

Preface

About this book

The idea for this volume arose in the context of a plenary lecture by Larry M. Hyman during the ALT Summer School in Linguistic Typology in Cagliari preceding the fifth meeting of the Association for Linguistic Typology, 2003. Mentioning the unique case of “affixation by place of articulation in Tiene” (cf. Hyman, this volume), Hyman argued that there should be a more consistent interest in rarities, as a counterpart to the widely practiced pursuit of broad-scale typological generalizations. In reaction, Jan Wohlgemuth, David Gil, Orin D. Gensler and Michael Cysouw organized an international conference around the topic of *rara* and *rarissima* which was held in Leipzig from 29 March to 1 April 2006. The present volume consists of a selection out of the fifty-two papers that were presented at that conference.

For the conference we invited papers would discuss and reflect on the impact of *rara* on linguistic theory and linguistic universals. Additionally, we explicitly solicited papers dealing with the description and analysis of (apparently) rare features in individual languages. The papers in this volume are of the latter kind: They address rare phenomena attested in specific languages, how such *rara* should be analyzed and how their existence can be made more sensible. Papers dealing with the former topic are collected in the companion volume “Rethinking Universals: How rarities affect linguistic theory”, also published by Mouton de Gruyter.

Survey of this book

The papers in this book describe *rara* in individual languages, covering an extraordinarily broad geographic diversity. There are papers about languages from all over the globe, ranging from Africa (Wolof, Tiene) and Eurasia (Dutch, French, Hungarian, Georgian), over South Asia (Sinhala, Dravidian), Southeast Asia (Weining Ahmao, Abui) and Australia (Mawng), leading into North America (Aleut, Ahtna Athabascan, Hocak, Algonquian) and Mesoamerica (Totonaco), all the way to South America (Movima). Also the range of theoretical subjects discussed shows an enormous breadth, rang-

ing from phonology through word-formation, lexical semantics to syntax and even some sociolinguistics.

Starting with the phonological papers, the contributions by MARIKA BUTSKHRIKIDZE “The nature of consonant sequences in Modern Georgian” and SIRI G. TUTTLE “Syllabic obstruents in Ahtna Athabaskan” illustrate how theoretically ‘impossible’ consonant clusters come into existence. More importantly, they also show how this conflict between theoretical predictions and empirical findings can be resolved by new descriptive-analytical approaches. In contrast, “The accentual system of Hocak” is characterized by JOHANNES HELMBRECHT as a phenomenon that is cross-linguistically rare in even more ways than previously assumed.

The paper by LARRY M. HYMAN, “Affixation by place of articulation: The case of Tiene” explores a rare link between phonology and morphology, in that the phonological makeup influences (or even establishes) the order of the morphology. He shows how a ‘motivated’ construction which is considered ‘unmarked’ can still be unexpectedly rare.

The topic of *rara* in the domain of tense morphology is discussed in two papers. First, KATHARINA HAUDE discusses nominal tense in Movima in her paper “She kisses her late husband = She kissed her husband”. Differently from known cases of nominal tense, Movima is able to express the tense of the whole clause with morphological marking in the noun phrase. Second, LEENA KELKAR-STEPHAN in “Rara in French: Future tense to express habitual past or present, and past tense to express immediate future” discusses the special tense marking in French variety spoken in Pondichery, which arose in contact with Tamil.

Two papers deal with special valency phenomena. First, SYLVIE VOISIN-NOUGUIER discusses in “Possessive voice in Wolof: a rare type of valency operator” a construction that is somewhat akin to possessor-raising. However, by a comparison with other languages that have such constructions she argues that Wolof is special, and presents a diachronic explanation for this *rarum*. Likewise, RICHARD A. RHODES presents a specific Algonquian construction in his paper “Relative root complement: a unique grammatical relation in Algonquian syntax”. This phenomenon somewhat resembles an applicative construction, though there are various special aspects that make this Algonquian construction unique.

In the domain of syntax, there are various papers dealing with different *rara* related to verbal person cross-reference, or ‘agreement’. First, ANNA BERGE in the paper “Unexpected non-anaphoric marking in Aleut” shows

that the apparently strange system of so-called ‘anaphora’ in Aleut is related to textual ambiguity. Seen from that perspective, the apparent *rarum* is not that strange anymore. In a similar vein, RUTH SINGER in her paper “Mawng lexicalised agreement in typological perspective” describes the lexicalisation of otherwise productive verbal agreement morphology with special emphasis on the Australian language Mawng, but drawing on related phenomena in various other languages. Finally, GUNTHER DE VOGELAER and JOHAN VAN DER AUWERA consider the rather unusual kind of person agreement attested in some Dutch dialects in their paper “When typological *rara* generate *rarissima*: analogical extension of verbal agreement in Dutch dialects”. These dialects mark person agreement on complementizers and answer particles. They explain this *rarum* through analogical extension during grammaticalization.

Using less complex forms to express semantically more complex concepts may at first glance seem unusual, counter-intuitive or contradicting universals. Yet, there are indeed languages where marked categories are segmentally less complex. First, EIKE NITZ and SEBASTIAN NORDHOFF shed light on the origins of plural marking by segment deletion in Sinhala in their paper “Subtractive plural morphology in Sinhala”. Second, CHRISTIANE PILOT-RAICHOOR discusses the diachronic context of the morphogenesis and conceptualization of the zero negative in some of the Dravidian languages.

Finally, there is a range of papers dealing with *rara* in the realm of word-formation and lexical semantics. MARIAN KLAMER and FRANTIŠEK KRATOCHVÍL describe the fascinating verb structure of the Papuan language Abui in their paper “Abui tripartite verbs: Exploring the limits of compositionality”. They argue that very many simple CVC verbs in Abui can actually be decomposed in morpheme-like elements that are the size of one phoneme. TERESA MCFARLAND in her paper “Ideophones and templatic morphology in Totonaco” describes a somewhat similar case of an ideophonic morphophonological template associated with manner of movement, color, or odor. MATTHIAS GERNER and WALTER BISANG analyze the internal structure of the classifiers in Weining Ahmao in their paper “Social-deixis classifiers in Weining Ahmao”. They show that the choice of classifier depends on social stratification in the population. Finally, MARI UUSKÜLA and URMAS SUTROP discuss Hungarian as a suggested counterexample to color terminology universals. In their paper “The puzzle of two terms for red in Hungarian” they debunk this myth and prove Hungarian color terms actually to be a non-*rarum*.

Acknowledgments

We would like to thank Bernard Comrie and the staff at the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology, especially Claudia Schmidt and Julia Cissewski, for their support with the organization of the *Rara & Rarissima* conference and the preparation of the present volume.

For their assistance in the editing and proofreading process we wholeheartedly thank our (anonymous) reviewers and Eike Lauterbach and Orin D. Gensler. Furthermore, we are indebted to Patrick Schulz and Johannes Reese for their invaluable part of the L^AT_EX typesetting and layout. Finally, we are grateful to the admirably persistent and indulgent staff of Mouton de Gruyter, especially Ursula Kleinhenz, Julie Miess, and Wolfgang Konwitschny.

Leipzig, September 2010

JAN WOHLGEMUTH & MICHAEL CYSOUW

Contents

Preface	v
List of contributors	xi
<i>Anna Berge</i>	
Unexpected non-anaphoric marking in Aleut	1
<i>Marika Butskhrikidze</i>	
The nature of consonant sequences in Modern Georgian . . .	23
<i>Gunther De Vogelaer & Johan van der Auwera</i>	
When typological <i>rara</i> generate <i>rarissima</i> : analogical extension of verbal agreement in Dutch dialects	47
<i>Matthias Gerner & Walter Bisang</i>	
Social-deixis classifiers in Weining Ahmao	75
<i>Katharina Haude</i>	
“She kisses her late husband” = “She kissed her husband”: nominal tense in Movima	95
<i>Johannes Helmbrecht</i>	
The accentual system of Hocak	117
<i>Larry M. Hyman</i>	
Affixation by place of articulation: the case of Tiene	145
<i>Leena Kelkar-Stephan</i>	
Future tense to express habitual past or present, and past tense to express immediate future	185
<i>Marian Klamer & František Kratochvíl</i>	
Abui tripartite verbs: exploring the limits of compositionality	209
<i>Teresa McFarland</i>	
Ideophones and templatic morphology in Totonac	235
<i>Eike Nitz & Sebastian Nordhoff</i>	
Subtractive plural morphology in Sinhala	247

x *Contents*

Christiane Pilot-Raichoor

- The Dravidian zero negative: diachronic context of its
morphogenesis and conceptualisation 267

Richard A. Rhodes

- Relative root complement: a unique grammatical relation in
Algonquian syntax 305

Ruth Singer

- Mawng lexicalised agreement in typological perspective . . 325

Siri G. Tuttle

- Syllabic Obstruents in Ahtna Athabaskan 341

Mari Uuskiöla & Urmas Sutrop

- The puzzle of two terms for red in Hungarian 359

Sylvie Voisin-Nouguier

- Possessive voice in Wolof: a rare type of valency operator 377

Index 401

Subject index 401

Language index 405

List of contributors

Anna Berge

Alaska Native Language Center
University of Alaska Fairbanks
403 N. Tanana Loop Dr.
Fairbanks, AK 99775-7680
U. S. A.
ffamsb@gmail.com

