the proportion goes up to over a half. Reflexes of *-MA “put” account for around 10% of the CVC types, while reflexes of *-BANJI “exchange” and *-KA “carry” fairly consistently account for about half of this fraction. Reflexes of *-WA “give” account for about 10% of CVC types in all languages in which this IV occurs in CVCs. The “go” and “get” IVs are also relatively consistent across the languages, despite the fact that they are not cognate in all modern languages. The main discrepancies are with reflexes of *-NI “sit, be”, which varies considerably in productivity from 2% (Bardi) to 20% (Yawuru), and that the “become” IV is restricted to Eastern Nyulnyulan, the “wander” IV to Western Nyulnyulan.

It seems reasonable to conclude from the above observations that the similarities in frequency of these productive IVs are not accidental, the result of independent innovations in the daughter languages, but have historical explanations. The two main possibilities are that the CVC diffused across Nyulnyulan languages after they separated from Proto-Nyulnyulan, or that it is inherited from Proto-Nyulnyulan.

There is evidence against the diffusion scenario. Diffusion could have happened in two ways: (a) the actual PV-IV pairings may have diffused; or (b) the ‘idea’ of the CVC, that is, the abstract construction itself may have been borrowed across a large

2. I give proportions rather than raw numbers because of the differences in sizes of attested PV lexicons in the five languages.

area. Against (a) is the fact that the “go” and “get” IVs are highly comparable across Nyulnyulan languages in terms of their cut of the CVC pie, despite the fact that the forms are not all cognate. Of course, this could be explained by replacement of a previous shared IV form by a group-specific form. But if this was so, the shared form and its collocational patterns would be reconstructible to Proto-Nyulnyulan, contradicting the diffusion hypothesis.

Scenario (b) is widely presumed in the Australianist literature as the diffusion mechanism: the CVC spread over a wide area, rather than particular pairings of PV and IV (e.g. McGregor 2002). There seems to be, however, much more similarity among the Nyulnyulan CVCs than can be reasonably accounted for by areal borrowing at the level of the construction. In support of this claim, consider the comparison presented in Table 2 of the relative break-up of the CVC pie in Nyulnyul and Gooniyandi,³ an unrelated but not too geographically remote language. In all likelihood the CVC was diffused during some period when the proto-languages of these two languages were spoken. It is evident from the table that there is rather little in common between the languages on the dimension at issue here. To begin with, there is much less agreement among the two languages as to the IVs that are typically used in CVCs than among the Nyulnyulan languages. Furthermore, there are only three likely cognate IVs – or, rather, look-alikes – and for two of them, the “carry” and “be” IVs, the percentages are quite different. It is only for the third, -M “put” (Nyulnyul) and -MA “say” (Gooniyandi) that the percentages are in agreement. But as it turns out, the categories marked by these two IVs do not correspond, but rather the correspondence is between Nyulnyul -J “say” and Gooniyandi -MA “say”, both of which are generic activity categories. The only place where there is significant agreement in the sizes of corresponding categories is in the case of the reflexive/reciprocals, marked by -BARNJ “exchange” (Nyulnyul) and -ARNI₂ “carry self” and -MARNI “say to self” (Gooniyandi). In any event, this agreement is likely not the result of diffusion, but probably reflects the general frequency of reflexive/reciprocals within the wider lexicon. What this comparison indicates is that where CVC systems have been borrowed little agreement in the type frequency of the IVs can be expected, even if the IVs are (putatively) cognate. Scenario (b) thus seems unlikely.

3. Gooniyandi does not have a CVC, but a single verbal construction involving a verb root and an inflecting classifying morpheme. This construction is evidently a fairly recent grammaticalization of a previous CVC, in which the IV ultimately became a verbal classifier. I present the Gooniyandi data here because it is readily available – in most cases frequency counts have not been done.

Table 2. Comparison of relative frequencies of IVs across CVC tokens in Nyulnyul and Gooniyandi (IVs expressing comparable meanings are indicated in the same row; grey background indicates that no such IV exists)

Nyulnyul		Gooniyandi	
-J “say”	33%	-MI “say”	13%
-NY “get, catch”	13%		
-M “put”	13%		
-W “give”	11%	-DI “give, put”	9%
-K “carry”	6%	-A “carry”	27%
-JID “go”	4%		
-BARNJ “exchange”	4%	-ARNI ₂ “carry self”	3%
		-MARNI “say to self”	2%
-N “be”	3%	-I “be”	17%
-KAL “wander”	2%		
-R “poke”	0.5%		
		-ANI “fall”	9%
		-BINI “hit”	7%
		-BINDI “become”	6%
		-ARRI “throw”	5%
		-ARNI ₁ “arrive”	2%
		-BIRLI “consume”	0.4%

In all languages each PV usually occurs with between one and five IVs; the number of distinct PV-IV combinations of each of these types reduces rapidly. Thus in Nyulnyul some 78% of PVs collocate with a single IV; 14% with two IVs; 5% with three; 2% with four; and just 1% with five IVs.

Where more than one IV collocates with a PV, the different collocations contrast in meaning as shown by the following collocations in Nyulnyul:

- (3) *baab* “open”
 -JID “go” “come open”
 -NY “get” “open something”
 -M “put” “cause something to open”
daarr “arrive”
 -R “poke” “arrive”
 -BARNJ “exchange” “meet together”
 -J “say” “arrive, arrive with”
 -*ma-R-anyj* “poke one another” “meet together”⁴
 -M “put” “(let) arrive”

4. The use of a reflexive/reciprocal form of an IV in a CVC is not attested anywhere in my own data, and can only be found in Nekes & Worms (1953).

<i>junk</i> “run”	
-JID “go”	“be running along”
-J “say”	“be running along”
-NY “get”	“run away, start running”
-M “put”	“cause or let something run”
<i>jarrbard</i> “lift up”	
-K “carry”	“lift up and carry”
-BARNJ “exchange”	“lift oneself up”
-NY “get”	“pick up”
-M “put”	“pick up”
<i>mijal</i> “sit”	
-N “be”	“be sitting”
-LAND “sit down”	“sit down”
-NY “get”	“start sitting down”
-M “put”	“put into sitting position”

Each pairing must be regarded as a distinct lexical entry, with a meaning that is not entirely predictable from the meanings of the components. However, there is a considerable amount of regularity in the meaning contrasts whereby general meanings can be associated with each IV across all of its uses in CVCs. For instance, in all uses of -J “say” the meaning component ‘dynamic event’ is present. This represents the semantic basis of the classificatory function of IVs in CVCs, whereby they categorize the PV and the event it denotes (see further §4 below).

Summing up, the evidence suggests that Proto-Nyulnyulan possessed a CVC which has reflexes in the modern languages. Quite likely at least eight IVs in Proto-Nyulnyulan were employed in CVCs: *-JU “say”, *-NI “sit, be”, *-MA “put”, *-BANJI “exchange”, *-KA “carry”, *-WA “give”, *-NYA “get”, and the proto-IV “go” (whatever its form might have been). With a few exceptions, these IVs have tended to retain their relative frequencies of occurrence in CVC types. In Nyikina the “say” verb became more promiscuous, ultimately accounting for half of the PV collocations; in Yawuru and Warrwa the “sit, be” IV expanded in its collocational potential. The “give” verb went out of use in CVCs in Nyikina and Yawuru. Possible innovations in Proto-Eastern Nyulnyulan and Proto-Western Nyulnyulan are, respectively, of the IVs *-NGARA “become” and *-KAL “wander”; the use of these IVs in CVCs was of necessity also an innovation. Nothing can be said with any degree of certainty about the “hit” and “poke” IVs.

3. Broad comparison of PV-IV collocations in the Nyulnyulan family

3.1 Reflexes of some proto-PVs and their collocating IVs in Nyulnyulan languages

The discussion of §2 shows that the relative bites of the CVC cherry by IVs in Nyulnyulan languages are quite comparable, presumably reflecting the situation in Proto-Nyulnyulan. The evidence in favor of the latter inference is based on a comparison of two languages between which borrowing of CVCs evidently must have occurred at some point in the distant past (before the languages differentiated). This tells us precious little about the properties of the proto-CVC beyond the fact that it would have been constituted of a pairing of PVs with IVs. To begin getting a handle on the nature of the beast, I now deploy an approach developed in McGregor & Rumsey (2009) in the context of arguing for the genetic unity of the Worroran languages of the northern Kimberley. This approach sets up pairings of a number of relatively basic PVs with IVs across a cluster of related languages, and examines the level of agreement in the pairings across the languages. This evidence can be further employed to reconstruct collocations in the proto-language.

Table 3 shows the attested collocating IVs in four Nyulnyulan languages (two from each of the two major branches) for lexemes – mainly PVs – that correspond to twenty seven PV-like lexical items that have been reconstructed for Proto-Nyulnyulan in Stokes & McGregor (2003). Included in the tabulation are not just reflexes of the proto-forms, but also forms innovated in some languages.

There are gaps in the data, where incomplete information is available on collocating IVs. Indeed, it is likely that a number of additional collocations would have been attested given larger corpora (McGregor 2002). Moreover, it must be cautioned that the data is not entirely reliable: it has been compiled from various sources, including modern descriptions and wordlists, as well as earlier sources including Nekes & Worms (1953), which latter work provides a number of PV and CVC attestations not otherwise documented,⁵ and it is quite possible that these forms and collocations were not traditional, but were engendered by their

5. This applies especially to Yawuru, where Hosokawa's data differs somewhat from that recorded earlier by Nekes and Worms in the 1930s and 1940s, and Stokes in the 1970s. Hosokawa puts the difference down to the fact that other investigators worked with individuals who showed a considerable influence from other Nyulnyulan languages, especially Nyikina (Hosokawa 1991).

Table 3. Examples of PV-IV collocations in four Nyulnyulan languages involving reflexes of Proto-Nyulnyulan PVs

(The tilde ~ replaces another instance of the previous PV; IVs given in square brackets express the given meaning in SVCs; where no IV is indicated in collocation with a given PV this means that the sources do not provide indication of the possible IV collocates; glosses given in brackets are meanings specific to the CVC – these are only given when highly irregular.)

Gloss	Proto-NN	Warrwa	Yawuru	Nyulnyul	Bardi
1 “be alive”	*nunyi	<i>nunjaya</i> + WANI “be”	<i>nunja</i> + NI “be”	<i>ninyj</i> + N “be”	
		<i>nunja</i> + NGARA “become”		~ + KAL “wander” ~ + J “say” ~ + JARRJARR “stand”	<i>ninyj</i> + KAL “wander”
2 “angry”	*bili	<i>bili</i> + MA “put”	<i>bili</i> + MA “put”		
		~ + NGARA “become”	~ + NGARA “become”	~ - <i>gaja</i> + JU “say”	<i>bili</i> + JOO “say”
		~ + NI “be”	~ + NGARI “leave”	<i>bil</i> + J “say”	
3 “blow”	*buu	<i>buu</i> + MA “put”	<i>buu</i> + MA “put”	<i>buu</i> + M “put”	<i>boo</i> + MA “put” (“blow away”)
		~ - <i>kay</i> + WANI “be”	~ + JU “say”	~ + J “say”	
4 “call out”	*kawu	<i>kawu</i> + JI “say”	<i>kaw</i> + JU “say”	<i>kaw</i> + J “say”	<i>kaw</i> + JOO “say”
		~ + BANJI “exchange”	~ + MA “put”	~ + M “put”	[-LIRMI]
5 “carry on shoulder”	*kundi	<i>kurndu</i> + KA “carry”	<i>kundu</i> + KA “carry”	<i>kurnd</i> + K “carry”	<i>kurnd</i> + K “carry”
		~ + JI “say”		~ + N “be” ~ + W “give” ~ + NY “get”	~ + N “be” [-MOONG-GARA]
6 “choke, strangle”	*kiny		<i>kiny</i> + NY “get”	<i>kiny</i> + NY “get”	<i>kiny</i> + NY “get”
			~ + R “poke”	~ + R “poke”	~ + R “kill lice”
		<i>kiny</i> + WA “give”		~ + W “give”	
		~ + JU “say”	<i>kinykiny</i> + J “say”		[-MINGGI]
		<i>kinykiny</i> + NYA “get”	~ + BARNJ “exchange”		

(continued)

Table 3. (continued)

Gloss	Proto-NN	Warrwa	Yawuru	Nyulnyul	Bardi
7 “climb”	*lakal	<i>lagarr</i> + YI “say”	<i>lakarr</i> + JU “say”	<i>lakal</i> + J “say”	
			~ + NYA “get”	~ + NY “get”	~ + NY “get”
				~ + JID “go”	~ + JIID “go”
				~ + WANYJ “climb”	<i>lakal</i> + GANYI “climb”
		[–WANYJI]	[–WANYJ]	[–GANYJI]	
8 “crawl”	*yadab	<i>yadab</i> + JI “say”	<i>yadap</i> + JU “say”	<i>yadab</i> + J “say”	<i>yadab</i> + JOO “say”
		~ + ARNDI “go”		~ + KAL “wander”	
9 “cry, weep”	*wangkurr	<i>wangkurr</i> + JI “say”	<i>wangkurr</i> + JU “say”		
		[–NGARLA]	~ - <i>gaja</i> + NI “be”	<i>wangirr</i> + [–NGALK]	<i>anggoorr</i> + MOOROO “spill” ~ + MA “put” (“mourn”) [–NGALGA]
10 “cut”	*jub	<i>jub</i> + ANDI “get”		<i>jub</i> + NY “get”	<i>yaarl</i> + NYA “get”
		~ + BANJI “exchange”		~ + BARNJ “exchange”	
		~ + WALI “cut”	<i>jad</i> + JU “say”	~ + M “put” ~ + W “give”	
11 “dance”	*burrb	<i>burrb</i> -JI “say”	<i>burrb</i> + JU “say”	<i>burrb</i> + J “say”	<i>boorboor</i> + JOO “say”
		~ - <i>kay</i> + WANI “be”			
12 “descend, sink”	*yuwurr	<i>yuwurr</i> + JI “say”	<i>jurjrurr</i> + JU “say”	<i>yurr</i> + J “say”	
				~ + M “put”	<i>yoorr</i> + MA “put”
13 “fly”	*dumbarra	<i>dulmarra</i> + JI “say”	<i>dumbarr</i> + JU “say”	<i>dumbar</i> + J “say”	
				~ + NY “get”	<i>dumbar</i> + NY “get”
					[–BIRRILI]
14 “hammer”	*da			<i>daa</i> + DAM “hit”	<i>da</i> + BI “hit”
			<i>dujul</i> + JU “say”	<i>dujuldujul</i> + J “say”	
				~ + W “give”	~ + R “kill lice”