Walter Bisang

Institut für Allgemeine und
Vergleichende Sprachwissenschaft
(FB 14.2)
University of Mainz
Jakob-Welder-Weg 18
55099 Mainz
GERMANY
wbisang@mail.uni-mainz.de

Marika Butskhrikidze

University FAMA
Rr. Gustav Majer nr. 7
10060 Prishtine
KOSOVA
butsmar@yahoo.com

Gunther De Vogelaer

Flemish Research Foundation
(FWO) / Ghent University
Department of Dutch Linguistics
Blandijnberg 2
9000 Ghent
BELGIUM
gunther.devogelaer@ugent.be

Matthias Gerner

Department of Chinese, Translation
and Linguistics (B 7704)
City University of Hong Kong
83 Tat Chee Avenue
Kowloon
HONG KONG
mgerner@cityu.edu.hk

Katharina Haude

CNRS (SEDYL / CELIA)
7 rue Guy Môquet
94801 Villejuif
FRANCE
haude@vjf.cnrs.fr

Johannes Helmbrecht

Institut für Information und Medien,
Sprache und Kultur (I: IMSK)
Philosophische Fakultät IV
Universität Regensburg
Universitätsstr. 31
93053 Regensburg
GERMANY
johannes.helmbrecht@sprachlit.
uni-regensburg.de

Larry M. Hyman

Department of Linguistics
Dwinelle 5218
University of California
Berkeley, CA 94720-2650
U. S. A.
hyman@berkeley.edu

Leena Kelkar-Stephan

Adolfstraße 4
06114 Halle (Saale)
GERMANY
leenaks@t-online.de

Marian Klamer

Leiden University Centre for Linguistics (LUCL)
Faculty of Humanities
Leiden University
P. O. Box 9515
2300 RA Leiden
THE NETHERLANDS
M.A.F.Klamer@hum.leidenuniv.nl

František Kratochvíl

School of Humanities and Social Sciences
Division of Linguistics and Multilingual Studies
Nanyang Technological University
HSS Level 3, Room 53
14 Nanyang Drive
Singapore 637332
SINGAPORE
fkratochvil@ntu.edu.sg

Teresa McFarland

Department of Spanish and Portuguese
Dwinelle 5218
University of California, Berkeley
Berkeley, CA 94720-2650
U. S. A.
terefar@berkeley.edu

Eike Nitz

Universität zu Köln
Institut für Indologie und Tamilistik
Pohligstraße 1
50969 Köln
GERMANY
suddhek@freenet.de

Sebastian Nordhoff

Department of Linguistics
Max Planck Institute for Evolutionary Anthropology
Deutscher Platz 6
04103 Leipzig
GERMANY
sebastian_nordhoff@eva.mpg.de

Christiane Pilot-Raichoor

Lacito – UMR 7107 – CNRS
7, rue Guy Môquet
94800 Villejuif
FRANCE
raichoor@vjf.cnrs.fr

Richard A. Rhodes

Department of Linguistics
University of California
Dwinelle 5218
Berkeley, CA 94720-2650
U. S. A.
rrhodes@berkeley.edu

Ruth Singer

Linguistics program
School of Languages and Linguistics
University of Melbourne
Parkville 3010
Victoria
AUSTRALIA
ruth.singer@gmail.com

Urmas Sutrop

Institute of the Estonian Language
Roosikrantsi 6
Tallinn 10119
ESTONIA
urmas.sutrop@eki.ee

Siri G. Tuttle

Alaska Native Language Center
University of Alaska Fairbanks
P. O. Box 757680
Fairbanks, AK 99775
U. S. A.
sgtuttle@alaska.edu

Mari Uusküla

Institute of the Estonian Language
Roosikrantsi 6
Tallinn 10119
ESTONIA
mari.uuskyla@eki.ee

Johan van der Auwera

Universiteit Antwerpen
Centre for Grammar, Cognition and
Typology
Prinsstraat 13
2000 Antwerp
BELGIUM
johan.vanderauwera@ua.ac.be

Sylvie Voisin-Nouguier

UMR 5596 – Dynamique du Langage
14 avenue Berthelot
69363 Lyon cedex 07
FRANCE
Sylvie.Voisin@univ-provence.fr

Editors

Michael Cysouw

Department of Linguistics
Max Planck Institute for Evolutionary
Anthropology
Deutscher Platz 6
04103 Leipzig
GERMANY
cysouw@eva.mpg.de

Jan Wohlgemuth

Department of Linguistics
Max Planck Institute for Evolutionary
Anthropology
Deutscher Platz 6
04103 Leipzig
GERMANY
wohlgemuth@eva.mpg.de

Unexpected non-anaphoric marking in Aleut

Anna Berge

1 Introduction

Aleut is the language historically spoken along the Aleutian Chain and on the Pribilof Islands off the coast of Alaska, and on the Commander Islands east of Kamchatka in the Russian Far East. It is the sole language in its branch of the Eskimo-Aleut language family although a number of distinct dialects are distinguished. Prehistoric developments in its case system have resulted in something typologically unusual and very different from the ergative-absolutive system found in Eskimo languages and assumed for Proto-Eskimo-Aleut; these developments profoundly affected not only the nominal but also the verbal inflectional system. The latter, for example, indexes subjects, and it can index objects, but the circumstances under which it does so are odd: a verb marks an unexpressed third person object (whether direct or oblique) or an unexpressed possessor, and the number of an unexpressed subject or object term or possessor thereof. Although documentation of this system, known as Aleut anaphora, has existed since the early 19th century, it is Knut Bergsland's work in particular that has brought the system some attention. From the 1960's on, this aspect of Aleut grammar has been studied more than any other, as is evident in the works of Bergsland (especially Bergsland 1997), Leer (1987, 1991), Fortescue (1985), and Sadock (2000). The focus of work on Aleut anaphora has been on where anaphora is used; on the link between anaphora and the absence of an independent way of expressing pronominal 3rd person; and on the independence of number marking and anaphoric marking. Other important aspects of the system have escaped attention, including common instances of what Bergsland calls 'zero-anaphora', by which he means the non-indexing of some unexpressed, anaphoric argument or possessor. In Bergsland's work as well as in extended Aleut texts, however, zero-anaphora is as common as anaphoric marking; furthermore, the canonical anaphoric marking that has become so well-known in the literature is canonical only in simple or decontextualized clauses.

In this paper, I will show that context is indispensable in the interpretation of anaphoric marking or lack thereof in Aleut: without taking context into ac-

count, we can do little more than describe the various environments in which anaphoric inflection is found. Aleut anaphora is part of a reference-tracking system which manages textual rather than clausal ambiguity. The more ambiguity there is in a text, the more likely there is to be anaphoric marking. Conversely, lack of anaphoric marking is related to the relative salience of the arguments: the more recoverable something is, the less likely it is to be tracked with anaphoric marking. In effect, anaphora and zero-anaphora must be explained together. While the particulars of the anaphoric system are unusual, the principle is not: what at first appears to be typologically unusual marking turns out to be a variant of topic-tracking strategies common in clause-chaining languages.

In Section 2, I will present a summary of the system of Aleut anaphora as discussed in the literature; in Section 3, I will present instances of zero-anaphora as discussed in Bergsland (1997); and in Section 4, with evidence from texts, I will show how context, referential ambiguity, and topic salience are crucial in the understanding of how this system works.¹

2 Canonical anaphoric reference marking in Aleut: summary of previous work

Some of the main features of Aleut nominal and verbal inflection discussed in the literature are presented below, and examples in Section 2, unless otherwise indicated, are common in the literature. Although I mostly refer to Bergsland (1997) as a primary source of Aleut data, other authors have made very significant attempts to explain the anaphoric system of Aleut.

2.1 Marking of dependency: absolutive, relative, and possessive inflection

There are several sets of inflectional endings, known respectively as the ‘absolutive,’ ‘relative,’ and ‘possessive,’ endings, and they are found on both nouns and verbs (as well as oblique endings found on both postpositions and some subordinate verb moods which will not be discussed in this paper). Although the endings do not indicate case or possession on verbs, Bergsland (1997) confusingly refers to verb endings by these terms, which are normally understood to refer to nominal functions. By doing so, he points to the identity of the inflectional morphemes regardless of the part of speech to which they attach. In this paper, I will use these terms in reference to nominal inflection only, and the terms ‘simple,’ ‘relational,’ ‘dependent,’ and ‘anaphoric’ for verbal inflection. Although they have not been explicitly analyzed as such, I

interpret these inflectional endings as markers of independence and dependence, or better, as markers indicating relational status. Thus, absolutive endings are found on nouns which are not in some kind of phrasal construction, and the corresponding inflectional ending on verbs are found in independent moods especially, as in example (1):

- (1) *Piitra-â asxinu-â kidu-ku-â*
 Peter-ABS girl-ABS help-IND-3SG
 ‘Peter is helping the girl’ (Atkan)

Note that *-â* on nouns is glossed as ‘absolutive’;² on verbs, it is traditionally glossed simply as ‘3SG’. In fact, it is the default verbal inflectional ending on all persons in the singular, and first and second persons have enclitic pronouns in addition to *-â*. Verbal endings which are morphologically identical to the absolutive nominal inflections will be referred to as simple verbal inflections.