(continued)

Table 3. (continued)

Gloss	Proto-NN	Warrwa	Yawuru	Nyulnyul	Bardi
15 “lick”	*yaly	<i>yaly</i> + JI “say” ~ + BANJI “exchange”	<i>yaly</i> + JU “say”	<i>yaly</i> + [-JILIK]	[-JILI]
16 “lift up and carry”	*jarrbard	<i>jarrbard</i> + JI “say” ~ + KA “carry” ~ + WA “give”	<i>jarrbard</i> + JU “say”	<i>jarrbad</i> + J “say” ~ + K “carry” ~ + M “put” ~ + NY “get” ~ + BARNJ “exchange”	<i>jarrbard</i> + KA “carry”
17 “like, want”	*liyan (“heart, emotion”)	<i>liyan-ngany</i> + MA “put”	<i>liyan</i> + MA “put” ~ + JU “say” ~ + NYA “get” ~ + KA “carry” ~ + RNDIRA “go” ~ + NGARA “become” ~ + NI “be” ~ + JALKU “fall”	<i>liyan</i> + M “put” ~ + J “say” ~ + NY “get” ~ + K “carry” ~ + BARNJ “exchange”	<i>liyan</i> + MA “put” ~ + JOO “say” ~ + NYA “get” ~ + KA “carry” ~ + BARNJ “exchange” ~ + NGULU “throw”
18 “pour out”	*jur(r)urr	<i>jurrurru</i> + JU “say” ~ + RA “poke” ~ + NA “be” <i>jili</i> + MA “put”	~ + NGULA “throw”	<i>jurur</i> + J “say” ~ + R “poke” (“flow”)	<i>joorr</i> + J “say” ~ + R “kill lice” (“flow”) <i>jozor</i> + N “be”
19 “run”	*jurnk	<i>kudii</i> + JI “say” ~ + ANDI “get” ~ + ARNDA “go”	<i>junku</i> + JU “say” ~ + NYA “get” ~ + NGARA “become”	<i>junk</i> + J “say” ~ + NY “get” ~ + M “put”	<i>joornk</i> + NYA “get” ~ + JARRALA “run”
20 “sing”	*jirrmu	[-WU]	<i>dyirrmu</i> + JU “say” ~ - <i>kaja</i> + NI “be”	<i>jirrm</i> + J “say” ~ - <i>kaj</i> + N “be”	<i>jirrm</i> + JOO “say” ~ + N “be”

(continued)

Table 3. (continued)

Gloss	Proto-NN	Warrwa	Yawuru	Nyulnyul	Bardi	
21 “be sitting”	*mijala	<i>mijala</i> + WANI “be”	<i>mijala</i> + NI “be”	<i>mijal</i> + N “be”	<i>miyala</i> + NI “be” (“be awake”)	
			~ + MA “put”	~ + M “put”		
			~ + JU “say”	~ + NY “get”		
			~ + NGARA “become”	~ + LAND “sit down”		
			~ + MINYJ “sit”		[-NI]	
			~ - <i>kaja</i> + NI “be”			
22 “split”	*ja(r)l	<i>jarl</i> + JI “say”	<i>jaarl</i> + JU “say”	<i>jarljarl</i> + J “say”	<i>gagal</i> + JIID “go”	
				<i>jarl</i> + JID “go”		
				<i>jarljarl</i> + JID “go”		
				~ + NY “get”		
23 “be standing”	*yalku	<i>yaalu</i> + WANI “be”	<i>yalku</i> + NI “be”	<i>yalk</i> + N “be”	<i>jirrjirr</i> + JOO “do” ~ + (I)NYA “get”	
			~ + JARRA “stand up”	~ + MIDJALA “arise”		~ + JARRNGAR “stand”
			~ + BULA “emerge”			
			~ + MA “put”			
		[+ JARRA]				
24 “steal, abduct”	*lanyb	<i>lanybu</i> + ANDI “get”		<i>lanyb-uk</i> + KAL “wander”	<i>laanyb</i> + NYA “get”	
			<i>malkin</i> +	[-LANYB]		
25 “suck”	*juny		<i>juny</i> + J “say”	<i>juny</i> + J “say”	<i>joony</i> + J “say”	
			<i>junyjuny</i> + JU “say”			
				~ + NY “get”		~ + NYA “get”
				~ + W “give”		
26 “turn”	*dibirr	<i>birrb</i> + JI “say” (“turn off”)	<i>birrb</i> + JU “say”	<i>dibirr</i> + J “say”	<i>dibirr</i> + JOO “say”	
			~ + BANJI “exchange”	<i>dibirr</i> + BANJI “exchange”		<i>dirray</i> + BANJI “exchange”
			<i>duwirr</i> + WARRI “throw”	~ - <i>gaja</i> + NI “be”		~ + M “put”
27 “twitch, twinkle”	*bilbil		<i>burdburd</i> + JU “say”	<i>birlbirl</i> + J “say”	<i>balbal</i> + JOO “say”	
				<i>birlbirl-kaja</i> + NI “be”		~ - <i>kaj</i> + N “be”

fieldwork methods.⁶ Be this as it may, it is also possible that the forms given in Nekes & Worms (1953) were acceptable in the 1930s and 1940s, but had gone out of use by the late twentieth century as the languages became increasingly moribund. There is no way we can be certain which circumstance obtained.

Table 3 reveals a good deal of agreement among the four languages. First, in almost every instance in each language there exists a PV (or similar lexeme) that can be employed in a CVC expressing the nominated meaning. In most cases the PVs are formally very similar, and indeed can be traced back to a Proto-Nyulnyulan PV. There are just two exceptions (indicated by boxed-in cells in the table), where in a particular language only an SVC expression is attested: for (15) “lick” in Bardi, and for (20) “sing” in Warrwa.⁷

Second, there is clearly a good agreement among the four languages in terms of the collocating IVs for each given PV. This is apparent from the number of greyed-in rows indicating shared IV collocates. To be sure, there are a number of pairings that are unique to particular languages. But these are outnumbered by pairings that are shared with at least one other language by a factor of greater than two to one overall – 157 pairings are shared out of a total of 218 pairings, while just 61 are unique. For Warrwa 64% of the pairings are shared; for Yawuru and Nyulnyul 72% are shared; and for Bardi 80% are shared. Perhaps even more striking is the fact that for each language every PV – with just one or two exceptions – shares at least one collocating IV with one or more other languages.

If more Nyulnyulan languages are brought into the picture the similarities become even more evident. Table 4 presents the collocating IVs for the reflexes of the twenty seven PV that are attested in at least two Nyulnyulan languages. Included in this table is data from three additional languages: Nyikina (Eastern Nyulnyulan), Jabirjabirr, and Nimanburru (Western Nyulnyulan).⁸

6. Nekes and Worms habitually employed a ‘round table’ setup in the field, in which speakers of a number of languages were seated together and asked for translations of prompts (Nekes & Worms 2006). This doubtless engendered a considerable amount of lexical borrowing and calquing in the speech of the interviewees. On the other hand, the collocations found in Nekes & Worms (1953) do not differ completely from the modern sources; there is in fact a good deal of agreement.

7. Certainly SVC expression sometimes predominates in a language, as in the case of (15) “lick” in Nyulnyul, which is almost invariably expressed by the IV -JILIK “lick”. Indeed, although the PV *yaly* “lick” is attested in the language, it is not attested in an actual CVC.

8. Nekes & Worms (1953) is the only source of information for Jabirjabirr and Nimanburru.

Table 4. Recurrent collocations of IVs with PVs attested in Nyulnyulan languages

	Gloss	Collocating IVs	Languages attested in
1	“be alive”	-NI “sit” -KAL “wander”	Warrwa, Nyikina, Yawuru, Nyulnyul Nyulnyul, Nimanburru, Bardi
2	“angry”	-JU “say” -NGARA “become”	Yawuru, Nyulnyul, Jabirrabirri, Bardi Warrwa, Yawuru
3	“blow”	-MA “put” -JU “say” -MA “put”	Warrwa, Yawuru Yawuru, Nyulnyul Warrwa, Yawuru, Nyulnyul, Jabirrabirri, Nimanburru, Bardi
4	“call out”	-JU “say” -MA “put”	Warrwa, Yawuru, Nyulnyul, Jabirrabirri, Bardi Nyikina, Yawuru, Nyulnyul
5	“carry on shoulder”	-KA “carry”	Warrwa, Nyikina, Yawuru, Nyulnyul, Jabirrabirri, Nimanburru, Bardi
6	“choke, strangle”	“get” -NI “be” -RA “poke” “get” -WA “give”	Nyulnyul, Jabirrabirri, Bardi Nyulnyul, Jabirrabirri, Nimanburru, Bardi Yawuru, Nyulnyul, Jabirrabirri, Bardi Yawuru, Nyulnyul, Jabirrabirri, Bardi Warrwa, Nyulnyul, Jabirrabirri
7	“climb”	-JU “say” -WANYJI “climb” “get” “go”	Warrwa, Nyikina, Nyulnyul Nyikina, Nyulnyul, Jabirrabirri, Nimanburru, Bardi Yawuru, Nyulnyul, Nimanburru, Bardi Nyikina, Nyulnyul, Jabirrabirri, Nimanburru, Bardi
8	“crawl”	-JU “say”	Warrwa, Yawuru, Nyulnyul, Jabirrabirri, Bardi
9	“cry, weep”	-JU “say”	Warrwa, Nyikina, Yawuru
10	“cut”	“get” -BANJI “exchange”	Warrwa, Nyulnyul, Bardi Warrwa, Nyikina, Nyulnyul
11	“dance”	-JU “say”	Warrwa, Nyikina, Yawuru, Nyulnyul, Bardi
12	“descend, sink”	-JU “say” -MA “put”	Warrwa, Yawuru, Nyulnyul Nyikina, Nyulnyul, Jabirrabirri, Bardi
13	“fly”	-JU “say” “get”	Warrwa, Yawuru, Nyulnyul Nyikina, Nyulnyul, Jabirrabirri, Nimanburru, Bardi
14	“hammer”	-JU “say” -RA “poke” “hit”	Yawuru, Nyulnyul Nyikina, Jabirrabirri, Bardi Nyulnyul, Jabirrabirri, Bardi
15	“lick”	-JU “say”	Warrwa, Yawuru, Jabirrabirri
16	“lift up and carry”	-JU “say” -KA “carry” “get”	Warrwa, Yawuru, Nyulnyul Warrwa, Nyikina, Nyulnyul, Jabirrabirri, Nimanburru, Bardi Nyulnyul, Jabirrabirri

	Gloss	Collocating IVs	Languages attested in
17	“like, want”	-MA “put” “get” -KA “carry” -JU “say” -BANJI “exchange”	Warrwa, Nyikina, Yawuru, Nyulnyul, Jabirrabirri, Bardi Yawuru, Nyulnyul, Bardi Yawuru, Nyulnyul, Bardi Yawuru, Nyulnyul, Jabirrabirri, Bardi Nyulnyul, Jabirrabirri, Bardi
18	“pour out”	-RA “poke” -JU “say”	Yawuru, Nyulnyul, Nimanburru, Bardi Yawuru, Nyulnyul, Nimanburru, Bardi
19	“run”	-JU “say” “get” “go” -NGARA “become”	Warrwa, Nyikina, Yawuru, Nyulnyul Warrwa, Nyikina, Yawuru, Nyulnyul, Jabirrabirri, Nimanburru, Bardi Warrwa, Nyikina, Yawuru
20	“sing”	-JU “say”	Yawuru, Nyulnyul, Jabirrabirri, Bardi
21	“be sitting”	-NI “sit” -MA “put”	Warrwa, Nyikina, Yawuru, Nyulnyul, Jabirrabirri, Nimanburru, Bardi Nyikina, Yawuru, Nyulnyul, Jabirrabirri
22	“split”	-JU “say” -MA “put” -JID “go”	Warrwa, Yawuru, Nyulnyul Nyikina, Yawuru Nyulnyul, Jabirrabirri, Bardi
23	“be standing”	-NI “sit” “stand (up)”	Warrwa, Nyikina, Yawuru, Nyulnyul Warrwa, Yawuru, Nyulnyul, Jabirrabirri
24	“steal, abduct”	“get”	Warrwa, Nyikina, Bardi
25	“suck”	-JU “say” “get”	Nyikina, Yawuru, Nyulnyul, Jabirrabirri, Nimanburru, Bardi Nyulnyul, Jabirrabirri, Nimanburru, Bardi
26	“turn”	-JU “say” -BANJI “exchange” -MA “put”	Warrwa, Yawuru, Nyulnyul, Jabirrabirri, Bardi Warrwa, Nyikina, Yawuru, Jabirrabirri, Bardi Nyikina, Nyulnyul, Jabirrabirri
27	“twitch, twinkle”	-JU “say”	Yawuru, Nyulnyul, Bardi

Fully sixty two IV pairings are attested for the reflexes of the twenty seven PVs that are shared by at least two languages; fifty six pairings (91%) are shared by at least three languages. In fact, fourteen collocations are shared by at least five languages. Perhaps even more strikingly, reflexes of a dozen proto-PVs show agreement among at least five languages in terms of one or more collocating IVs. Given that the data for at least two languages is quite poor, this seems highly significant.

3.2 Reconstruction of PV-IV pairings in Proto-Nyulnyulan

The level of agreement amongst Nyulnyulan languages cannot be put down entirely to semantic similarities or associations between the PVs and collocating IVs. True, in a number of instances pairings are not unexpected.⁹ For instance, for each of (5) “carry on shoulders”, (7) “climb”, and (21) “sit”, one collocating IV is virtually identical in meaning with the PV: the IV is effectively a repetition of the PV (though by a different, non-cognate, lexical form). But this repetition is far less frequent than it might have been: many IVs that can express the given meaning in SVCs cannot occur with a semantically identical PV in a CVC. For instance, the Nyulnyul IV -LANYB “steal” does not collocate with the PV *lanyb-uk* “steal”; nor does the “cry” IV collocate with the PV *wangkurr* “cry” in any known language. And even though (7) “climb” shows a repeater pairing, this is found in just the two Western Nyulnyulan languages Nyulnyul and Bardi; it is not attested in Warrwa or Yawuru, even though the cognate IV is attested in at least Warrwa.

More significant for our purposes are unexpected collocations. The most unusual of these are doubtless those involving the Bardi verb -R “kill lice (by use of a pair of delousing sticks)”. Most of the collocating IVs express quite basic meanings; this IV is perhaps the most semantically specialized, even more so than IVs like -KAL “wander” and -BANJI “exchange”. However, -R “kill lice” in Bardi is cognate with -RA ~ -R “poke, spear” of the three other languages, and a reflex of Proto-Nyulnyulan *-RA “poke, pierce, spear” (Stokes & McGregor 2003). The Bardi IV has recently undergone semantic specialization (the connection with poking is that lice were killed by squashing them between a pair of delousing sticks). Three of the four Bardi collocations involving -R “kill lice” correspond with collocations in other languages involving the IV -R(A) “poke, spear”, namely with reflexes of (6) *kiny “choke, strangle”, (14) *da* “hammer”, and (18) *jur(r)urr “pour out”. Even the collocations of these three PVs with the “poke” IV do not appear particularly strongly motivated semantically. Given that the PV-IV collocations are shared by Yawuru or Nyikina and Nyulnyul or Jabirrabirri, it seems reasonable to suggest that they reflect the situation in Proto-Nyulnyulan. (-RA “poke” does not occur in CVCs in Warrwa; it seems likely that this reflects loss in the combinatorial potential of this IV.)

It is also significant that, in some cases where IVs are not cognate across all of the Nyulnyulan languages, corresponding non-cognate IVs are found in collocation with some of the same PVs. This is especially evident for the “get” IV, for which Warrwa and Nyikina have replaced the inherited reflex of *-NYA with -ANDI. Nonetheless, reflexes (or replacements) of four Proto-Nyulnyulan PVs – (10) *jub

9. Some pairings are more expected than might appear at first, due to the fact that the glosses for the IVs are in some instances rather misleading. This is particularly the case for -JU “say”, which includes “think” and “do” among its senses. There is no single good gloss in English for this IV.

“cut”, (13) *dumbarra “fly”, (19) *jurnk “run”, and (24) *lanyb “steal” – collocate with -ANDI “get” in Warrwa and/or Nyikina, and with a reflex of *-NYA “get” in one or more other Nyulnyulan languages. And these pairings are not expected, except perhaps for (24) “steal”. It would seem most likely that the collocations with the “get” IV can be traced back to Proto-Nyulnyulan, and that they were retained in Warrwa and Nyikina even when the IV itself was replaced.

The “hit” IV plays a very minor role in CVCs in Nyulnyulan languages, a fact differentiating them from neighbouring languages, in which it usually plays a fairly significant role. However, in a number of Western Nyulnyulan languages it occurs with (14) *da* “hammer”, even though no IV is reconstructible. Of course, this collocation is not unexpected. On the other hand, its significance is enhanced by the fact that no corresponding collocation exists in Eastern Nyulnyulan languages.

To wind up this section, I suggest in Table 5 the likely collocations of the Proto-Nyulnyulan PVs in that language and also in the two daughter languages, where they seem to have innovated. Collocations are attributed to Proto-Nyulnyulan only when supporting evidence comes from four or more languages, including at least one language from each of the two primary branches. In a few cases where three modern languages (representing both branches) share a collocation this is included in the table with a question mark. Collocations are attributed to the two daughter proto-languages when three modern languages from the corresponding group show the same collocation.

Table 5. Reconstructed CVCs in the proto-languages

	Proto-PV	Gloss	Collocating *IVs	Proto-language of collocation
1	*nunyji	“be alive”	*-NI “sit” *-KAL “wander”	Proto-Nyulnyulan Proto-Western Nyulnyulan
2	*bili	“angry”	*-JU “say” *-NGARA “become” *-MA “put”	Proto-Nyulnyulan Proto-Eastern Nyulnyulan Proto-Eastern Nyulnyulan
3	*buu	“blow”	*-MA “put”	Proto-Nyulnyulan
4	*kawu	“call out”	*-JU “say”	Proto-Nyulnyulan
5	*kundi	“carry on shoulder”	*-KA “carry” *-NYA “get” *-NI “be”	Proto-Nyulnyulan Proto-Western Nyulnyulan Proto-Western Nyulnyulan
6	*kiny	“choke, strangle”	*-RA “poke” *-NYA “get” *-WA “give”	Proto-Nyulnyulan Proto-Nyulnyulan Proto-Nyulnyulan (?)
7	*lakal	“climb”	*-JU “say” *-WANYJI “climb” *-NYA “get” “go”	Proto-Nyulnyulan (?) Proto-Nyulnyulan Proto-Nyulnyulan Proto-Nyulnyulan (?)

(continued)

Table 5. (continued)

	Proto-PV	Gloss	Collocating *IVs	Proto-language of collocation
8	*yadab	“crawl”	*-JU “say”	Proto-Nyulnyulan
9	*wangkurr	“cry, weep”	*-JU “say”	Proto-Eastern Nyulnyulan
10	*jub	“cut”	*-NYA “get”	Proto-Nyulnyulan
			*-BANJI “exchange”	Proto-Nyulnyulan
11	*burrb	“dance”	*-JU “say”	Proto-Nyulnyulan
12	*yuwurr	“descend, sink”	*-JU “say”	Proto-Nyulnyulan (?)
			*-MA “put”	Proto-Nyulnyulan
13	*dumbarra	“fly”	*-JU “say”	Proto-Nyulnyulan
			*-NYA “get”	Proto-Nyulnyulan
14	*da	“hammer”	*-JU “say”	Proto-Nyulnyulan (?)
			*-RA “poke”	Proto-Nyulnyulan
			“hit”	Proto-Western Nyulnyulan
15	*yaly	“lick”	*-JU “say”	Proto-Nyulnyulan
16	*jarrbard	“lift up and carry”	*-JU “say”	Proto-Nyulnyulan
			*-KA “carry”	Proto-Nyulnyulan
17	*liyan	“heart, emotion”	*-MA “put”	Proto-Nyulnyulan
			*-NYA “get”	Proto-Nyulnyulan
			*-KA “carry”	Proto-Nyulnyulan
			*-JU “say”	Proto-Nyulnyulan
			*-BANJI “exchange”	Proto-Western Nyulnyulan
18	*jur(r)urr	“pour out”	*-RA “poke”	Proto-Nyulnyulan
			*-JU “say”	Proto-Nyulnyulan
19	*jurnk	“run”	*-JU “say”	Proto-Nyulnyulan
			*-NYA “get”	Proto-Nyulnyulan
			*-JID “go”	Proto-Western Nyulnyulan
			*-NGARA “become”	Proto-Eastern Nyulnyulan
20	*jirrmu	“sing”	*-JU “say”	Proto-Nyulnyulan
21	*mijala	“be sitting”	*-NI “sit”	Proto-Nyulnyulan
			*-MA “put”	Proto-Nyulnyulan
22	*ja(r)l	“split”	*-JU “say”	Proto-Nyulnyulan
			*-MA “put”	Proto-Eastern Nyulnyulan
			*-JID “go”	Proto-Western Nyulnyulan
23	*yalku	“be standing”	*-NI “sit”	Proto-Nyulnyulan
			“stand (up)”	Proto-Nyulnyulan
24	*lanyb	“steal, abduct”	*-NYA “get”	Proto-Nyulnyulan
25	*juny	“suck”	*-JU “say”	Proto-Nyulnyulan
			*-NYA “get”	Proto-Western Nyulnyulan
26	*dibirr	“turn”	*-JU “say”	Proto-Nyulnyulan
			*-BANJI “exchange”	Proto-Nyulnyulan
			*-MA “put”	Proto-Nyulnyulan (?)
27	*bilbil	“twitch, twinkle”	*-JU “say”	Proto-Nyulnyulan (?)

3.3 PV-IV collocations in other Kimberley languages

To conclude the broad comparison of PV-IV collocations in Nyulnyulan languages I attempt to give an idea of the significance of the habitual collocations in Nyulnyulan by comparing them with collocations in nearby languages belonging to different families. Unfortunately detailed and reliable data is available only for a few languages, and for a number of the meanings listed in Table 3 (e.g. (1) “be alive”, (18) “pour out” and (27) “twinkle, twitch”) it is possible to find information in just a few languages. Partly for this reason, and partly to keep the comparison within reasonable limits, instead of using the twenty seven meanings of §3.1 I use the smaller sample of ten basic verbal meanings employed in McGregor & Rumsey (2009).

Table 6 shows the expressions of the ten basic meanings in nine non-Pama-Nyungan Kimberley languages, three Nyulnyulan (Warrwa, Nyikina, Bardi), two Jarrakan (Miriwoong, Kija), one Bunuban (Bunuba), and three Worroran (Worrorra, Ngarinyin, Wunambal). This is a compilation from Tables 17 and 18 in McGregor & Rumsey (2009: 57–58, 60–63), with some additional more recent information included.

This tabulation reveals less agreement in verbal expression than what is found within Nyulnyulan. First, of the ten basic meanings, two – “see” and “hear” – are expressed by SVCs only in the three Nyulnyulan languages. Second, where collocations are shared with other languages, in almost all instances the IV involved is the “say, do” one. The few exceptions involve the positional PVs “sit” and “stand” which, predictably, collocate with “be” in virtually all languages, and the PV “die” which collocates with “go” in a range of languages. (And the CVC expression of this meaning is likely to be an innovation of Proto-Eastern Nyulnyulan.) Third, the more lexically specific IVs – such as “get”, “hit”, “climb”, “fall” – are not involved in any shared collocations involving Nyulnyulan and any other family. Similarly, the IV “exchange” – a suppletive form of the “give” IV – is unique to Nyulnyulan languages. Where another language shows a corresponding reflexive/reciprocal IV in a collocation this is of “say” or “hit”, not “give”. By contrast, in the collocations shared among Nyulnyulan languages are a number involving the more lexically specific IVs, and not just the “say” IV.

Table 6. Expression of ten basic verbal meanings in nine Kimberley non-Pama-Nyungan languages

	Warrwa	Nyikina	Bardi	Miriwoong ^a	Kija ^b
“climb”	<i>lagarr</i> + JI “say” [–WANYJI]	<i>lakarr</i> + I “say” <i>lakarr</i> + WANYJI “climb” <i>lakarr</i> + NDA “go” [–WANYJI]	<i>lakal</i> + GANYI “climb” <i>lakal</i> + JIID “go” <i>lakal</i> + NY “get” [–GANYI]	<i>berdij</i> + “do” <i>berdij</i> + “go”	<i>pertij</i> + “do” <i>pertij</i> + “be”
“cry, weep”	<i>wangkurr</i> + JI “say” <i>wangkurr</i> + NGARLA “cry” [–NGARLA]	<i>wangkoorr</i> + I “say” <i>wangkoorr</i> + LOOKA “cry” [–LOOKA]	<i>anggoorr</i> + MOOROO “spill” <i>anggoorr</i> + MA “put” [–NGALGA]	<i>nyuringbe</i> + “do”	<i>ngartawu</i> + “do”
“cut”	<i>jub</i> + ANDI “get” <i>jub</i> + BANJI “exchange” <i>jub</i> + WALI “cut”	<i>job</i> + BANJI “exchange”	<i>yaarl</i> + NYA “get”	<i>gad</i> + “go” [–I]	<i>katij</i> + “do to each other”
“die”	<i>kud</i> + ARNDA “go”	<i>koord</i> + MA “go” <i>koord</i> + NI “sit”	[–JIIBI] [–BANYI]	<i>juwarig</i> + “sit” <i>juwarig</i> + “fall”	<i>tijpila</i> + “fall” <i>tijpila</i> + “be”
“dig”	<i>widij</i> + YI “say”	<i>widij</i> + I “say”	[–GALBOO]	<i>guraj</i> + “do” <i>guraj</i> + “gather”	<i>jitijata</i> + “do”
“eat”	<i>kab</i> + YI “say” <i>kab-kay</i> + WANI “be” [–LI]	<i>kab</i> + I “say” [–LI]	[–(A)RLI–]	<i>jangab</i> + “sit” [–NGAN]	<i>jang</i> + “do” <i>jang</i> + “take” [–NGUN]
“hear”	[–LARRA]	[–LIKARRA]	[–LAMANKA]	<i>rangga</i> + “sit” <i>ranga</i> + “take” <i>ranga</i> + “hit”	<i>rangka</i> + “be” <i>rangka</i> + “take” <i>rangka</i> + “do” ^d
“see”	[–JALA]	[–BA(RA)]	[–JALA]	<i>balaj</i> + “do” <i>balaj</i> + “hit”	<i>tek</i> + “do” <i>tek</i> + “get” <i>tek</i> + “take” <i>tek</i> + “do to one another”