Relative endings are routinely found on nouns in phrasal constructions such as possessive phrases and postpositional phrases, as in example (2), and on verbs in clauses with some dependency on neighboring clauses, as in examples (3) and (4). In example (4), the use of the indicative mood with the relative ending rather than the conjunctive mood shows that the first clause is not being negated along with the second (Bergsland 1997: 221). The relative ending on both nouns and verbs is traditionally glossed as ‘relative.’ In verbs, it is not entirely predictable from current descriptions, although it tends to be found in clause combinations which show succession or contrast (Bergsland 1997: 252) or which indicate irrealis (in clauses I have collected); in texts it is somewhat rare, and this ending, which we could refer to as ‘relational’ verbal inflection, will not be further discussed here:

- (2) *Piitra-m ada-a asxinu-â kidu-ku-â*
 Peter-REL father-3SG.POS.ABS girl-ABS help-IND-3SG
 ‘Peter’s father is helping the girl’ (Atkan)
- (3) *Sa-â uchigi-hli-ku-m txin îga-s îga-na-â*
 duck-ABS swim-until-IND-3SG.REL self be.scared-CONJ take.off-PART-3SG
 ‘the duck was swimming until he got scared and took off’ (Atkan; Bergsland 1997: 244)
- (4) *Mariiya-â unguchi-ku-m qa-na-ġ-ulax*
 Mary-ABS sit.down-IND-3SG.REL eat-PART-3SG-NEG
 ‘Mary sat down but did not eat’ (Atkan; Bergsland 1997: 221)

Possessive endings are routinely found on possessed nouns, as in example (2), repeated below as example (5), where the ending is glossed as ‘possessive’, and on verbs with certain of the subordinate moods, as in example (6). In subordinate, or dependent moods, these verbal endings indicate simple person and number, and they are glossed accordingly; however, they must be distinguished from the same morphological forms which are also used to indicate anaphora in moods which otherwise take simple or relational inflections. I will refer to these endings as ‘dependent’ endings:

- (5) *Piitra-m ada-a asxinu-â kidu-ku-â*
 Peter-REL father-3SG.POS.ABS girl-ABS help-IND-3SG
 ‘Peter’s father is helping the girl’ (Atkan)
- (6) *Piitra-â asxinu-â kidu-gu-u . .*
 Peter-ABS girl-ABS help-COND-3SG
 ‘If / when Peter helps the girl. . .’ (Atkan)

2.2 Marking of anaphoric elements

The system just described is complicated by several overlapping but nevertheless separate systems of marking, including most famously the marking of non-subject anaphoric elements. When all arguments, both grammatical and oblique, are expressed, as in example (1) on the preceding page, repeated here as example (7), absolutive nominal inflection and simple verb inflection are used. Because absolutive is used for both subject and object, Aleut is a relatively strict SOV language:

- (7) *Piitra-â asxinu-â kidu-ku-â*
 Peter-ABS girl-ABS help-IND-3SG
 ‘Peter is helping the girl’ (Atkan)

If the subject is unexpressed, but the object is overt, as in example (8), there is no change:

- (8) *asxinu-â kidu-ku-â*
 girl-ABS help-IND-3SG
 ‘He is helping the girl’ (Atkan)

When a non-subject argument is unexpressed, anaphoric reference is made to it through the use of relative marking on the subject and what looks like dependent marking on the verb, as in examples (9) and (10):

- (9) *Piitra-m kidu-ku-u*
 Peter-REL help-IND-3SG.AN
 ‘Peter is helping her’ (Atkan)

- (10) *Kidu-ku-u*
 help-IND-3SG.AN
 ‘He is helping her’ (Atkan)

In this case, the inflectional ending on the verb is glossed as before according to its person and number, but additionally, it is indicated as being anaphoric. In fact, it is not morphologically distinct from the non-anaphoric verbal inflection in dependent moods; however, it must be functionally distinguished from dependent marking. Following existing convention, I will refer to these as the ‘anaphoric’ verbal endings. Anaphoric and dependent inflection are essentially in complementary distribution.

This system of anaphoric reference extends to oblique objects, as can be seen in examples (11) and (12), and to the unexpressed possessors of objects, as in examples (13) and (14):³

- (11) *qalgada-â stuuluġi-m kug-an a-ku-â*
 food-ABS table-REL top-LOC be-IND-3SG
 ‘The food is on the table’
- (12) *qalgada-m kug-an a-ku-u*
 food-REL top-LOC be-IND-3SG.AN
 ‘The food is on it’ (Atkan)
- (13) *Liidiya-â Piitra-m ada-a kidu-ku-â*
 Lydia-ABS Peter-REL father-3SG.POS.ABS help-IND-3SG
 ‘Lydia is helping Peter’s father’ (Atkan)
- (14) *Liidiya-m ada-a kidu-ku-u*
 Lydia-REL father-3SG.POS.ABS help-IND-3SG.AN
 ‘Lydia is helping his father’ (Atkan)

2.3 Marking of anaphoric number

Another complicating factor in the use of the inflectional types (absolutive, relative, and possessive on nouns, or simple, relational, dependent, and anaphoric on verbs) is the marking of plurality, which is characterized by anaphoric marking of the number of an unexpressed argument (subject, object, or oblique) or possessor. This is illustrated in examples (15) and (16). In the

former, all arguments are overt, while in example (16), the plural subject is overt but the singular object is not and is therefore indexed on the verb with singular anaphoric inflection:

- (15) *asxinu-s Piitra-â kidu-ku-s*
 girl-PL Peter-ABS help-IND-3PL
 ‘The girls are helping Peter’ (Atkan)

- (16) *asxinu-s kidu-ku-u*
 girl-PL help-IND-3SG.AN
 ‘The girls are helping him’ (Atkan)

In cases of multiple anaphora with a conflict in number, plural number seems to take precedence. Thus, in example (17), assuming it is semantically related to examples (15) and (16),⁴ both the plural subject and the singular object are unexpressed; this triggers plural anaphoric marking on the verb, despite the singular object:

- (17) *kidu-ku-ngis*
 help-IND-3PL.AN
 ‘They are helping him’

Similarly, with a plural anaphoric possessor, as in example (18), the verb is marked with plural inflection, despite the singular subject. It is not marked for anaphora because there is no object:

- (18) *ada-ngis awa-ku-s*
 father-3PL.POS.ABS work-IND-3PL
 ‘Their father is working’ (Atkan)

To the extent possible, anaphoric marking captures what is otherwise unrecoverable. In example (19), the verbal inflection is singular and anaphoric because of the 3SG unexpressed possessor of the object:

- (19) *ada-a kidu-ku-ng*
 father-3SG.POS.ABS help-IND-1SG/3SG.AN
 ‘I am helping her father’ (Atkan)

In example (20), the verbal inflection is singular and anaphoric, despite the plural object, because the anaphoric possessor is singular.

- (20) *huyu-ngis kidu-ku-ng*
 brother-3PL.POS.ABS help-IND-1SG/3SG.AN
 ‘I am helping her brothers’ (Atkan)

Leer (1991) calls this system ‘promiscuous number marking’ and suggests that this is the result of areal borrowing between languages in the region of Southeastern Alaska. For a more complete discussion of the oddities of the number marking system, see Sadock (2000).

2.4 Marking of focused elements

Elements which are in some way focused may be marked as if they were external or independent of the clause marking described above. Focused elements of all types may be found, with concomittent effects on the syntax. Thus, in example (21), the focused object is preposed, and therefore grammatically outside of the SOV clause, and takes absolutive case. The clause is marked as if the object were anaphoric:

- (21) *asxinu-s Piitra-m kidu-ku-ngis*
 girl-PL Peter-REL help-IND-3PL.AN
 ‘The girls, Peter is helping them’ (Atkan)

3 Zero-anaphora, or unexpected non-canonical marking (Bergsland)

The discussion up to now essentially summarizes the elements of Aleut reference marking which have been most famously discussed in the literature. Other important aspects of the system have escaped wider attention, including common instances of what Bergsland calls ‘zero-anaphora’. By ‘anaphora’, Bergsland does not simply refer to the typologically unusual anaphoric system of marking as explained above; he refers to any instance in which reference is made to anaphoric information, including, for example, subject agreement on the verb. When he uses the term ‘zero-anaphora,’ he refers to any instance in which no anaphoric marking is used, whether this be in reference to common subject agreement or to the less common anaphoric non-subject or number agreement special to Aleut. In the discussion that follows, I limit myself to the latter.

Under the system just presented, examples (22) and (23) show seemingly regular anaphoric marking: in example (22), all noun phrases are overt, whereas in example (23), the unexpressed oblique object ‘Peter’ is anaphorically marked on the verb:

- (22) *Piitra-â qa-ku-â*
 Peter-ABS eat-IND-3SG
 ‘While Peter was eating,’

tayaġu-ŝ Piitra-m naga-n qangu-na-ŝ
 man-ABS Peter-REL into-LOC come.in-PART-3SG

‘a man came in to Peter’ (Atkan; Bergsland 1997: 254)

- (23) *Piitra-ŝ qa-ku-ŝ*
 Peter-ABS eat-IND-3SG

‘While Peter was eating,’

tayaġu-m naga-n qangu-qa-a
 man-REL into-LOC come.in-PART.AN-3SG.AN

‘a man came in to him’ (Atkan; Bergsland 1997: 254)

However, this is only one of a number of strategies for identifying participants in clause combinations; there are a number of variations of anaphoric marking, as well as zero-anaphora, which turns out to be almost as common as expected anaphoric marking. The system presented so far is valid primarily for simple and decontextualized sentences. In almost any instance of clause combination, the anaphoric system may be manipulated in ways which differ from the commonly understood rules of Aleut anaphora presented in Section 2. Examples (24) through (30) illustrate some of these unusual uses of anaphora or instances of zero-anaphora, and examples (31) and (32) reexamine the meanings of examples (22) and (23). Bergsland (1997: 253) suggests the choice of anaphora or zero-anaphora in these cases is determined by some level of coreferentiality. While some of the examples are pragmatically understandable, many are much less clear.

In simple clause combinations, if two clauses have a coreferential object, and this object is overt in the first clause, but anaphoric in the second, the verbs may simply not be marked for anaphora; the object in either clause may be direct or oblique. This is especially, but not exclusively, true if the first clause is in the conjunctive mood, as in example (24):

- (24) *ulu-ŝ una-l qa-ku-s*
 meat-ABS cook-CONJ eat-IND-1PL

‘We cooked the meat and ate it’ (Atkan; Bergsland 1997: 222)

(one would have expected *uluŝ unal qakungis*)

The conjunctive mood commonly links subordinate clauses with sequential or cotemporal semantics in a chain, tends to indicate coreferential subjects between clauses, and tends to have defective conjugation. The lack of anaphoric marking, therefore, could be attributed to the features of this particular mood; however zero-anaphora is also found between clauses with other moods, and

conjunctive mood is used in clauses with non-coreferential subjects. Conjunctive mood cannot therefore be the sole factor in zero-anaphora in (24).

We also find the same sorts of clause combinations with anaphoric marking, although the anaphora may not follow the expected pattern. Example (25) is similar to example (24) in having an overt object in the first clause; however, in this case, the first clause has anaphoric marking, despite its overt object, while the second clause does not.⁵ Bergsland explains anaphora here as marking the coreferential nature of the subject and object of both clauses:

- (25) *hla-s ilga-ku-ning uku-na-ġula-q*
 boy-PL look.for-IND-1SG/3PL.AN find-PART-NEG-1SG
 ‘I looked for the boys but did not find them’ (Atkan; Bergsland 1997: 251) (one would have expected *hlas ilgakuq ukuqaningulax*)

If the object is unexpressed in both clauses, then both clauses may be marked for anaphora, as in examples (26) and (27); this last includes a conjunctive clause (which does not need anaphoric marking even in the expected pattern of anaphoric use, cf. examples (36) and (37), and see Bergsland 1997: 223):

- (26) *ilga-ku-ning uku-qa-ning-ulax*
 look.for-IND-1SG/3PL.AN find-PART.AN-1SG/3PL.AN-NEG
 ‘I looked for them but did not find them’ (Atkan; Bergsland 1997: 251)
- (27) *ama-an ayaga-qan ngaan slaaġa-kan*
 DEM.INV-ABS wife-3SG.POS.REL 3SG.ABL come.out-CONJ.AN

aaluusa-qali-ku-u awa
 laugh.at-begin-IND-3SG.AN DEM.PROX
 ‘His wife came out to him and began to laugh at him’ (Eastern, Bergsland 1997: 223)

One might be led to suppose that zero-anaphora (and other anaphoric marking patterns) reflects coreferentiality of objects (and / or adjuncts of objects, cf. example (33)), at least one of which is overt, in clause combination; but this is not the case. Zero-anaphora is also used as a strategy when the overt subject of the first clause is the unexpressed object of the second, as in (28):

- (28) *aniqdu-ġ hit-naaġi-ku-ġ hiti-chġi-na-ġula-q*
 child-ABS go.out-try-IND-3SG go.out-let-PART-NEG-1SG
 ‘The child tried to go out but I didn’t let it go out’ (Atkan; Bergsland 1997: 254)
 (one would have expected *aniqduġ hitnaaġikuġ, hitichġiqangulax*)

At the same time, anaphoric marking on a verb may be used in reference to the non-overt subject of the previous clause. In example (29), the subject of the first clause is overt, and there is no anaphoric marking on either verb. In example (30), however, the first clause's subject is not expressed independently (it is reflected in the verbal inflection) — and the verb of the second clause receives anaphoric marking. Once again, anaphoric marking does not show the expected pattern:

- (29) *hla-s tunum-kada-ku-z-iin ting saĝani-na-q*
 boy-PL talk-CESS-IND-PL-ENCL 1SG go.to.sleep-PART-1SG
 'When the boys stopped talking, I went to sleep' (Atkan)
- (30) *tunum-kada-ku-z-iin ting saĝani-qa-ning*
 talk-CESS-IND-PL-ENCL 1SG go.to.sleep-PART.AN-1SG/3PL.AN
 'When they stopped talking, I went to sleep' (Atkan; Bergsland 1997: 248) (one would have expected *tunumkadakuziin, ting saĝaninaq*)

From these examples, it is also obvious that anaphoric marking is not directly related to transitivity.⁶ Anaphoric marking is not limited to transitive verbs, and transitive verbs with unexpressed objects do not always indicate the objects in any way.

The interplay between anaphora and zero-anaphora seems to involve some recursivity in the application of rules of anaphoric marking, as in example (23), repeated here as example (31), and examples (32) through (34). In examples (31) and (32), the oblique object in the second clause is unexpressed, which should trigger the use of anaphora in that clause. When the subject of the second clause is overt, as in example (31), there is anaphoric marking, and when the subject is not overt, as in example (32), there is zero-anaphora:

- (31) *Piitra-â qa-ku-â*
 Peter-ABS eat-IND-3SG
 'While Peter_i was eating,'
tayaĝu-m naga-n qangu-qa-a
 man-REL into-LOC come.in-PART.AN-3SG.AN
 'a man came in to him_i' (Atkan; Bergsland 1997: 254)
- (32) *Piitra-â qa-ku-â naga-n qangu-na-â*
 Peter-ABS eat-IND-3SG into-LOC come.in-PART-3SG
 'While Peter_i was eating, he_j came in to him_i' (Atkan; mod. from Bergsland 1997: 254)⁷
 (one would have expected *Piitraâ qakuâ nagan qanguqaa*)

In each of the following two sentences, the second clause has no overt subject. In example (33), the possessor in the first clause is coreferent with the subject of the second clause, while in example (34), the two are noncoreferent:

- (33) *(Piitra-m) ada-a qa-ku-â*
 (Peter-REL) father-3SG.POS.ABS eat-IND-3SG
 ‘While (Peter_i)’s father_j was eating,’
naga-n qangu-na-â
 into-LOC come.in-PART-3SG
 ‘he_i came in to him_j’ (Atkan; Bergsland 1997: 254)
 (one would have expected *nagan qanguqaa*)
- (34) *ada-a qa-ku-â*
 father-3SG.POS.ABS eat-IND-3SG
 ‘While his_j father was eating,’
(Piitra-m) naga-n qangu-qa-a
 (Peter-REL) into-LOC come.in-PART.AN-3SG.AN
 ‘he_i (Peter) came in to him_j’ (Atkan; Bergsland 1997: 254)

In examples (32) through (34), two clauses have involved different subjects without overt indications of switch-reference, unless the manipulation of anaphoric marking in the second clause is taken to show changes in subject. However, Aleut does in fact have inflectional marking to distinguish same and different subjects, so this kind of ambiguity should not have to arise. It is also possible for the anaphoric object or oblique in a first clause to be the subject of a second clause, as in example (35) (this should not be misconstrued as evidence of ergative syntax):

- (35) *Piitra-m tuga-ku-u qida-qali-na-â*
 Peter-REL hit-IND-3SG.AN cry-start-PART-3SG
 ‘Peter hit him_j and he_j started to cry’ (Bergsland 1997: 257–258)

These examples have been provided by Bergsland to show how anaphoric marking and zero-anaphora may work in various types of clause combinations, with and without coreferential objects, obliques, or subjects; however, the rules regarding the marking of coreference are not obvious from the examples given above, and coreference does not appear to be a defining factor in the use or lack thereof of anaphora, as we see from examples (30) and (35). Further, the understanding of what is coreferential would need to be broadened to account for coreference with oblique objects or with possessors, and

for coreference beyond immediately neighboring clauses. Bergsland rarely makes categorical claims about the use of anaphora or lack thereof in clause combinations; in general, he simply lists all possible environments in which a particular structure is found. Most examples above could be countered with examples of perfectly ‘regular’ use of anaphoric marking.

4 Explaining zero-anaphora (and anaphoric marking) in reference to context

In fact, there is a simple explanation of anaphora and zero-anaphora when context is taken into account. Aleut, like the related Eskimo languages, is a clause chaining language, and within a clause chain, referential marking is best explained with respect to the topic being tracked, where topic is understood to be the prominent entity, perhaps restricted to participants, across a stretch of continuous discourse, and not specifically to subjects or objects between two clauses (see Berge 1997).⁸

In the examples of zero-anaphora above, ‘unexpected’ non-anaphoric marking is found with apparent topic continuity, as in example (32) in which the participants are both given (and also definite) as opposed to example (31), in which the overt new participant in the second clause coincides with anaphoric marking. Zero-anaphora is possible because of a high level of tolerance within Aleut for referential ambiguity in individual clauses within a greater text, and it is left to the context to adequately disambiguate the various arguments and possibilities of interpretation. As long as the topic of a clause chain remains unchanged, and there are no significant new participants introduced within the chain, there is no need for special anaphoric inflection to refer to unexpressed objects or possessors, regardless of actual changes in grammatical role of these participants (e. g. subject, object, or oblique): zero-anaphora signals topic continuity.