Bunuba ^c	Worrorra	Ngarinyin	Wunambal
<i>bara</i> + NI “poke?”	<i>baay</i> + NU/NI “be”	<i>baj</i> + YI “be”	<i>baiba</i> + N “be”
<i>wala</i> + RA “sit”	<i>wala</i> + NU/NI “be” <i>wala</i> + YI “say, do” <i>wala</i> + MA “get”	<i>warda</i> + YI “be” <i>warda</i> + MA “say, do” <i>warda</i> + MA(RA) “get, take” <i>warda</i> + YILA “hold”	<i>wala</i> + MA “say, do” <i>wala</i> + MIRRA “go to”
<i>gayga</i> + MA ₂ “say”	<i>durr</i> + ^B WU “hit” <i>durr</i> + ^B WEE “hiton” self”	<i>durr</i> + WU “act” <i>durr</i> + A “go”	<i>dirr</i> + WU “hit” <i>dirr</i> + WU ₁ “hit self”
<i>duluga</i> + WU “hit”	<i>debarr</i> + YA “go”	<i>debarr</i> + A “go”	<i>debarr</i> + YA(N) “go”
<i>wirriyga</i> + MA ₂ “say”	<i>jarri</i> + ^B WU “hit”	<i>jarri</i> + WU “act on”	<i>jarri</i> + WU “hit” <i>jarri</i> + MA “say, do” <i>jarri</i> + YA(N) “go”
<i>nga(g)</i> + RA “sit” <i>nga(g)</i> + RA ₂ “carry”	<i>minjarl</i> + NU/NI “be” <i>minjarl</i> + ^B WU “hit”	<i>minjarl</i> + YI “be” <i>minjarl</i> + WU “act on”	<i>minja(l)</i> + N “be” <i>minja(l)</i> + MA “say, do”
<i>winyi</i> + RA ₂ “carry”	<i>nguru</i> + NU/NI “be” <i>nguru</i> + AANGURRU “carry”	<i>nguru</i> + YI “be” <i>nguru</i> + MINDA “take”	<i>nguru</i> + N “be” <i>nguru</i> + MINDA “take”
<i>mila</i> + RA ₂ “carry”	<i>mara</i> + ^B WU “hit”	<i>mara</i> + WU “act on”	<i>mara</i> + WU “hit”

(continued)

Table 6. (continued)

	Warrwa	Nyikina	Bardi	Miriwoong ^a	Kija ^b
“sit”	<i>mijala</i> + WANI “be”	<i>mijala</i> + NI “sit” [-NI] <i>mijala</i> + NGANYJ “sit” <i>mijal</i> + MA “put”	<i>miyala</i> + NI “be” [-NI]	<i>lulu</i> + “sit” <i>lulu</i> + “fall” [-NI]	<i>rurt</i> + “do” <i>rurt</i> + “be” <i>rurt</i> + “fall” [-N]
“stand”	<i>yaalu</i> + WANI “be” <i>yaalu</i> + JARRA “stand” <i>yaalu</i> + BULA “emerge” <i>yaalu</i> + MA “put” [+ JARRA]	<i>yalkoo</i> + NI “be”	<i>jirrijir</i> + JOO “do” <i>jirrijir</i> + NYA “get”	<i>bare</i> + “sit” <i>bare</i> + “fall”	<i>that</i> + “do” <i>that</i> + “be” <i>that</i> + “fall”

Note

- a All Miriwoong data comes from Kofod (1976, 1978). Because of the large amount of root suppletion and morphophonemic alternation in Miriwoong verb forms, Kofod does not try to give underlying forms for the roots. Therefore just their glosses (“do”, “be”, “fall”, etc.) are listed, except in the case of simple verb constructions, where only one of the allomorphs is given.
- b All Kija forms showing the preverbs paired with English glosses for their auxiliary verbs come from Taylor (1967), Taylor & Hudson (1976), and Kofod (n.d.). Again, as for Miriwoong, just the glosses of the IVs that occur with the preverbs are listed, and one of the allomorphs when the inflecting verb is used independently in a simple verb construction.
- c None of the Bunuba inflecting verbs whose roots are shown here (-NI, -WU, etc.) ever occurs outside of a compound verb construction, so it is impossible to give a gloss for them that is comparable in lexical specificity to the ones for other languages shown in this table. What has been given instead is a gloss for the hypothesized free IV that existed previously in some earlier pre-Bunuba.
- d Textual examples in Taylor (1967) show *rangka* + “do” glossed as “listen”.

Bunuba ^c	Worrorra	Ngarinyin	Wunambal
<i>yatha</i> + RA “sit”	<i>aja</i> + NU/NI “be” <i>aja</i> + ^B WA “fall” <i>ajag</i> + ^B WA “fall”	<i>ada</i> + YI “be” <i>ada</i> + WA “fall” <i>adag</i> + WA “fall”	<i>atha/ada</i> + N “be” <i>ada</i> + MA “say, do” <i>athag/adag</i> + (A)WA “fall”
<i>wara</i> + RA “sit”	–	<i>darr</i> + MA “say, do” <i>darr</i> + ININGA “put”	<i>dad</i> + MA “do” <i>dad</i> + NINU? “put”

4. Comparison of CVC systems

The evidence presented in the previous section strongly suggests that at least some pairings of IVs with PVs in Nyulnyulan languages are inherited from Proto-Nyulnyulan and its two daughter proto-languages. We now adopt a system-level perspective on CVC-based verb classification in the modern languages. We focus attention on two geographically and genetically distant languages, Nyulnyul and Warrwa, these being the languages I have worked most intensively on myself, and have access to most data for. We begin in §4.1 by comparing the two systems in terms of the semantics, then in §4.2 make some comparative observations about the extensions of the categories.

4.1 Semantics of Nyulnyul and Warrwa systems of verb classification

The Nyulnyul and Warrwa systems of verb classification are virtually identical in terms of the categories they distinguish and the semantic principles behind the categorization. This is the case for the most productive categories, the categories that have at minimum five member PVs – where the categories are the sets of PVs attested in collocation with a given IV. As shown in Table 7, the two languages agree for eight of the ten categories, both in terms of the IVs that mark them, and their semantic specifications. Just two categories are unique to each language, and these are among the smallest in terms of the number of collocating PVs. Moreover, only one

of the four IVs marking these categories can be reconstructed to Proto-Nyulnyulan; the other three were most likely innovated in the proto-languages of the two main groups in the family.

Table 7. Semantic characteristics of categories defined by the ten primary classifying IVs in Nyulnyul and Warrwa

Nyulnyul	Warrwa
Atelic	
-N “be” [stative]	-NI ~ -WANI “be” [stative]
-JID “go” [activity progresses over time]	-ARNDA “go” [activity progresses over time]
-KAL “wander” [action not uniquely directed towards a specific goal]	-NGARA “become” [inchoative]
-K “carry” [move something by constantly applied force to new location]	-KA “carry” [move something by constantly applied force to new location]
Telic	
-R “pierce” [action taking place in a straight line, impacting on something at a point]	-NKA “hit” [impact violently on something]
-W “give” [action directed outwards from actor, making contact with something]	-WA ~ -ø “give” [action directed outwards from actor, making contact with something]
-NY “get” [acquire or achieve an entity or condition by active means]	-ANDI “get” [acquire or achieve an entity or condition by active means]
-M “put” [induce something to enter new state, condition, or location]	-MA “put” [induce something to enter new state, condition, or location]
± Telic	
-BARNJ “exchange” [reflexive/reciprocal action]	-BANYJI “exchange” [reflexive/reciprocal action]
-J “say, do” [dynamic activity]	-JI “say, do” [dynamic activity]

As we have already seen (Table 2), there is much less agreement between Nyulnyul and Gooniyandi in terms of the relative sizes of the categories, and their markers (virtually none are plausible cognates). More importantly, the semantic bases for the categories are rather divergent, although the same general parameters (telicity, valency, and vectorial configuration) are relevant. Thus, for instance, while the Aktionsart contrast between telic and atelic is maintained throughout the Gooniyandi system (see e.g. McGregor 1990: 557, 2002: 58, where the contrast is construed as between accomplishments and extendible events), it is neutralized for two of the most general categories in the Nyulnyul system, which specify nothing about telicity. As a result, the “say” categories of the two languages are quite

different in terms of their semantic specification: in Nyulnyul they are activities, while in Gooniyandi they are verbs of effect. And while the -K “carry” category in Nyulnyul is quite small, the corresponding +A category in Gooniyandi (the marker of which derives from a former IV *-GA “carry”) is the largest and most generic in Gooniyandi, specifying nothing more than atelicity and bivalence. The Nyulnyul -K “carry” category is much more semantically specific, and includes events in which an entity is relocated in space, state, or condition, by constant application of energy from an instigator that remains in contiguity with it.

4.2 Comparison of the extensions of the Nyulnyul and Warrwa categories

Although the intensional semantics of the shared major categories are virtually identical, the two systems of classification show a considerable amount of divergence in terms of their actual assignment of PVs to the categories. This happens partly because the intensional semantics of the categories do not just overlap (the most general categories also include a number of specific ones semantically), but also a given PV might equally be assigned to two or more different categories according to which aspect of the event is attended to. For instance, the Warrwa and Nyikina PV *muk* “hit” could be assigned to the general dynamic activity category -JI “say” (as it in fact usually is). But it is also consistent with the specifications of the Warrwa categories marked by -WA “give”, -NKA “hit”, and of course -BANYJI “exchange”.

In this section we overview similarities and differences in the collocate sets of the categories in the two languages, i.e. the extensions of the categories.¹⁰ A convenient way of doing this is via category correspondences: two categories are said to correspond in terms of extension if there exist a pair of PVs, one in each language, that – in CVCs involving the IVs marking the categories – express nearly

10. This exercise is fraught with difficulties. Numerous verbal meanings are represented by PVs in just one language. This is sometimes because the available corpora in both languages show accidental gaps, and the meaning is attested in just one of them. Sometimes it is because the meaning expressed by a CVC in one language is expressed by an SVC in the other – that is, the meaning expressed by a PV-IV pairing in one language is represented by a single IV in the other. For the purposes of our comparison of category extension we restrict attention to meanings that are (or may be) expressed by PVs (in CVCs) in both languages, discarding those cases in which (for whatever reason) just one of the languages uses a PV. Of course, the result of this decision is that the categories may appear more similar in extension than they really are. On the other hand, to do otherwise would potentially make too much of accidental differences (both in terms of accidents of the corpus, and accidents of language history). The effect of this decision is that over half of the PV-IV collocations in Nyulnyul were discarded, as were over a third in Warrwa.

synonymous meanings.¹¹ If we take the category correspondences for all of the PVs belonging to each of a pair of categories in the two languages, the result is an indication of the degree of extensional similarity of the categories.

When this is done for the correspondences for the eight major shared categories of Nyulnyul and Warrwa, a rather confusing picture emerges, reminiscent of the correspondences between category extensions of Gooniyandi and Nyulnyul, as per McGregor (2002: 179). That is to say, the correspondences between Nyulnyul and Warrwa categories are many:many.

However, if the situation is examined a little more closely, it turns out to be not quite as messy as it first appears. Specifically, the shared categories generally also emerge as the dominant extensional correspondents. That is, in numerical terms the majority of correspondences for any shared categories are between these two categories. Usually only a handful of correspondences exist with other categories.

To demonstrate these observations, let us first consider two shared category pairs, -JID/-ARND “go”, and -K/-KA “carry”. Figures 1 and 2 depict the level of correspondences between the respective categories in the two languages in terms of shared extensions, as well as the correspondences with other categories in terms of shared extensions. (Note that the sets of correspondences for the other categories are not indicated – e.g. we show the correspondences of the “say” category to the target categories, but not to any other categories.) Thick black lines indicate the most frequent correspondences; grey lines show less frequent correspondences, and broken lines indicate correspondences manifested by just a couple of PVs.

Both shared categories agree best in extension with the corresponding category in the other language, although for the “go” category another correspondence is also almost as numerous, i.e. the correspondence between Warrwa “go” and Nyulnyul “say”. In fact, restricting attention to the sets of PVs in the correspondences of Figure 1, overall less than half of the shared PV pairings in each language involve the “go” IV. Nonetheless, the numerically dominant correspondence in extension remains between the two “go” categories. For “carry”, the shared collocations are rather more numerous, and pairings of the PVs outside of the two extension sets are less frequent, and (unlike the situation for “go”) none reaches anything like as

11. The need for this requirement of synonymy at the level of the CVC derives from the fact that a single PV is often multiply categorized in a given language. Thus corresponding synonymous PVs might well occur in the extensions of two or more categories in each language, making it impossible to sensibly associate any specific categories. However, as remarked in §2 above, multiple categorizations express different senses, so the category correspondences should be restricted to cases in which the CVC involving the semantically comparable PV in both languages are near synonyms. In effect, this means that we should count as a correspondence in extension of a category ‘be closed’ and ‘become closed’, but not ‘be closed’ and ‘close something’.