However, even this tolerance has its limits. In texts, we consistently see anaphora used to mark some nominal information being referred to without being expressed, and which there is perhaps no other way to recover. Anaphoric marking is found where ambiguity would lead to an obviously false interpretation or no possible interpretation. Furthermore, when anaphora is used for disambiguation, it is to disambiguate which of the participants may be considered topical: the referent of the anaphora is typically topical, whereas, for example, the subjects of anaphoric structures are typically not topics (this is perhaps the case in examples (33) through (35), and seems to explain the

switch of subjects in example (35); see Berge (2009) for more recent work on anaphora and topic in Aleut). The link between anaphoric marking and topicality has been made previously by Fortescue (1985), using a rather different understanding of the concepts of topic and theme, where ‘topic’ is a constituent within the predication, and theme is one outside of the predication. Fortescue limits his discussion to examples of anaphoric agreement such as those in Section 2, and to simple clauses or simple clause combinations; his definition of topic is ultimately clause-dependent; and in offering a strictly rule-governed analysis, he finds it difficult to account for the very wide range of acceptable syntactic structures. However, his conclusions, that anaphoric marking is related to some pragmatic need ‘outside the predication proper’ (Fortescue 1985: 121), support the ideas I am proposing here.

Since most examples from Bergsland (1997) are decontextualized in his presentation of the grammar, this explanation is not immediately apparent from the examples given so far. However, this interpretation is supported by examples of anaphora or lack thereof in context, as in examples (36) through (38) below.⁹

Example (36) illustrates the use of anaphora for disambiguation. The narrator describes her experiences on the way to an internment camp during World War II. The topic of this text is her mother. In the second last line, there is an anaphoric conjunctive, despite no previous overt identification of the subject ‘we’, referring to the narrator and someone else, and a following non-anaphoric conjunctive, despite the same arguments and an argument structure which is understood to be similar (3SG passive in Eastern dialect is used for 1PL active). In fact, one could have expected the passive form of the first conjunctive, *suġalix* ‘we took [her]’ (or ‘she was taken’), structurally the same as the second conjunctive, *kumsiġalix*. However, if a non-anaphoric form were used, it would be interpreted as the mother taking (something):

- (36) *Aang maama-ng aygag-duuka-lakan*
 yes mother-1SG walk-PROB-CONJ.NEG
 ‘Yes, my mother could not really walk’
alima-x txin umlikaada-asa-lix
 leg-DU 3SG.OBJ be.sick-with-CONJ
 ‘Her legs made her sick,’
atxaŋ-six txin ayug-ni-duuka-lakaŋ
 be.straight-CONJ 3SG.OBJ move-CAUS-PROB-IND.NEG.3SG
 ‘she couldn’t move them straight’

su-âta-kan *agacha quchxi-ngin* *kumsi-âa-lix*
 take-CONT-CONJ.SG.AN only between-3PL.POS.ABS lift-PASS-CONJ
 ‘sometimes holding her up between us,’

agach[a] ayug-ni-
 only move-cause-
 ‘we used to walk her’

(Eastern, 1984; transcribed by Dirks in 1984 and 2004) (the final verb was not completed by the speaker)

It is useful to compare the use of anaphora and zero-anaphora in analogous structures, although it is not always possible in the same text. Examples (36) and (37) do belong to the same text, and both involve the same verb, *su-* ‘to take’, and the same mood, the conjunctive. In example (37), the mother, still the topic, has been scared by some army planes, left the house, and hidden under a kayak with a young girl; her daughter and a friend are searching for her. *Adan kimkadangin* anaphorically refers to an unexpressed complement; in traditional descriptions of Aleut grammar, it could refer both to the kayak under which the mother is hiding (anaphoric reference to the object of the postposition) and to the mother and young girl (with the use of the plural form of the anaphoric inflection); but the verb, *su-* ‘take’, has zero-anaphora, referring to the same arguments. Aleut minimizes its referential specificity: neither the subject nor the object are necessarily made explicit via verbal inflection; nevertheless, the meaning of *su-âa-lix* is clear enough in context:

- (37) *Eva agiita-lix* *ada-n kim-kada-ngin* *su-âa-lix*
 Eva together.with-CONJ to-LOC go.down-CESS-[1,3]PL.AN take-PASS-CONJ
 ‘Together with Eva, after [we] had gone down to [it / them], [we] took [her / them]’

ala-âu-m achida-n a-qa-a *a-âta-na-â*
 ocean-REL shore-LOC be-PART.AN-3SG.AN AUX-apparently-PART-3SG
 ‘she had apparently been down to the shore,’

i-âata-âsi-ida-lix
 fear-make-DIM-CONJ

‘becoming scared, the poor thing’
 (Eastern, 1984; transcribed by Dirks 1984 and 2004)

If anaphoric marking is used for disambiguation, zero-anaphora is used for known or recoverable information, rather than coreference, and for topic

continuity rather than the disambiguation of possibly competing topics. Example (38) is a very simple example of zero-anaphora with topic continuity across extended clauses. Here, there is in fact no anaphora at all, despite the consistent use of the valency-increasing suffix *-usa-*, as in *idgiġtusa-* ‘to pull (fish) up’, or *akuunusa-* ‘to take something up’, or especially in the final clause, *aġaasa-* ‘to place something’. This is not just a more complex version of example (24) on page 8, in which coreference between an expressed object of one clause and an unexpressed object of another eliminates the need for anaphoric marking. In this case, *qaġ*, ‘fish’, is originally introduced as a subject; it is never expressed as an object, no anaphoric reference is made to it, and there is an intervening clause with no implied reference to it. No anaphoric marking is present because there is no potential ambiguity here.

- (38) *chamchux-six ali-ġta-ku-qing-aan,*
 fish.from.land-CONJ still.be-CONT-IND-1SG-ENCL
 ‘When I’d been fishing for a while,
- qa-ġ duġta-ng adġa-ku-ġ-aan ting amani-lix*
 fish-ABS hook-1SG.POS.ABS bite-IND-3SG-ENCL 1SG startle-CONJ
 ‘a fish bit my hook and startled me.’
- chamchxi-ng qangli-ng kanga-n aġ-six*
 fish.line-1SG.POS.ABS shoulder-1SG.POS.ABS top-LOC put-CONJ
 ‘Putting my fish line on my shoulder’
- nung idgi-ġtusa-lix akuunusa-lix*
 to.1SG pull.out-INTNS-CONJ take.up-CONJ
 ‘Pulling [it] up to me, I took [it] up’
- chuguulġu-n ku-ngin aġa-asa-ku-qing-aan*
 gravel-PL on.top-LOC.PL put-with-IND-1SG-ENCL
 ‘I placed [it] on the gravel / beach.’
 (one would have expected *aġaasakungaan* ‘I put it’)
 (Eastern, 1984; transcribed by Dirks in 1984 and 2004)

In Example (39), there is too much ambiguity in the reference to the participants. The text concerns help given to a starving youth, and there are at least three participants (and possibly a fourth), the youth, Albert, and the person who went to fetch Albert. There are several instances of switches in subjects; it appears that the fourth person form *igiim* ‘to him (4SG)’ refers exclusively to the youth, despite the fact that he has not been mentioned

since the beginning of the narration (some 17 or 18 lines previously); he is reintroduced anaphorically in the fourth line through *ngaan* ‘to him (3SG)’.

The marking of anaphora or lack thereof in these clauses differs from descriptions of Aleut anaphora in a number of ways. One would expect *aslixsix* in the second clause to be marked for anaphora, since the following clause is. In the fifth line, the clause contains an anaphoric form of the conjunctive, *nuusakan*; the next clause contains no anaphora, despite the fact that *igiim* refers to the still unspecified youth. In this text, the use of anaphora is found in clauses referring to three participants, at least two of whom are unexpressed. In fact, it typically seems to refer to Albert. Once again, therefore, anaphora is used to disambiguate; zero-anaphora is used where the reference is clear (e. g. through 4th person marking as in *igiim*).

- (39) ... *Inga-ya ama-agan aygax-six, Canneri-â u-lix*
 DEM.PROX-DEIC DEM.INV-ABL walk-CONJ cannery-ABS reach-CONJ
 ‘... That one, [he] walked from there, going to the cannery’
- a-ku-m ama-ya ama-an Albert aslix-six*
 be-IND-REL DEM.INV-DEIC DEM.INV-ABS Albert meet.with-CONJ
 ‘and there [he] met with that Albert’ or ‘and there that Albert met [him]’
- awa-ya ama-ya agitaasa-a ngaan*
 DEM.DIST-DEIC DEM.INV-DEIC friend-3SG.POS.ABS to.3SG
uya-qa-a a-âta-na-â
 bring-PART.AN-3SG.AN AUX-apparently-PART-3SG
 ‘there, that friend of his brought him to him_j’
- ilaâta-agiim nu-usa-kan*
 have.as.companion-4SG.ANT go.to-with-CONJ.SG.AN
 ‘when he_i had [him] as a companion, going with him,’
- igiim kim-na-n a-âta-na-n*
 to.4SG go.downhill-PART-PL AUX-apparently-PART-PL
 ‘they went down to him_i’
- Muluka-m ila-ada-a ama-y,*
 milk-REL part-DIM-3SG.POS.ABS DEM.INV-DEIC
- ingay anguna-â qa-châi-duuka-lka-kin*
 DEM.PROX-DEIC be.big-ABS eat-let-PROB-NEG-CONJ.PL.AN
 ‘they couldn’t feed [him] very much, a little bit of milk’
- muluka-m ila-a amasix*
 milk-REL part-3SG.POS.ABS with / and

ataqan saahmla-â agachiida-a igiim qa-châi-sxa-ku-â-nga
 one egg-ABS only-3SG.POS.ABS to.4SG eat-let-PASS-IND-3SG-ENCL
 ‘only a bit of milk and one egg were fed to him’
 (Eastern, undated; transcribed by Dirks, in the 1980’s)

Examples of anaphora or zero-anaphora given in Bergsland (1997) which are readily traceable to their source texts, such as those taken from the Jochelson materials, can be tested against the ideas presented here. Thus, example (40) illustrates one form of non-canonical anaphoric agreement, with the final predicate marked for anaphoric reference to the singular subject of a preceding clause but agreeing in number with its direct subject:

- (40) *Niiğuî-n tana-ngin nu-ku-ğ-aan*
 Atkan-PL land-3PL.POS reach-IND-3SG-ENCL
 ‘When he reached the island of the Atkans,’
Niiğuî-n unglu-m kangan txichi taxa-qa-ngin
 Atkan-PL pinnacle-REL on self.3PL gather-PART.AN-3PL.AN
 ‘the Atkans gathered on a pinnacle.’ (Eastern, J:57:11; Bergsland 1997: 248)

The complete text from which this is taken (see Bergsland and Dirks 1990: 402–404) contains seven or eight instances of anaphora and zero-anaphora, representing many of the problematic and sometimes seemingly fickle features of anaphora I have discussed here. For example, a preceding sentence in this text shows anaphoric agreement identical to example (40), except that it agrees in number with the subject of the first clause, not with its own subject. An analysis of the complete text is too lengthy to present here, but the text-based interpretation of anaphora and zero-anaphora I am proposing works well. Thus, in example (40) there are three groups of participants, one of which is underspecified in this sentence and which is the plural referent of the anaphora: the underspecified group is the hero of the text’s men.

5 Conclusion

Aleut allows context to determine the interpretation of a great many parts of the syntax, and this appears to include the use of anaphoric marking or zero-anaphora. Broadly, there are several main findings, the first two of which are certainly implicit, if not explicit, in Bergsland:

1. Anaphoric marking or lack thereof is not directly tied to grammatical roles, such as subject or object. In Aleut texts, subjects and objects may switch frequently and rather freely without overt marking, unless ambiguity would ensue; and this also holds true of oblique objects and possessors.
2. Anaphoric marking and zero-anaphora do not specifically refer to coreference or switch-reference in the traditional sense: to relate anaphoric marking to coreference in Aleut requires that coreference be broadly interpreted to include subject, object, possessor, and oblique coreferential possibilities.
3. Anaphoric marking and zero-anaphora are intimately related, and one cannot understand the use of anaphoric marking in Aleut without also explaining zero-anaphora.
4. Anaphora and zero-anaphora are related to topic salience; the more salient something is, the less likely it is to be marked. Zero-anaphora is the preferred form when the interpretation of pronominal reference is obvious.

The advantage to an approach which considers discourse context, and which considers anaphora in conjunction with zero-anaphora, is that we can simplify the explanation; a strictly clause-based approach, or one in which only simple clause combinations are considered, requires a long and confusing enumeration of different environments in which anaphoric inflection is found.

By contextualizing examples of anaphora and zero-anaphora, it also becomes possible to view this inflectional system as just another way of tracking nominals through discourse. It is not so very different in principle from tracking information (Du Bois 1987) or topics in clause chaining languages, in particular the related Eskimoan languages (Berge 1997). The differences are of course in the particulars: Aleut marks anaphora, not grammatical role.

In addition to contextualizing the examples being analyzed, it is equally important to place the typologically unusual anaphoric system of Aleut in context. Anaphoric marking is in fact relatively rare in Aleut texts; anaphoric inflection on verbs occurs about one tenth of the time, on average, and the overt expression of participants in discourse can be even rarer: it is not unusual for participants to be underspecified even when newly introduced. In fact, anaphoric marking is only a small part of a system of typologically unusual marking in clause combination, including other aspects of the reference tracking system as well, such as plurality and dependency marking.

All of these show the same characteristic ambiguity and underspecification. Most inflectional morphemes in Aleut are underspecified for function, even down to part of speech to which an inflectional morpheme attaches (hence the confusing terminology, as mentioned in Section 2). To give just one example from the pronominal paradigm of both nouns and verbs, *-ngis* (Atkan) or *-ngin* (Eastern) is used for 3PL dependent (non-anaphoric) endings and for 1st or 3rd subjects (singular or plural) and / or plural objects of anaphoric endings on verbs, and for plural possession on nouns. It is tempting to imagine that inflectional morphemes in Aleut may best be seen as floating morphemes, not specifically attached to a particular grammatical role, or rather, attached to a generalized meaning, e. g. *-ngis* ‘plural’, or ‘plural anaphoric’. At any rate, reference tracking has to be seen in the greater context, as a discourse-level feature, and ultimately, understanding how it functions will require an understanding of linguistic units larger than one or two clauses.

Aleut is greatly different in its reference system from the only other language group to which it is related, Eskimo. Some of the features which characterize its reference system, such as number marking, have been explained as the result of areal diffusion (Leer 1991), and features such as underspecification of arguments may be as well. On the other hand, the importance of discourse, topic tracking, and topic over sentential subject is shared with Eskimo; the notable difference here being the lack of topic marking, reflecting Aleut’s preference for underspecificity, rather than topic marking (cf. Berge 1997; Fortescue 1985, for a different formulation of this, in which Aleut marks theme and topic, whereas Eskimo marks topic and subject). Aleut may reflect one end of a continuum of topic-tracking capabilities in languages for which discourse, rather than syntax, is the linguistic level of primary importance.

Abbreviations

ABL = ablative; ABS = absolutive; AN = anaphoric; CESS = cessative; COND = conditional; CONJ = conjunctive; CONT = continuative; DEIC = deictic; DEM.DIST = demonstrative distal; DEM.INV = demonstrative invisible; DEM.PROX = demonstrative- proximal; DIM = diminutive; DU = dual; ENCL = enclitic; IND = indicative; INTNS = intensifier; LOC = locative; NEG = negative; PART = participial; PASS = passive; POS = possessor; PL = plural; PROB = probability; REL = relative; SG = singular

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0343968.

Notes

1. The examples from Bergsland (1997) are mostly Atkan, because of his familiarity with that dialect. Bergsland also gives copious examples from Eastern, but because they are usually taken from extended texts rather than direct elicitation, they tend to be more complex. While Bergsland references the date of a text, he does not always identify the speaker (speakers are referenced in Bergsland 1994 but not Bergsland 1997) or the text from which the example is drawn unless the text had been previously published (e. g. Aleut translations of the Bible, or the Jochelson materials). For this reason there is an apparent but nonetheless spurious dialectal difference in the data presented here.
2. Absolutive and relative inflection are often not distinguished in plural forms, as a result of which it is not unusual for singular absolutive endings to simply be glossed as ‘absolutive’, rather than ‘absolutive singular’.
3. Unexpressed possessors of subjects do not trigger anaphoric marking; compare example (14) and the following:

- (41) *ada-a hila-kuĥ*
 father-3SG.POS.ABS reading-IND-3SG
 ‘His/her father is reading.’ (Atkan, Bergsland 1981:21)

4. The anaphoric ending *-ngis* is multiply ambiguous, and *kidukungis* can also mean ‘he/she/it is helping them.’
5. Note that this is not a case of embedding; the two clauses are juxtaposed. This is evident in the following example, in which *yaasikam kugan* is a phrase, and the phrase cannot be the object of *chaĥtinaĥ*:

- (42) *tayaġu-ĥ yaasika-m kuga-n txin unguti-ku-un chaĥ-ti-na-ĥ*
 man-ABS box-REL on-LOC 3SG sit.on-IND-3SG.REL.AN crack-cause-PART-3SG
 ‘When the man sat on the box, he cracked it’ (Eastern; Bergsland 1997: 252)

6. This point was also made in Fortescue (1985: 114).
7. This is modified from the following example, for the sake of simplifying the data presented here:

- (43) *Piitra-ĥ qa-ku-ĥ naga-n qangu-na-q*
 Peter-ABS eat-IND-3SG into-LOC come.in-PART-1SG
 ‘While Peter was eating, I came in to him’ (Bergsland 1997: 254)

8. For the theoretical background for this claim, Bergsland (1989) proposed that Aleut was originally a canonical ergative-absolutive system; its present typologically unusual anaphoric system results from the attrition of most oblique cases. Aleut continues to show traces of its original ergative system (Fortescue 1994). Following discourse-based

explanations of ergative case marking (e.g. Du Bois 1987), I proposed that problems of coreference and switch-reference in ergative clause chaining languages (specifically West Greenlandic) could be understood by viewing referential marking as a topic tracking device, rather than a strictly syntactic feature for tracking subjects between two clauses (Berge 1997). See also Payne (1982) for similar arguments with respect to Central Alaskan Yup'ik. Fortescue (1985) proposed an account of Aleut anaphora and thematicity from the perspective of Functional Grammar.