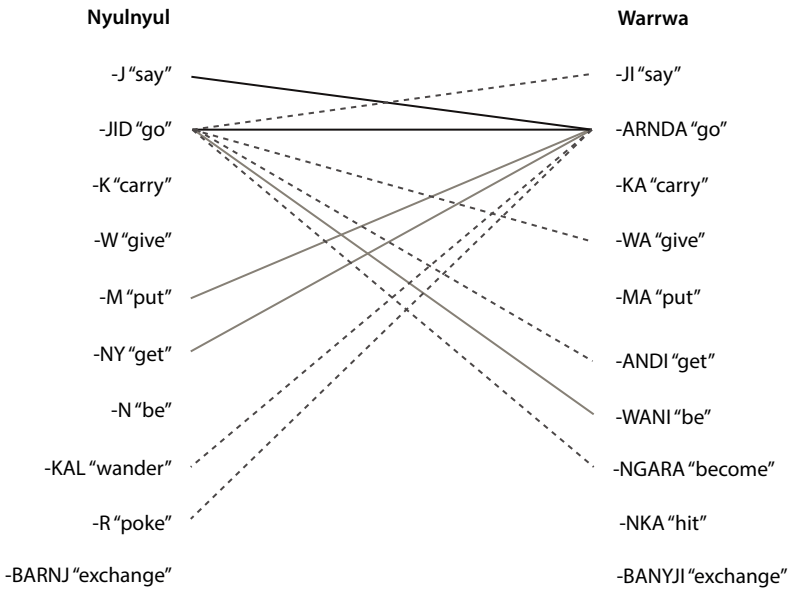


Figure 1. Correspondences of collocate sets for “go” categories in Nyulnyul and Warrwa

many instances as the “carry” correspondences. A similar picture emerges for the other shared categories. This is supported by Figure 3, which shows the correspondences for the “put” categories. In this instance, over half of the correspondences in extension are shared between the “put” categories in the two languages, with the next most frequent being represented by less than a sixth of the correspondences – and the others by just a handful.

Rather than tediously illustrate this for each of the shared categories, we turn now to the most generic category in the two languages, the “say” category. In both languages a comparable range of meaning domains is covered, including: (1) **vocalizations and noises** (of both animates and inanimates); (2) emission of light and heat; (3) attention and cognition; (4) emotions; (5) **body moves and behavior** (including gestural signs); (6) **motion events**; (7) social activities; (8) **inchoatives and induced states**; and (9) **violent actions on an object**. The bolded domains represent the largest sets of PVs in both languages. Despite this high level of agreement, the collocate sets of PVs in the nine groups show a considerable amount of divergence. Much of this can be attributed to differences in the corpora (the Nyulnyul collocate set is much larger than the Warrwa one); a small amount is due to a given meaning being expressed just by an IV in Nyulnyul, but by a CVC in Warrwa. In just a few instances a PV collocates with the “say” IV in one language, while one with a comparable meaning in the other language collocates instead with another IV. For

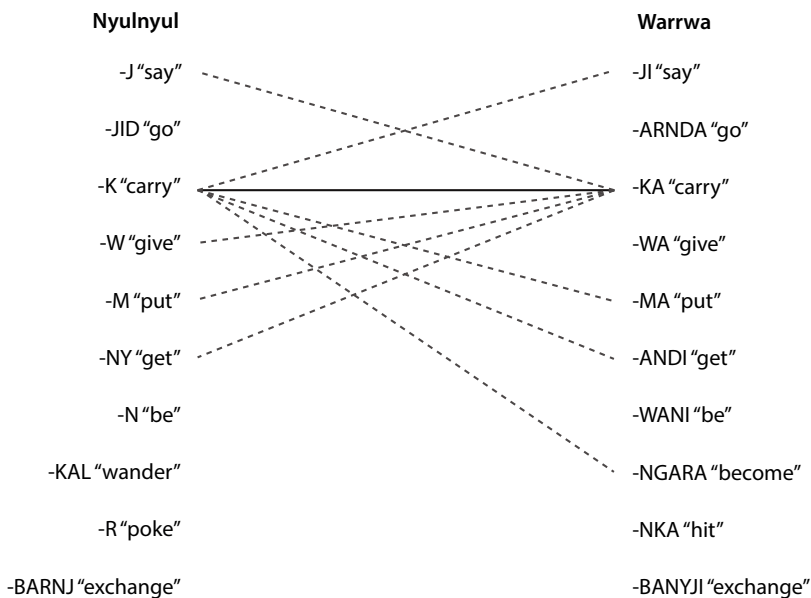


Figure 2. Correspondences of collocate sets for “carry” categories in Nyulnyul and Warrwa

instance, the Warrwa IV *jurrmak* “punch” collocates with -JI “say”, whereas the corresponding Nyulnyul PV *dud* “punch” collocates with -K “carry” and -W “give” but not with -J “say”. The majority of correspondences involve the “say” categories in each language, other correspondences being relatively few in number. The pattern is essentially the same as for the “carry” and “put” categories.

The level of agreement between Nyulnyul and Warrwa in the “say” category is far more significant than between either language and Gooniyandi. For although many of these domains are also covered by the corresponding + MI “effect” category in Gooniyandi, only those events that are accomplishments, telic events, are included. So while in Nyulnyul *kaw* ... -J “call out ... say” can be used in reference to an atelic event of calling out that is not represented as specifically directed to an addressee, in Gooniyandi the corresponding *baa*- ... + MI “call out ... effect” refers only to an accomplishment; + I “be, go” is used with *baa*- “call out” if the referent event is an atelic one that is not directed at a specific addressee (see also McGregor 2002: 191).

The cognate “exchange” IVs in Nyulnyul and Warrwa show less agreement than one might expect given that they mark generic reflexive/reciprocal categories. Just ten correspondences in extension are attested from more than twenty collocating PVs in each language. Moreover, some of these are quite tenuous (e.g. *barabar* “hit”

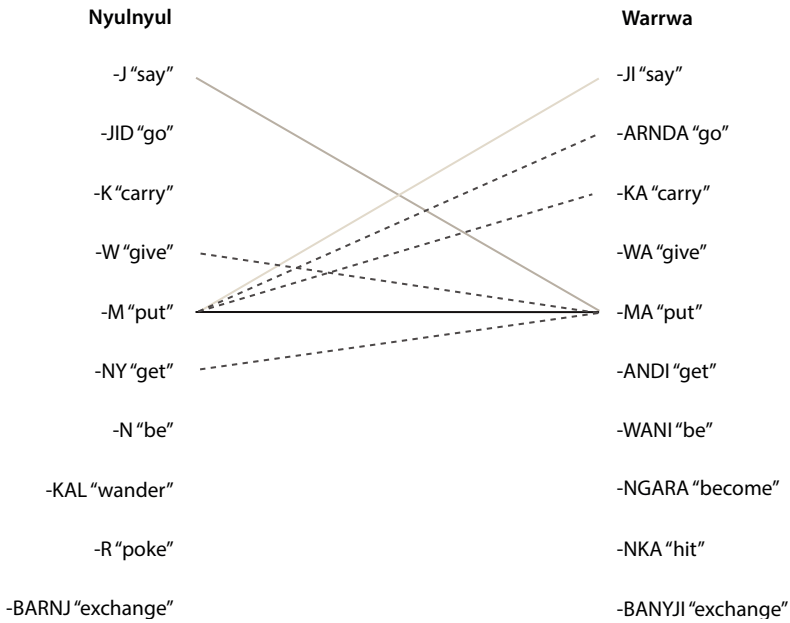


Figure 3. Correspondences of collocate sets for “put” categories in Nyulnyul and Warrwa

in Nyulnyul with *ruk* “hit” in Warrwa; the former PV is in fact inherently reflexive in meaning – McGregor 2012: §11.4.1.1). Nevertheless, only a scattering of extensional correspondences exists with any other category, so that again the strongest correspondence is between the two “exchange” categories.

To summarize the discussion of this section, what we have seen is that Nyulnyul and Warrwa share eight categories that are effectively identical in their semantic specifications. Although the extensions of the categories are rather divergent, once we allow for accidental differences in the corpora, the most predominant correspondences in extensions are between the shared categories. Overall, shared categories are more like one another in extension than they resemble any other category: they share more in common with one another in terms of extension than they share with any other category. Despite the fact that the correspondence picture initially looks as messy as that for Gooniyandi and Nyulnyul (as per McGregor 2002: 179), for Nyulnyul and Warrwa usually just one correspondence predominates. This contrasts with the situation for the former pair of languages, where in many cases a number of correspondences are about equally strong quantitatively.

5. The evolution of verb classification in Nyulnyulan languages

5.1 The Proto-Nyulnyulan CVC as a verb classification system

In §3.2 we suggested that it is possible to reconstruct some PV-IV collocations in Proto-Nyulnyulan (and also in Proto-Eastern and Proto-Western Nyulnyulan), indicating the existence of the CVC in the proto-language. Given the paucity of data, in particular the small number of PVs that have so far been reconstructed for Proto-Nyulnyulan, it is difficult to say anything with certainty about the properties of the proto-CVC. Nonetheless, it is possible to say something, and the evidence suggests that indeed the CVC was already a verb classifying system in the proto-language. Let us review the evidence.

- a. It is evident from the collocations shown in Table 5 that the IVs were not always being used in their full lexical meanings. The IV *-BANJI “exchange”, for instance, was evidently already being used of reflexive/reciprocal events that did not involve exchange of physical objects; the corresponding root *-WA “give” was perhaps also being used without its full lexical meaning in collocation with *kiny “choke, strangle”. Nor was *-NYA “get” being used exclusively for events of capturing or collecting, but also for inceptive events. And similarly for *-RA “poke”, events showing an abstract straight line vector already involved expressions with this IV.
- b. More importantly, as revealed in Table 5, it appears that already in Proto-Nyulnyulan a number of PVs were collocating with more than one IV. This is a common feature of verb classification systems in northern Australia (McGregor 2002), as in modern Nyulnyulan languages, and presumably there existed a concomitant semantic difference. For instance, for (13) *dumbarra “fly” and (19) *jurnk “run” it seems likely that the collocation with *-JU “say” denoted the activity, while the collocation with -NYA “get” denoted the act of starting the motion event, as in the modern languages. For (14) *da “hammer”, the collocation with *-JU “say” doubtless denoted any activity of pounding, while its collocation with *-RA “poke” would presumably have denoted a more specific type of pounding with an elongated instrument such as a pestle. Those proto-PVs that collocate with *-BANJI “exchange” all collocate with some other IV, and the former evidently indicated a reflexive/reciprocal event in contrast to a transitive or intransitive event for the latter.
- c. Many of the pairings of PV and IV are unpredictable, or at least unusual. For instance, of the motion PVs just one appears to have been at all likely to have collocated with the “go” IV in Proto-Nyulnyulan, (7) *lakal “climb”. And this is perhaps the least expected of them, given that motion is in the vertical

dimension. Almost all collocated with *-JU “say”, which while not remarkable in itself, is unusual in the context of the rarity of collocations with the generic “go” IV. The collocation of (10) *jub “cut” with *-NYA “get” is also unexpected, especially given that (22) *ja(r)l “split” did not collocate with this IV, but rather with the generic *-JU “say”. One also most naturally expects (14) *da “hammer” to collocate with the “hit” IV, though this is only evidenced in Proto-Western Nyulnyulan.

- d. Our reconstructed CVCs include few instances where the PV and IV express almost identical meanings, as is the case for (7) *lakal “climb” and *-WANYJI “climb”, (21) *mijala “sit” and *-NI “sit”. These are reminiscent of repeaters, such as are found in many classification systems including verb classification systems (McGregor 2002). In addition, there are a few cases in which the IV expresses a closely related, though more generic, meaning than the PV, as in the case of collocations of *-KA “carry” with *kundi “carry on shoulder” and *jarrbard “lift up and carry”.

These four features are common in grammaticalized systems of classification, although they are not diagnostic features. In the absence of direct evidence in the shape of large numbers of PV-IV collocations, this adds some weight to the suggestion that the Proto-Nyulnyulan CVC was a verb classifying system. Two other observations are pertinent, representing recurrent features of Australian verb classification systems, lexicalization of the collocations and the generic semantics of the IV.

- e. Some PV-IV pairings were evidently already habitual in Proto-Nyulnyulan, and semantically non-compositional; they were presumably part of the language’s resources for construction of new lexemes. Indeed, at least one of the proto-CVCs looks as though it may have been an idiomatic expression, namely the collocation of (17) *liyan “heart, emotion” with *-MA “put”, and perhaps also the collocation with *-KA “carry”.
- f. The evidence strongly supports the notion that the IVs were being used generically in CVCs, which gives us at least a starting point for a system of verb classification.

Unfortunately, due to the paucity of data from a number of Nyulnyulan languages it may prove impossible to reconstruct much more of the Proto-Nyulnyulan CVC, except very tentatively. Given the rather tenuous status of the Proto-Nyulnyulan CVC as a verb classification system, one needs to be cautious in drawing conclusions about it. Nevertheless, the collocations in Table 5 suggest that already by this stage in the evolution of the languages the classifying potential of the IVs was shaping into something like what we see today:

*-BANJI “exchange”	reflexive/reciprocal events
*-JU “say”	activities
*-NI “sit”	states
*-NYA “get”	inchoatives/inceptives
*-MA “put”	causatives and inchoatives
*-RA “poke”	events involving a salient straight line vector
*-KA “carry”	motion in association with
*“go”	motion

In fact, given the identity of the semantic specifications of the shared categories in Nyulnyul and Warrwa, it is not implausible to presume that they were the same in Proto-Nyulnyulan. The case is strengthened by the findings of §4.2 that corresponding shared categories are most similar to one another in their extensions.

5.2 Subsequent evolution of verb classification in Nyulnyulan

The CVC system of Proto-Nyulnyulan almost certainly had a lower functional load than in any of the modern languages: there were presumably many more IVs than in any of the daughter languages, PVs were probably only just beginning to come into use, and CVCs just beginning to replace SVCs in the expression of verbal meanings. This is supported by the fact that reconstructed Proto-Nyulnyulan PVs are so few in number. Quite likely this mode of expression was an areal feature shared with neighboring languages, Proto-Bunuban and/or Proto-Worrوران.

By the time of separation of Proto-Eastern Nyulnyulan from Proto-Western Nyulnyulan, we see evidence of expansion of the system in the shape of emergence of new IVs in CVCs, of *-NGARA “become” in Eastern Nyulnyulan and *-KAL “wander” in Western Nyulnyulan. The Proto-Nyulnyulan “go” IV was replaced in one or both of the groups, though the replacing IV remained in use in CVCs, presumably in the same ones.