9. The textual examples are taken from recordings made by Bergsland and Dirks in the 1980's, transcribed by Dirks on several occasions, and analyzed as part of my field studies of Aleut.

References

- Berge, Anna
1997 *Topic and discourse structure in West Greenlandic agreement constructions*. Doctoral dissertation, Department of Linguistics, University of California at Berkeley.
- Berge, Anna
2009 Tracking topics: a comparison of topic in Aleut and Greenlandic discourse. In *Variations on Polysynthesis: The Eskaleut Languages*, Marc-Antoine Mahieu and Nicole Tersis (eds.), 185–200. Amsterdam: Benjamins.
- Bergsland, Knut
1989 Comparative Aspects of Aleut Syntax. *Journal de la Société Finno-ougrienne* 82: 7–80.
- Bergsland, Knut
1997 *Aleut Grammar : Unangam Tunuganaan Achixaasiġ*. Fairbanks: Alaska Native Language Center.
- Bergsland, Knut, and Moses Dirks
1981 *Atkan Aleut School Grammar*. Anchorage: National Bilingual Materials Development Center.
- Bergsland, Knut, and Moses Dirks, (eds.)
1990 *Unangam Ungiikangin kayux Tunusangin / Unangam Uniikangis amaTunuzangis / Aleut Tales and Narratives: Collected 1909–1910 by Waldemar Jochelson*. Fairbanks: Alaska Native Language Center.
- Bergsland, Knut
1997 *Aleut Grammar : Unangam Tunuganaan Achixaasiġ*. Fairbanks: Alaska Native Language Center.
- Du Bois, John W.
1987 The discourse basis of ergativity. *Language* 63 (4): 805–855.
- Fortescue, Michael
1985 Anaphoric Agreement in Aleut. In *Predicates and terms in Functional Grammar*, ed. by M. A. Bolkestein, C. de Groot, and J. L. Mackensie, 105–126. Dordrecht: Foris.

Fortescue, Michael

- 1994 The Role of Typology in the Establishment of the Genetic Relationship between Eskimo and Aleut. In *Hokkaido University Publications in Linguistics*, no. 7. Sapporo, Japan: Department of Linguistics, 9–36.

Leer, Jeff

- 1987 *Aleut anaphora*. M.A. thesis. Chicago: University of Chicago.

Leer, Jeff

- 1991 Evidence for a Northwest Coast Language Area: Promiscuous Number Marking and Periphrastic Possessive Constructions in Haida, Eyak, and Aleut. *International Journal of American Linguistics* 57 (2): 158–193.

Payne, Thomas E.

- 1982 Role and Reference Related Subject Properties and Ergativity in Yup'ik Eskimo and Tagalog. *Studies in Language* 6 (1): 75–106.

Sadock, Jerrold M.

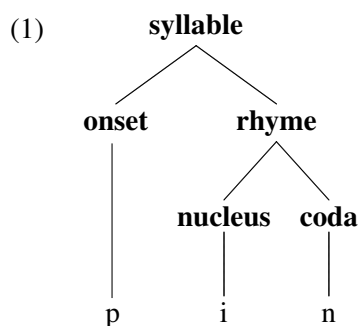
- 2000 Aleut Number Agreement. In *Proceedings of the Berkeley Linguistics Society*, 26, *Special session on Syntax and semantics of the Indigenous languages of the Americas (February 18–21, 2000)*, Andrew K. Simpson (ed.), 121–138. Berkeley: University of California.

The nature of consonant sequences in Modern Georgian

Marika Butskhrikidze

1 Introduction

There is a long and controversial debate concerning the role of the syllable in phonological theory. During the *Sound Pattern of English* (SPE) period (Chomsky and Halle 1968) the notion of syllable had no formal recognition. However, subsequent research within the framework of generative phonology has shown that the syllable is an essential concept for understanding phonological structure. Three kinds of justification have been offered in favour of recognition of the syllable. First, the syllable is a natural domain for the statement of many phonotactic constraints. Second, phonological rules are often more economically and insightfully expressed when they explicitly refer to the syllable. Third, the organization of segments into syllables serves as the base over which the rules and principles of prosody are defined (Kenstowicz 1994). Various structural models for the structure of the syllable have been proposed, of which the onset-rhyme model (1) is the most widely adopted one (cf. Fischer-Jørgensen 1975; Anderson 1981; Selkirk 1982; Cairns and Feinstein 1982).



The representation in (1) is an example of a syllable with a very simple structure. More complex syllables demand the incorporation of additional consonantal material into the onset and the coda. Typically, the creation of

such complex onsets and codas is severely constrained. There are universally two requirements for complex onsets and codas, as formulated in (2) and (3) respectively:

- (2) The number of consonants in onsets and codas should not exceed more than two or three.
- (3) Onsets and codas should obey the Sonority Sequencing Principle.

The Sonority Sequencing Principle (abbreviated: SSP) was originally proposed by Jespersen (1904) and later recognised either as an absolute condition or simply as a preference condition expressing universal markedness values (Blevins 1995).

Georgian challenges both these two purportedly universal requirements. Examples given under (4) illustrate words with complex onsets in Georgian. In these examples, a dash refers to a morphological boundary and a full stop indicates a syllable boundary. These words illustrate two important characteristics of the Georgian syllable structure: 1. Syllable onsets in Georgian may contain more than two consonants (actually as many as six); and 2. The Sonority Sequencing Principle is violated.

- | | | | |
|-----|------------|-------------------|--------------|
| (4) | a) CCV | <i>sp'i.lo</i> | 'elephant' |
| | | <i>sxi.v-i</i> | 'ray' |
| | | <i>rbi.l-i</i> | 'soft' |
| | b) CCCV | <i>mk'la.v-i</i> | 'arm' |
| | | <i>zrdi.lo.ba</i> | 'politeness' |
| | c) CCCC | <i>prtxi.l-i</i> | 'careful' |
| | | <i>msxvi.l-i</i> | 'thick' |
| | | <i>prēxi.l-i</i> | 'nail' |
| | d) CCCCCV | <i>msxvre.va</i> | 'to brake' |
| | | <i>pšxvna</i> | 'to crumble' |
| | e) CCCCCCV | <i>prckvna</i> | 'to peel' |
| | | <i>brdyvna</i> | 'to tear' |

The violations of these purportedly universal phonological principles by Georgian data raise several questions, one of which concerns the issue of arbitrariness. There are two possibilities: either one has to accept the existence of the special phonotactic patterns in Georgian as a random fact of linguistic variation or one has to find new principles to explain the constituency and distributional characteristics of the Georgian consonant sequences. This paper will take the latter approach.

To account for these clusters, I propose the Gradual Consonant Analysis (abbreviated: GCA) incorporating the following paradigmatic, syntagmatic, phonetic, historical and comparative evidence. Examination of consonantal stems of Georgian and the application of the Gradual Consonant Analysis shows that a combination of language-external and language-internal evidence provides a useful basis for exploring “complex” structures of a language.

In most cases, the long consonant sequences of Georgian can be explained as the result of a productive vowel-reduction process combined with a process of complex segment formation. Georgian consonant sequences are generally derived from CVC stems with added vowel-initial affixes. Thus, the phonological complexity is primarily due to morphological complexity because the structures of the CCC type in Georgian can be explained to be the result of reductions of structures of the CVCVCV type.

The rest of the paper consists of five sections. Section 2 introduces Georgian and its phoneme system. Section 3 concerns previous analyses of Georgian consonant clusters. A detailed description of Georgian consonant clusters is given in Section 4. Section 5 concerns Gradual Consonant Analysis and in Section 6 some conclusions are drawn.

2 Georgian and its phoneme system

Georgian belongs to the Kartvelian (South Caucasian) language family. The name of the language group is related to the ethnonym *kartveli* ‘Georgian’. Besides Georgian, the Kartvelian language family consists of Laz, Megrelian, and Svan. Kartvelian languages are spoken within the territory of Georgia, the Central and West Caucasus and the vast territory of the South Caucasus.

The phonemic inventory of Modern Georgian consists of 33 phonemes: 28 consonants and 5 vowels (see Table 1 on the following page). Some notational remarks concern the following: <’> denotes glottalization, <~> on a segment denotes friction in a consonant (this symbol appears on the fricative and affricate consonants).

All 33 phonemes of Georgian can appear in the word-initial position. Word-initial position is characterized by clusters of up to six members. There are no consonant clusters in word-final position in underived words. There are consonant clusters in word-medial position, but since they have sub-patterns of the word-initial ones, only the word-initial consonant clusters will be the focus of the paper.