The trajectories of development of the CVC subsequently diverged. In Eastern Nyulnyulan languages many SVCs were replaced by CVCs, and many IVs went out of use. Western Nyulnyulan languages preserved more IVs, and CVC expression did not take over to the extent that it did in the sister Eastern group. Nevertheless, both branches of the family maintained a considerable number of IVs – more than in neighboring non-Pama-Nyungan languages – even though CVC expression was sometimes available. Fewer IVs were retained in the Eastern group than in the Western group, but in both groups only a small number of IVs had the potential of occurring in CVCs.

Areal factors were evidently at play. The Eastern Nyulnyulan languages are in direct contact with mainland Kimberley languages, in which, as just mentioned,

the classifying systems are more entrenched. Although not completely insulated from either the Kimberley mainland or islands, the Western Nyulnyulan languages were relatively more isolated, being locked into the Dampier Land peninsula, and perhaps subjected to less areal pressure to adopt the CVC as the dominant mode of expressing verbal meanings.

6. Conclusions

This paper has presented evidence that the CVC systems of Nyulnyulan languages are retentions from Proto-Nyulnyulan, which shared the system with neighboring languages, or perhaps even borrowed it from them. Whether the proto-system was a verb classifying system may well be impossible to ever determine with complete certainty. Nevertheless, the evidence supports the idea that even in the proto-language IVs were used as generic verbs in CVCs which had already reached a degree of conventionalization; this is consistent with the status of the proto-CVC as a verb classifying system. Under this assumption, we can make a beginning on reconstructing the system, including its semantic basis.

Thus Bowern (2008) is far too negative about the possibilities of tracing the diachrony of the Nyulnyulan CVC, and allows the differences among the systems – genuine as they obviously are – to obscure the very real underlying similarities. I have shown that there is good overall agreement among the modern languages in terms of the category-marking IVs, most of which are cognates. Moreover, there is widespread agreement in the IVs collocating with a set of basic, frequent, and cognate PVs, which permits us to establish a set of proto-language PV-IV collocations. Some of these are relatively unexpected, or at least marked in the sense that they are not the most obvious collocation; these provide the best evidence for retention from the proto-language.

At a higher level, the level of the CVC, the similarities point to a common origin in Nyulnyulan, not borrowing. Significantly, the intensional characterizations of most categories in two languages from opposite geographical extremes and from different groups are virtually identical. The extensions of these categories also look promisingly comparable, and include a number of shared unexpected members. Overall, Warrwa and Nyulnyul show much more similarity at the system level than either language does with Gooniyandi. More impressionistically, the Nyulnyul and Warrwa systems are formally and semantically typically Western Nyulnyulan and Eastern Nyulnyulan, while the Gooniyandi system diverges in crucial respects, and is in turn much more similar to the Bunuba system (McGregor 2002: 96–98). Indeed, the discussion of McGregor (2002) shows that the verb classification system of northern Australian languages are most similar within families; this is supported

by more recent work which proposes metrics for measuring the distance between CVCs in different languages (McGregor 2013). If the CVC-based system of verb classification had spread areally subsequent to the splitting up of the families, there would be no reason to expect that the clustering of formal and semantic similarities would respect family divisions – as is evidently the case for the Warrwa desiderative complement construction, not found in other Nyulnyulan languages, which evidently spread into that language from its Worrوران neighbors after the separation of Warrwa from Nyikina (McGregor 2007).

One final observation is in order. Although Bownern (2008) is correct that ‘PVs’ often come from other word classes, this fact does not undermine my story. One ‘PV’ in our proto-collocations was evidently a nominal, *liyan* “heart, emotion”; looking further we also find *mada* “play, and *ningarra* “true”. As it turns out, we can reconstruct collocations in the proto-language for all three lexemes. The lexical category assignment of the initial item in the CVC does not affect either the analysis of the construction or the unpacking of its diachrony.

Acknowledgments

An earlier version of this paper was presented at the conference *The diachrony of classification systems* held at the Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS), Wassenaar, 12–13 March 2009, under the auspices of COST Action 31 *Stability and Adaptation of Classification Systems in Cross-Cultural Perspective*. I am grateful to the participants for comments on the presentation, and to an anonymous referee for comments on a previous draft of the paper. The usual disclaimers apply.

References

- Aklif, Gedda. 1991a. “Bardi Material”. Unpublished manuscript, Canberra, AIATSIS Aboriginal Studies Electronic Data Archive (ASEDA).
- Aklif, Gedda. 1991b. “Fieldnotes on the Bardi Language”. Unpublished manuscript, Canberra, AIATSIS.
- Aklif, Gedda. 1999. *Ardiyooloon Bardi Ngaanka: One Arm Point Bardi dictionary*. Halls Creek: Kimberley Language Resource Centre.
- Bownern, Claire. 2004a. “Diagnostic Similarities and Differences between Nyulnyulan and Neighbouring Languages. *Australian Languages: Classification and the comparative method* ed. by Claire Bownern & Harold Koch, 269–290. Amsterdam: John Benjamins.
doi:10.1075/cilt.249.16bow
- Bownern, Claire L. 2004b. *Bardi Verb Morphology in Historical Perspective*. Ph.D. dissertation, Harvard University.
- Bownern, Claire. 2008. “The Reconstruction of Nyulnyulan Complex Predication”. *Diachronica* 25(2).186–212. doi:10.1075/dia.25.2.04bow

- Hosokawa, Komei. 1991. *The Yawuru Language of West Kimberley: A meaning-based description*. Ph.D. dissertation, Australian National University.
- Kofod, Frances M. 1976. "Miriwung". *Grammatical Categories in Australian Languages* ed. by Robert M. W. Dixon, 646–653. Canberra: Australian Institute of Aboriginal Studies.
- Kofod, Frances M. 1978. *The Miriwung Language (East Kimberley): A phonological and morphological study*. M.A. thesis, University of New England.
- Kofod, Frances M. n.d. *Introduction to Kija Grammar*. Unpublished manuscript, Halls Creek, Kimberley Language Resource Centre.
- McGregor, William B. 1990. *A Functional Grammar of Gooniyandi*. Amsterdam: John Benjamins. doi:10.1075/slcs.22
- McGregor, William B. 2002. *Verb Classification in Australian Languages*. Berlin: Mouton de Gruyter. doi:10.1515/9783110870879
- McGregor, William B. 2007. "A Desiderative Complement Construction in Warrwa". *Language Description, History and Development: Linguistic indulgence in memory of Terry Crowley* ed. by Jeff Siegel, John Lynch & Diana K. Eades, 27–40. Amsterdam: John Benjamins. doi:10.1075/cll.30.o8mcg
- McGregor, William B. 2012. *The Nyulnyul Language of Dampier Land, Western Australia*. Canberra: Pacific Linguistics.
- McGregor, William B. 2013. "Comparing Linguistic Systems of Categorisation". *Approaches to Measuring Linguistic Differences* ed. by Lars Borin & Anju Saxena, 387–427. Berlin: De Gruyter Mouton. doi:10.1515/9783110305258.387
- McGregor, William B. & Alan L. Rumsey. 2009. *Worrorrnan Revisited: The case for genetic relations among languages of the Northern Kimberley region of Western Australia*. Canberra: Pacific Linguistics.
- Metcalfe, Christopher D. 1975. *Bardi Verb Morphology*. Canberra: Pacific Linguistics.
- Nekes, Hermann & Ernest A. Worms. 1953. *Australian Languages*. Fribourg: Anthropos Institut.
- Nekes, Hermann & Ernest A. Worms. 2006. *Australian Languages*. Berlin: Mouton de Gruyter.
- Stokes, Bronwyn. 1982. *A Description of Nyigina: A language of the West Kimberley, Western Australia*. Ph.D. dissertation, Australian National University.
- Stokes, Bronwyn & William B. McGregor. 2003. "Classification and Subclassification of the Nyulnyulan Languages". *The Non-Pama-Nyungan Languages of Northern Australia: Comparative studies of the continent's most linguistically complex region* ed. by Nicholas Evans, 29–74. Canberra: Pacific Linguistics.
- Taylor, Peter. 1967. "Kitja Materials". Unpublished manuscript, Canberra, AIATSIS.
- Taylor, Peter & Joyce Hudson. 1976. "Metamorphosis and Process in Kija". *Talanya* 3.25–36.

Author index

B

- Bates, Daisy 318
Bender, Bryon W. 183
Benton, Richard A. 173, 181-3
Bisang, Walter 2, 244, 246,
258-61 *passim*
Blust, Robert A. 174
Bowern, Claire 316-7, 349, 350
Brown, Dunstan 166
Brownie, John 179-80
Brownie, Marjo 179-80
Bybee, Joan L. 296-7

C

- Clark, Ross 184, 185
Craig, Colette G. 79, 80, 245,
251-3
Crowley, Terry 174

D

- Dammel, Antje 4
Donohue, Mark 5, 191

F

- François, Alexandre 178, 179,
189
Frenda, Allesio S. 2, 4

G

- Gil, David 145, 157
Greenberg, Joseph H. 145, 153,
158-9
Grinevald, Colette 202-3,
244-5, 262

H

- Hull, Kerry 4, 85-6
Hyslop, Catriona 178
Hosokawa, Komei 324

K

- Kaufman, Terrence 81, 194
Kießling, Roland 3
Klamer, Marian 190-1, 192, 193,
195, 196
Kofod, Frances M. 338
Krasnoukhova, Olga 202
Kürschner, Sebastian 4

L

- Lass, Roger 311
Lichtenberk, Frantisek 3, 4-5,
169, 174, 187
Lynch, John 171-5 *passim*,
184-88 *passim*, 195

M

- Mache, Avazeh 139, 140, 157-8
Mańczak, Witold 284
McGregor, William B. 6, 10,
254-6, 265-7, 315, 316, 317, 324,
349-50

N

- Nekes, Hermann 324, 329

O

- Ozanne-Rivierre, Françoise
180

P

- Palmer, Bill 166
Pawley, Andrew 178, 186-7, 194

R

- Ramat, Paolo 286
Rumsey, Alan L. 324, 335

S

- Sakel, Jeanette 256-7, 268
Schapper, Antoinette 5, 191
Seifart, Frank 1-2, 6, 9-10, 34,
38, 202, 203, 253-4, 264-5
Sohn, Ho-min 183
Stilo, Donald L. 3
Stokes, Bronwyn 317-25 *passim*

T

- Taylor, Peter 338
Thomason, Sarah Grey 194

V

- van den Berg, René 4-5, 187-9
van der Voort, Hein 5-6, 235-6

W

- Wichmann, Søren 84, 90
Wilson, William H. 184-5
Worms, Ernest A. 324, 329

Z

- Zavala, Roberto 80, 94, 262

Language index

A

- Adamawa languages 35, 64
- African languages 3, 33, 34, 63, 69, 202
- Afrikaans 284
- Afroasiatic languages 3, 33
- Akatek 80, 94, 262
- Aikanā 5, 204, 205, 207
- Amarakaeri, *see* Harakmbut
- Amazonian languages 5, 7, 9, 34, 201–63 *passim*
 - Northwestern 5, 9, 201, 202, 203, 204
 - Southeastern 7
 - Southwestern 5, 201–63 *passim*
- Anejoñ 179
- Apinajé 232
- Araki 179
- Aramaic, *see* Neo-Aramaic
- Araxes linguistic area 3, 135–62
- Arikapu 204, 205, 211, 214, 218, 228, 230, 232, 233, 234
- Armenian 3, 135–62 *passim*
 - Lorri 136, 159–60
 - Yerevan 136
- Aroma 172
- Australian languages 6, 242, 243, 254, 256, 272, 315–6, 318, 346, 347, 349
 - Northern 6, 242–3, 254, 315–6, 318, 346
- Austronesian languages 166, 167, 168, 174, 186–96
 - non-Austronesian, *see* Papuan
 - proto-Austronesian 174
- Avestan, *see* Old Avestan
- Azerbaijani 3, 135, 136, 138–42, 144–6, 149–52, 154–55, 160–2
- Bayadistan 141
- Azeri, *see* Azerbaijani

B

- Bafanji 36, 37, 39
- Bali-Mungaka 36–73 *passim*
- Bardi 317, 318, 319, 320, 325–8, 329, 330–1, 332, 335, 336, 338
- Bauan 170, 171, 174, 176, 178, 185
- Busa 36, 37, 39, 42, 43, 50, 57, 64, 65, 73
- Bwamu 36, 42
- Bamileke languages 53, 57, 63
 - proto-Bamileke 63
- Bantu languages 24–5, 34, 37, 38, 39, 40, 63, 68, 73
- Bantoid languages, *see* Bantu languages
- Baure 5, 6, 204–36 *passim*
- Benue-Congo languages 34–69 *passim*
- Boko 34, 64
- Bora-Miraña 9, 11, 234
- Bunuba 335, 337–9, 349
- Bunuban languages 243, 335, 348
 - proto-Bunuban 348

C

- Caddo 11, 16
- Cameroonian languages 40, 68
- Cameroonian Pidgin English 40
- Cariban languages 235
- Caspian languages 136, 138, 139, 142, 143, 160
- Cavineña 5, 204, 205, 210, 222, 226, 227–8, 230, 233
- Celtic languages 2, 4, 107, 108, 111, 114, 129
- Central Pacific languages 184
 - Proto-Central Pacific 184
- Chadic languages 35
- Chinese, *see* Mandarin Chinese
- Chiquitano 235, 236

Chuukese 4, 172–3, 178, 180, 181–3

- Ch'ol 78, 81, 82, 93, 99
- Ch'olan languages 77, 81
 - proto-Ch'olan 81, 82
- Ch'orti' 77–103
- Cornish 4, 108, 129, 130
- Cross River languages 37, 74

D

- Dampier Land languages 317, 349
- Danish 4, 283–310 *passim*
 - Old 293, 298, 303
- Denya 36, 37, 39, 40, 41, 42, 44, 45, 47, 51, 56, 57, 58, 59, 64, 65, 66, 67, 73
- Dutch 283–311 *passim*
 - Middle 293, 298, 303
 - Old 293
- Djeoromitxi 211, 228, 238, 233, 234

E

- East Asian languages 157, 242, 244, 258, 260, 270, 271, 276
 - Southeast Asian languages, *see* Mainland Southeast Asian languages
- East Sutherland Gaelic 108, 129
- Eastern Nilotic languages 34
- Ejagham 36–65 *passim*, 73
- Ekoid languages 35, 37, 73
- English 108, 130, 140, 284, 290, 297, 301, 309
- Ese Eja 228

F

- Farsi, *see* Persian
- Fijian, *see* Bauan

G

- Gaelic, *see* East Sutherland Gaelic
see also Cornish, Irish, Manx
- Gavião 213
- Gawrajui 136, 137, 142, 160
- Gban 36, 42, 44, 46, 57, 64, 73
- Georgian 3, 135, 136, 138, 142,
 158, 160, 161
- German 4, 19, 20, 25–7, 283–311
passim
 Middle High 298, 299, 302,
 304, 306
 Modern High 292, 293, 299,
 300, 302, 305
 Old High 290, 291–2, 293,
 298, 302
- Germanic languages 4, 283–310
passim
 North 284, 294, 295, 307
 Northwest 300
 proto-Germanic 284–310
passim
 West 292, 294, 295, 305, 307
- Gilaki 136, 137, 138, 142, 143, 158
- Gooniyandi 243, 254, 255, 256,
 265–8, 271, 277, 278, 317, 321,
 322, 340–1, 342, 344, 345, 349
- Gran Chaco linguistic area 235
- Grassfields languages 37, 39,
 44, 54, 55, 63, 64, 72, 73, 74, 75
 Eastern Grassfields languages
 54, 63, 64
 proto-Eastern Grassfield
 63
- Guaporé-Mamoré languages
 5, 201, 203, 211, 213, 231, 232,
 233, 236–7
- Gur languages 34–9 *passim*,
 44, 50, 57, 63, 75

H

- Harakmbut 234
- Harakmbut-Katukinan
 languages 234
- Hmong 243, 245, 246, 249–50,
 259–60, 272
- Houailou 176
- Hup 203, 243, 245, 250–1, 262

I

- Iaai 4, 180–1
- Ibibio 36
- Indo-European languages
 3, 135, 136, 285, 289, 310
 proto-Indo-European
 162, 285
- Indo-Aryan languages 269
- Iranian, *see* Persian
- Iranian languages 3, 135–56
passim, 161, 162
- Irish 4, 28, 107–31
 Cois Fhairrge 124
 Connemara 111
 Corca Dhuibhne 124
 Donegal 108
 Dublin 110
 Iorras Aithneach 110
 Munster 111
- Isu 36–74 *passim*
- Itonama 5, 204, 205, 209, 217,
 222, 229, 230

J

- Jabirrabirri 317, 318, 329, 330–1,
 332
- Jabutian languages 5, 205, 211,
 229, 231, 232, 233–4, 236
- Jacalteco 2, 11, 16, 79, 243, 244,
 245, 251–3, 262–3, 274, 275
- Jakaltek, *see* Jacalteco
- Japanese 2, 21–3
- Jarrakan languages 335
- Jawi 317
- Jukun 317

K

- Kana 36–74 *passim*
- Kanoè 5, 204–33 *passim*
- Karajá 232
- Karo 5, 204, 205, 213, 215, 222,
 230
- Kartvelian languages 3, 135, 136
- Kegboid languages 37, 52, 58, 63
- Khoisan languages 33
- Kija 335, 336, 338
- Kru languages 35, 64
- Kuliak languages 35
- Kurdish 136, 160
 Central 136, 160
 Kurmanji 160

Kurmanji, *see* Kurdish

- Kuruaya 213
- Kwa languages 35, 64, 65

L

- Lakondè 5, 204–33 *passim*
- Latundè 209, 232
- Lemerig 178
- Lolovoli 171, 178
- Löyöp 178

M

- Mainland Southeast Asian
 languages 156, 242, 244,
 245–6, 258–9, 260, 270, 271,
 276
- Makurap 213, 236
- Malayo-Polynesian languages
 4, 5, 166, 167, 174, 175, 183–6,
 187–8, 191–5 *passim*
- Central Malayo-Polynesian
 languages 166, 187, 191,
 192
- proto-Malayo-Polynesian
 167, 174
- Eastern Malayo-Polynesian
 4, 5, 189, 191, 192, 193, 194,
 195
- proto-Eastern Malayo-
 Polynesian 4, 5, 189,
 192, 193, 194, 195
- South Halmahera-
 West New Guinea
 languages 4, 166,
 187–8, 190–5 *passim*
- proto-South Halmahera-
 West New Guinea
 187–9
- Polynesian languages
 175, 183–6
- Proto-Nuclear Polynesian
 184
- Proto-Polynesian
 183, 184, 185
- Western Malayo-Polynesian
 languages 166, 167, 187
- Mamaindè 209, 214, 219,
 220–1, 226
- Manam 166, 168, 169–70, 176
- Mandarin Chinese 2, 21, 140,
 259, 270–1, 275

- Mande languages 34–73 *passim*
 Eastern 39, 61, 63, 64, 73
- Manx 4, 108, 129, 130
- Maori 184, 185
 Cook Islands 184
 New Zealand 185
- Masulei 136, 138, 142, 143, 144,
 146, 147, 166
- Mazanderani 136, 138, 139, 142,
 143, 157, 158
- Mesoamerican languages
 242, 262
- Mekens 213, 235–6
- Micronesian languages
 165, 172, 180, 181, 182, 195
- Miriwoong 335, 336, 338
- Miriwoong, *see* Miriwoong
- Mosetén 243, 254, 256–7, 265,
 268–9, 271, 277, 278–9
- Movima 5, 204–34 *passim*
- Muinane 234
- Munduruku 203, 213, 232
- Muinane 234
- Muslim Caucasian Tat 138, 142,
 143–4, 147, 148, 154
- Mussau 3, 179–80
- Mwotlap 178
- N**
- Nambikwaran languages 5,
 204, 205, 209, 218–36 *passim*
 proto-Nambikwaran 226
 Southern 209, 221, 226
 Northern 209, 210, 221, 233
- Neo-Aramaic 3, 135, 136, 142,
 159, 160
 Urmi Jewish 136, 159
 Sanandaj Jewish 136, 159,
 160
- Ngarinyin 335, 337, 339
- Ngiemboon 36–75 *passim*
- Ngumbarl 317
- Ngwe 36, 37, 44, 46, 48, 53, 57,
 59, 60, 61, 63, 69, 74
- Niger-Congo languages 3, 24,
 34–69 *passim*, 242
- Nilotic, *see* Eastern Nilotic
- Nimanburru 317, 318, 329, 330
- non-Pama-Nyungan 254, 317,
 335, 336, 348
see also Bunuban languages,
 Jarrakan languages,
 Nyulnyulan languages,
 Pama-Nyungan languages,
 Worrوران languages
- Northern Kurdish, *see* Kurmanji
- Nun languages 64
- Nyang languages 35, 37, 73
- Nyikina 317, 318, 319, 320, 323,
 324, 329, 330, 331, 332, 333, 335,
 336, 338, 341, 350
- Nyulnyul 243–78 *passim*,
 316–49 *passim*
- Nyulnyulan languages 6, 254,
 255, 324–50 *passim*
 proto-Nyulnyulan 6, 315–49
passim
 Eastern 255, 318, 320, 323,
 329, 333, 334, 335, 348
 proto-Eastern Nyulnyulan
 320, 323, 333, 334, 335, 348,
 349
 proto-Nyulnyulan 320, 323,
 333, 334, 348
 proto-Western Nyulnyulan
 320, 323, 333, 334, 346,
 347, 348
 Western 318, 320, 323, 329,
 333, 334, 346, 347, 348, 349
- O**
- Oceanic languages 3–5, 165–96
passim
 Proto-Oceanic 4–5, 165–96
passim
 non-Oceanic Austronesian
 languages 165, 168, 169,
 174, 186–7, 191–2, 193
- Obolo 36
- Old Avestan 157
- Old Norse 292
- P**
- Pama-Nyungan languages
 254, 290, 317, 335, 336, 348
see also non-Pama-Nyungan
- Pana 36, 42, 43, 50, 51, 57
- Pano 210
- Papuan languages 5, 165, 168,
 190–4, 196
- Persian 135–62 *passim*
 Classical 157
 Colloquial 135, 136, 138, 141,
 145, 148, 151, 152, 153–4
 Middle 156–7, 158
 Old 157
- Pohnpeian 183
- Polynesian languages, *see*
 Malayo-Polynesian languages
- Q**
- Q'anjob'alan languages 262
- R**
- Ramarama languages 205, 213
- S**
- Sabanê 5, 209, 210, 226–7
- Sanuma 203
- Scottish 129
- Selaru 187, 192, 193
- Sinitic languages 260
- Sirionó 233–4
- So 34
- South American languages 202
- South Halmahera-West New
 Guinea languages, *see*
 Malayo-Polynesian languages
- Southeast Ambrym 3, 178
- Southeast Asian languages, *see*
 Mainland Southeast Asian
 languages
- Swahili 22–3, 24
- Swedish 4, 28, 284–311 *passim*
 Middle 306
 Old 294, 302–3
- T**
- Talyshi dialects 136, 137–8,
 142–3, 144
- Tamabo 178
- Tat, *see* Muslim Caucasian Tat
- Tati 136, 137, 139, 142, 144, 146,
 147, 151–2, 155, 156
- Tatic languages 136–56 *passim*
- Thai 249, 259, 260
- Tobati 177

- Toussian 34–75 *passim*
 Toqabaqita 174, 177, 194
 Tsafiki 254, 271
 Tupian languages 5, 203, 204,
 205, 213, 232, 235, 236
 Turkic languages 3, 135, 136,
 138, 139, 162
 Tzeltal 79, 81, 99
 Tzotzil 82
- U
 Ubangi 35
 Ulithian 180, 183
 Urdu 269–70
- V
 Vafsi 3, 135–62 *passim*
 Vanuatu languages 3, 171, 178,
 179, 189
 Northern 178, 189
 Vietnamese 243, 246–9, 259,
 274, 275
- W
 Warrwa 317–50 *passim*
 Wayan 172
 Welsh 4, 108, 114, 129
 Witoto 234
 Worrorra 335, 337, 339
- Worroran languages 316, 324,
 335, 348, 350
 proto-Worroran 316, 348
 Wunambal 335, 337, 339
- Y
 Yanomami languages 203
 Yanomam 203
 Yanomami 203
 Yawuru 317–32 *passim*
 Yemba 36–69 *passim*
 Yucatec, *see* Yukatek
 Yukatek 99
 Yurakaré 235, 236

Subject index

A

ablauting 285–310 *passim*
abstractness 206, 287–8
accomplishment of action, *see*
telicity
adjectivity 50, 52, 58, 60, 80
see also agreement >
adjectival agreement
aggregation, *see* semantic
properties of classifiers and
class markers
agreement 4, 23–8 *passim*,
56–7, 62, 64–5, 107–31, 142,
202, 212, 230, 254, 262
adjectival 112, 117–119,
126–7, 128, 129, 130, 131
alliterative 24
article 112, 115–7, 126–7,
128–131 *passim*
clash 125
classifier 62, 202, 212, 230
gender 107–31, 202
hierarchy *see* Agreement
Hierarchy
noun class 23–4, 56–7, 62,
64–5, 202
pronoun 107, 111, 113, 119–25,
126–7, 128, 129, 130, 131
semantic 24–5, 117, 118, 122,
123, 124, 125, 126–7, 128,
129, 131
syntactic 27, 109, 117, 122,
126, 127, 128, 130
Agreement Hierarchy 27–8,
109, 128
Aktionsart, *see* telicity
allomorphy 283–311 *passim*
anaphora 1–2, 9, 11–2, 14, 15–17,
23, 26–27, 29, 80–1, 107, 108,
119, 120–1, 202, 211, 215–6,
226, 262–3, 276
see also reference tracking

animacy 4, 12–16 *passim*, 19,
24–7, 36–44 *passim*, 82–3, 88,
89, 91–5, 99, 100, 107–8, 111,
124–5, 129, 139, 148–51, 161,
247, 249, 274, 288, 303, 304,
308, 310, 311
hierarchy 150–1, 162, 274
see also humanness, semantic
properties of classifiers
and class markers
antipassives 278–9
applicatives 278–9
areal 3, 35, 135–6, 159–61, 196,
203–4, 229–31, 232–3, 269,
307, 315, 316, 348–9, 350
convergence 35, 307
diffusion, *see* feature
fade-out phenomena
135–6, 159–61
feature 3, 135, 159–61, 196,
203–4, 229–31, 232–3, 269,
315, 316, 348–9, 350
articles 107, 111, 112, 115–7,
128–31 *passim*, 145–8, 150,
152–3, 174–5, 184–5, 189
common 174–5, 185
indefinite 145–8, 150, 152–3,
160, 185
specific 184, 189
aspect 267, 271, 273, 278, 296
assimilation 306, 307
associative 33, 50–1, 52–4, 56,
58–9, 60, 62, 68–9
atomization 258, 275
attrition, *see* language attrition
augmentatives 259, 260, 261
autonomy, *see* syntactic autonomy

B

bilingualism 108, 110, 196
borrowing 5, 22–3, 28, 29, 78,
136, 138, 162, 194, 201–4 *passim*,
231, 316–321 *passim*, 349

botanical terms, *see* semantic
properties of classifiers and
class markers

C

calquing 196, 201, 329
case 112, 284–309 *passim*
accusative 288
genitive 112, 297
nominative 112, 255, 288
cataphora 119
causation 269
causative 289–92 *passim*, 348
chaining 252–3
classifiers
body part 39–42, 45–7
passim, 206, 207–12
passim, 224–8 *passim*
composite 206–11 *passim*,
227, 230
core 247, 277
general, *see* generic
generic 3, 4, 5, 12–3, 15,
23–4, 38, 42, 44, 47, 78–90
passim, 169–75 *passim*, 186,
189, 192, 202, 206, 209, 213,
221, 230, 252, 255, 256, 316,
321, 341–49 *passim*
mensural 77, 79, 80, 81, 139,
203, 244
multifunctionality of 5, 202
noun 2, 11, 16, 79, 203, 242,
243, 245, 250–3, 262–3
numeral 3, 4, 11, 33–69
passim, 77–103 *passim*,
135–62 *passim*, 165, 172–3,
178, 179, 183, 203, 209,
217, 222, 242–50 *passim*,
258–61, 274, 276
possessive 3, 35, 165–196
passim
sortal 77, 79, 80, 81, 203

- specific 202, 206, 218, 221,
252–3
- verbal 2, 6, 11, 203, 209, 217,
222, 241–4 *passim*, 254–69,
275–9
- see also* semantic properties
of classifiers and class
markers
- cliticization 49, 79, 142, 155–6,
210, 297
- collocations 10, 128–31
- see also* verbs > preverb-
inflecting verb compounds
in Northern-Australian
languages
- concord, *see* agreement
- consistency, *see* semantic
properties of classifiers and
class markers
- consonantal templates 289
- constituent order, *see* word order
- contact, *see* language contact
- contrastive focus 18, 23, 25–6,
27, 28, 242, 258, 259, 260, 275
- control, *see* subject/actor control
- countability 64, 66, 69, 80, 140,
145, 253, 258, 264, 275, 301
- D**
- declension 283–310 *passim*
- definiteness 66–7, 137, 145–8, 150,
158, 258, 259–60, 261, 262, 275
- demonstratives 175, 259, 260
- degree of activity 243, 257,
265–8 *passim*, 278, 323
- dental suffix 290, 294, 306, 307
- diffusion, *see* areal
- dimensionality, *see* semantic
properties of classifiers and
class markers
- diminutive 58, 61, 66, 73, 259,
260, 261, 287
- direction of movement, *see*
directionality
- directionality 223–4, 234, 242,
265, 266, 267, 268, 270–2,
277, 278
- dummy elements 5–6, 211, 212,
220–9 *passim*, 234
- dynamic verbs, *see* degree of
activity
- E**
- edibility, *see* semantic properties
of classifiers and class
markers
- empty roots, *see* dummy
elements
- enumeration, *see* number
- etymology 33, 34, 38, 40–1,
48–9, 53, 63, 68, 86, 174, 178,
179, 182, 195, 206, 208, 220,
224
- folk 86, 224
- transparency in 33, 34,
40–1, 48–9, 53, 195, 206
- exaptation 311
- expletives, *see* dummy elements
- extendibility of action, *see*
telicity
- F**
- fade-out phenomena, *see* areal
- fluidity 170, 171–3, 176–7, 186
- frequency 3, 4, 16, 39–40,
115–27, 128–31 *passim*, 148–51,
157, 159–60, 161, 279, 301,
305, 311, 318–322 *passim*, 332,
342–4
- cross-linguistic 3, 4, 39–40,
279
- token, 16, 115–27, 128–31
passim, 148–51, 157, 159–60,
161, 301, 305, 311, 318–322
passim, 332, 342–4
- folk etymology, *see* etymology
- fusion 33, 49, 57, 59, 69, 186,
261, 286, 287, 295, 296, 297,
298, 300
- G**
- gender, *see* grammatical gender
see also natural gender
- grammatical gender 4, 6, 11, 19,
20, 25, 27, 62, 64, 65, 107–31,
202, 203, 243, 244–5, 288, 301–
3, 304–5, 307, 308, 310, 311
see also natural gender
- grammaticalization 2, 28, 29,
34, 49, 56, 57, 60–2, 174, 195,
202–3, 206, 217, 220, 236, 241,
242–4, 246, 248, 272, 273–5,
277–9, 286, 289–90, 308–9
- H**
- hierarchy of relevance 296–7,
300, 309
- humanness 24, 42–3, 66, 73, 75,
77, 87, 88–92, 94–6, 97–100,
102, 135, 137, 139, 143–53
passim, 160, 247, 249, 303
- hyponymy 221
- I**
- immediate constituency
49, 57–8, 59
- inchoative 289, 290, 291, 292,
340, 343, 348
- incorporation 5, 185, 204–12
passim, 215–8, 220–27 *passim*
- individuation 64, 65, 66, 177,
248, 258–9, 262–4 *passim*,
275–6
- inflection 4, 111, 117, 119–20,
216, 229, 233, 283–311 *passim*
- initial mutations in Insular
Celtic 111–3, 114–30 *passim*
- K**
- kinship 95–6, 97, 169, 182–3,
183, 195
- L**
- language attrition 4
- language contact 5, 9, 28–9, 35,
65, 78, 79–80, 108, 109, 165,
168, 190–6, 231–2, 236–7, 324
- language revitalization 78, 84,
85, 89, 94–101 *passim*
- language shift 5, 194, 196
- lexical coverage 37, 38, 44, 64, 65
- lexicalization 137, 226, 227, 347
- light verbs, *see* verbs
- linguistic area 3, 135, 190, 196,
201, 235, 236
- loanword, *see* borrowing
- location, *see* semantic properties
of classifiers and class
markers
- M**
- medial 260, 261
- meronymy 206, 244, 246
- metaphors 206, 267, 308
metaphorical extension 206

- metonymy 46, 169, 308
see also synecdoche
- metatypy 196
- modality 308, 311
- mutations, *see* initial mutations
 in Insular Celtic
- N**
- natural gender 16, 19–20, 25,
 61, 259
 of speaker 259
- Natural Morphology 284
- neologisms 78, 101, 303
- nominal
- bound nominal morphemes
 in Amazonian languages
 5, 18, 201–37 *passim*
 - juxtaposition 218, 219
 - status 61, 232
see also noun class,
 classifiers > noun classifiers
- nominalization 5, 174, 195,
 204–15 *passim*, 218–9, 221, 230
- noun, *see* nominal
- noun class 11, 17–25 *passim*,
 33–8 *passim*, 44, 56, 57, 59, 61,
 63–5 *passim*, 68, 69, 202, 203,
 242–5 *passim*, 253–4, 263–5
see also classifiers > noun
 classifiers
- noun classifiers, *see* classifiers
- number 13, 22–3, 27, 36–41
passim, 42, 61–6 *passim*, 137,
 151, 156–9 *passim*, 253, 260–4
passim, 293, 298–309 *passim*
- collective, *see* singulative
- contra-expectual deficiency in
 37–41 *passim*, 66
- dual 13, 263
- plural 22–3, 42, 61–6
passim, 137, 151, 156–7,
 158–9, 260–4 *passim*, 293,
 298, 302
- singular 27, 42, 61–6 *passim*,
 157, 159, 253, 260–4 *passim*,
 293, 303, 304
- singulative 242, 258–63
passim, 275
- numerals 49–79 *passim*, 137,
 138, 144–51 *passim*, 154–61
passim, 248, 249, 254, 261
- enumeration (of nouns)
 33–69 *passim*
- 'one' 54, 55, 66, 137, 138, 145,
 254, 261
- see also* quantifiers, classifiers >
 numeral classifiers
- O**
- obligatoriness 5–6, 28, 37, 58,
 60, 128, 137–60 *passim*, 174,
 206–36 *passim*, 246–72 *passim*
- 'one function, one form'-principle,
see uniformity
- P**
- palatality 108, 110, 111, 124, 298
- periphrasis 290, 297, 308
- plural, *see* number
- pluralizability 49, 61, 62, 158,
 159
- politeness 242, 259, 261
- possession 3, 4–5, 6, 35,
 50–69 *passim*, 113, 155, 165–96
passim, 206–36 *passim*, 242,
 243, 258–64 *passim*, 275, 287,
 297
- A/O-possession in Polynesian
 languages 183–6
- active/passive 171–2, 186
- alienable/inalienable 3, 4–5,
 35, 165–196 *passim*, 224,
 235, 236, 259
- associative 50–1, 52–4, 56,
 58, 59, 60, 62, 68
- direct/indirect, *see* alienable/
 inalienable
- impersonal 6, 229, 233, 234
- irrealis 185
- obligatory/optional 209–12
passim, 217, 228, 229, 234
- see also* classifiers >
 possessive classifiers,
 pronouns > possessive
 pronouns
- preterite 289–310 *passim*
- pronouns 108–13 *passim*,
 119–25, 126–130 *passim*, 142,
 154, 155, 177, 230
- personal 14, 15, 25–8 *passim*,
 111, 119, 120–25 *passim*, 129,
 154, 155, 230
- possessive 3, 120–25 *passim*,
 142, 177
- prepositional 120–25 *passim*
- Q**
- quantifiers 52, 55, 80, 144, 148,
 151–3, 154, 218, 259, 262
- R**
- reanalysis 22–3, 56, 60, 224,
 233, 248–9, 261, 289–94
passim, 303, 316
- reciprocity 255–6, 265, 266, 321,
 322, 335, 340, 344, 346, 348
- reconstruction 3, 5, 6, 7, 38,
 63, 81, 82–3, 89, 173, 174–5,
 186–96 *passim*, 211, 232, 285,
 289–90, 315, 316, 320–1, 324,
 332–4, 340, 346–50 *passim*
- reference tracking 11–2, 15–7,
 27, 29
- reflexivity 255–6, 265–6, 275,
 321, 322, 335–48 *passim*
- register 139–40, 141, 151, 157
- relative clauses 218–9, 258,
 262–5 *passim*, 275
- repeater 21–2, 180, 181, 182,
 204, 212, 217, 236, 253, 254,
 264, 332, 347
- revitalization, *see* language
 revitalization
- S**
- scope 82, 97, 284, 302, 308
- referential 82, 97, 284
- syntagmatic 302, 308
- semantic
- category 12–3, 20, 28, 35,
 37–47 *passim*, 69, 85, 107, 111,
 122, 181, 183, 205, 242, 245,
 247, 291–2, 301–2, 303–4,
 305, 308, 310, 339–45 *passim*
 - change 46, 77–83 *passim*,
 85–103 *passim*, 284, 286
 - generality 9–29 *passim*, 90,
 206–13 *passim*, 221, 223,
 224, 244, 247, 332, 341
 - properties of classifiers and
 class markers
 aggregation 36, 37, 38, 44,
 45, 47

- botanical 38, 44, 45
 consistency 16, 217, 229, 274
 dimensionality, *see* shape
 edibility 16, 72, 186, 188
 location 217, 274
 shape 16, 18, 36, 37, 209, 217, 229, 247, 249, 263, 264, 274
 size 16, 36, 217, 274
see also animacy, humanness
 transparency 2–3, 6, 9–29 *passim*, 292, 311
see also dummy elements, agreement > semantic agreement
 shape, *see* semantic properties of classifiers and class markers
 shift, *see* language shift
 singular, *see* number
 singulative, *see* number
 size, *see* semantic properties of classifiers and class markers
 Sprachbund, *see* linguistic area
 specificity 66, 67, 102, 243, 258, 259, 262, 338, 341
see also article > specific article, semantic generality
- stative verbs, *see* degree of activity
 stem alternation, *see* ablaut
 subject/actor control 242, 266, 267, 268, 276, 278
 suppletion 301, 335, 338
 synecdoche 46
 syntactic autonomy 49, 61, 62, 278, 290, 300, 309
- T**
 telicity 242–79 *passim*, 340–1, 344
 tense 142, 216, 255, 277, 289–309 *passim*
 thematicity in proto-Indo-European lexemes 285–8 *passim*, 306
 transparency, *see* semantic transparency, transparency in etymology
- U**
 umlaut 292–301 *passim*, 309
 uniform symbolization, *see* uniformity
 uniformity 283, 299, 300, 309
 unitization, *see* individuation
- V**
 vectorial configuration, *see* directionality
 verb
 preverb-inflecting verb compounds in Northern Australian languages 6, 315–50 *passim*
 proto-compound verb constructions 324, 346, 347, 349
 light verbs 269–272
 simple verb constructions in Northern Australian languages 316, 325, 329, 332, 335, 348
 see also classifier > verbal classifier
 verbal classifier, *see* classifier
 valence 204, 208, 215–17 *passim*, 242, 265–68 *passim*, 276, 278, 289, 291, 301, 305, 308, 340, 341
 valency, *see* valence
 volitionality 269
- W**
 word order 49–59, 60, 69, 78, 80, 110, 191, 196

Classification is a popular topic in typological, descriptive and theoretical linguistics. This volume is the first to deal specifically with the diachrony of linguistic systems of classification. It comprises original papers that examine the ways in which linguistic classification systems arise, change, and dissipate in both natural circumstances and in circumstance of attrition. The role of diffusion in such processes is explored, as well as the question of what can be diffused. The volume is not restricted to nominal systems of classification, but also includes papers dealing with the less well-known phenomenon of verbal classification. Languages from a wide spread of world regions are examined, including Africa, Amazonia, Australia, Eurasia, Oceania, and Mesoamerica. The volume will be of interest to linguistic typologists, descriptive linguists, historical linguists, and grammaticalization theorists.

ISBN 978 90 272 0067 9



9 789027 200679

JOHN BENJAMINS PUBLISHING COMPANY