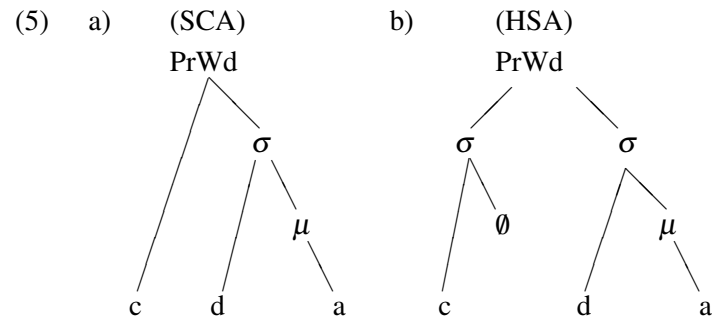
Table 1. The consonant inventory of Georgian

Bilabial	<i>b</i>	<i>p</i>	<i>p'</i>	<i>m</i>				<i>v</i>			
Alveolar	<i>d</i>	<i>t</i>	<i>t'</i>	<i>n</i>	<i>j</i>	<i>ǰ</i>	<i>c'</i>	<i>z</i>	<i>s</i>	<i>l</i>	<i>r</i>
Palatal-alveolar					<i>c</i>	<i>č</i>	<i>č'</i>	<i>ž</i>	<i>š</i>		
Velar	<i>g</i>	<i>k</i>	<i>k'</i>					<i>ɣ</i>	<i>x</i>		
Uvular									<i>χ'</i>		
Laryngeal									<i>h</i>		

3 Previous analyses of Georgian consonant clusters

Two types of analyses have been proposed to account for the Georgian consonant clusters: the Syllabified Consonant Analysis (abbreviated: SCA, Bush 1997) and the Headless Syllable Analysis (abbreviated: HSA, Nepveu 1994). The former recognizes the arbitrariness of the constituency of consonant clusters while proposing strong constraints on the distribution of segments. The latter proposes an analysis in which the Sonority Sequencing Principle is still maintained; hence the constituency of consonant clusters is not arbitrary. To account for the Georgian clusters a sub-cluster structure has to be added.

The structures for *cda* ‘attempt’ as suggested by these two analyses are depicted in (5). These two approaches to Georgian consonant clusters will be discussed extensively below.



3.1 Syllabified Consonant Analysis (SCA)

In the Syllabified Consonant Analysis (as illustrated in 5a) all consonants are “normally syllabified”, i. e. the syllable always has a vocalic nucleus (Bush

1997). As a consequence, the Sonority Sequencing Principle is not always obeyed in Georgian. Bush's account is formulated in the framework of Optimality Theory (OT), using Correspondence Theory (McCarthy and Prince 1995). There are various empirical and theoretical problems with this approach to Georgian consonant clusters.

An empirical problem with the analysis of *cda* 'to attempt, to try, to wait' in (5a) is that the cluster /cd/ cannot be analysed as a true cluster, since there are wordforms with the same root, but with a vowel in between the consonants, e. g. *e-cad-e* 'try you', *v-ecad-e* 'I tried'. Thus, the form *cda* is derived from the stem *cad*. To account for this fact, an approach like described in Bush (1997) would need to introduce something like an unspecified vowel inside the derived word *cda*. A minor problem with the Syllabified Consonant Analysis is that Bush considers the cluster /cd/ to be an instance of a violation of the Sonority Sequencing Principle. However, the consonant /c/ is a voiceless affricate in Georgian and the consonant /d/ is a voiced stop. So, it seems that Bush (1997) recognizes a sonority ranking in which all affricates are more sonorous than stops, regardless of laryngeal specification. Another problem with the Syllabified Consonant Analysis is empirical as well as theoretical in nature. As Bush notes:

"Without the Sonority Sequencing Principle, we expect to have clusters of arbitrary contours and length, like *rmqb*, *tdtdtd*, or *trblqsvntkb*. While it is true that only a small percentage of the clusters that would arise from random combinations of consonants are actually attested, I claim that this fact does not need to be accounted for by the phonology." (Bush 1997: 48)

To summarize, Bush proposes that the first syllable of Georgian words allows for arbitrary combination of consonants, which should not be within the scope of phonology. However, the arbitrariness of the complex clusters in the first syllable of a word is not empirically justified (see Section 4) so a phonological theory must account for them. To account for the Georgian clusters, Bush (1997) proposes several distributional generalizations, as summarized in (6).

- (6) a. Only the first syllable of a morpheme allows SSP-violating clusters.
- b. The second syllable of a morpheme allows only clusters that do not violate the Sonority Sequencing Principle.
- c. The third syllable of a morpheme does not allow any clusters at all.

Loanwords are considered as problematic for his analysis, but the examples he cites are not problematic. One possible problematic word is a proper name *dos.to.ev.ski*, but cross-linguistically proper names are special in many respects and there are likewise many Georgian proper names with complex structure, e. g. [me.-č'ur.č'l-e-t-u-xu.c-e.si] with a complex cluster in the third syllable. As for the other problematic words cited by Bush (1997), viz. [or.to.-gra.pi.-a] and [kro.no.-me.tr-i], they are morphologically complex words. Bush's (1997) analysis of the Georgian facts is otherwise strictly based on underived words. Without this restriction there are a lot of words in Georgian with complex clusters in the third or even in the fourth syllable of a word, for example [ma.ma.švi.lu.ri] 'parentally'; [sa.k'a.co.bri.o] 'worldly'. These examples suggest that the stem should be the relevant domain for formulating restrictions on syllable structures.

The main problem with Bush's analysis lies in considering the clusters to be a mismatch between morphological and phonological prominence. Bush (1997) almost completely disregards the asymmetry between lexical and grammatical morphemes and accounts for the distributional patterns of complex clusters only in terms of counting syllables. This strategy raises the question why the violation of the Sonority Sequencing Principle occurs only in the first syllable, and not in the second syllable, and why there are no clusters in the third syllable. There is no explanation in Bush's analysis of where these patterns could come from, and I believe there is no way to explain them without referring to the asymmetry of lexical vs. grammatical morphemes.

3.2 Headless Syllable Analysis (HSA)

In Headless Syllable Analysis (as illustrated in 5b) Nepveu (1994) proposes that all Georgian oversized clusters, as well as those clusters not obeying the Sonority Sequencing Principle, can be broken down into series of sub-clusters of the form T (obstruent), R (resonant) or TR (obstruent + resonant), with at most one sub-cluster occurring at each morpheme boundary. Such sub-clusters can be the well-formed onsets of headless syllables. Their existence and distribution are accounted for within Optimality Theory and Prosodic Morphology. There are various problems with this approach to explaining Georgian consonant clusters.

First of all, Nepveu (1994) rightly notices that the domain of the distribution of complex clusters is morphological in nature. However, there is a great deal of variation about the precise specification of this domain in his analysis.

Throughout his analysis the domain where the sub-clusters are found varies between: the morpheme, the prosodic word, and the root.

Another problem that arises with the adoption of the Headless Syllable Analysis is that it is no longer possible to capture regularities within consonant sequences after positing that the entities (sub-clusters) which are the constituents of the oversized long consonant complexes are independent of each other. Thus, an account of the restrictions attested within complex clusters is completely missing in this type of analysis. Even though Nepveu (1994) begins his analysis with denying the arbitrariness of the composition of the Georgian complex clusters, he proposes an analysis in which the relationship between the sub-constituents of the long consonant sequences is completely overlooked.

Even though in both analyses it is noticed that clusters that do not obey the Sonority Sequencing Principle are restricted to a particular morphological constituent the morphological constituency is not properly appreciated as a factor for the analysis. In the following section I will introduce some relevant phonotactic generalizations of Georgian consonant clusters. These generalizations will be the basis for the subsequent explanation of their structure.

4 Georgian consonant clusters

In general, combinations of identical consonants are unusual in Georgian. This evidence can be used as a starting point when discussing the natural classes of phonemes.

The first restriction concerns labial phonemes (including the labio-dental /v/). It is not permitted to have two adjacent labial phonemes mono-morphemically. Any combination between the members of this class is unacceptable in initial position. Thus /**mb-*, **mp-*, **mp'-*, **mv-*, **bm-*, **bp-*, **bp'-*, **bv-*, **pm-*, **pb-*, **pv-*, **vm-*/ etc. are not accepted. We refer to this class as class of bilabials. It contains: /b, p, p', m, v/.

The second class contains coronals: /d, t, t', j, c, c', z, s, ʃ, č, č', ž, š/, in which three types of place of articulation are distinguished as follows: dentals /d, t, t'/; alveolars /j, c, c', z, s/; palato-alveolars /ʃ, č, č', ž, š/. The class of coronals has two restrictions: 1. Combinations of homorganic phonemes are not accepted (i.e. the same restriction as for labials); 2. A coronal with a relatively backward place of articulation can precede a coronal with a relatively forward place of articulation, but never follow it. Thus, combinations: /št, ʃd, cd/ are attested, while /**tš*, **dʃ*, **dc*/ never occur.

The third class has the following restriction: the members of the class never combine (i. e. the same restriction as for the first and the second class). It contains sonorants: /r, l, n/.

Velar and uvular consonants can be grouped together. They never combine with each other. We refer to this class as the class of velar-uvulars. It contains /g, k, k', ɣ, x, χ', h/ (Uturgaidze 1976).

The common restriction for all these four classes is as follows:

- (7) Members within each class are never adjacent.

Two additional factors are of importance when discussing adjacent consonant combinations:

- (8) a. The successive places of articulation of the members of the cluster,
and
b. the laryngeal specification.

With regard to the first factor, clusters are distinguished as regressive or non-regressive. Regressive clusters combine consonants with front and back place of articulation, e. g. labial-coronal, coronal-dorsal, labial-dorsal. Non-regressive clusters are clusters with back-to-front place of articulation, e. g. dorsal-coronal, coronal-labial, dorsal-labial. With regard to the second factor, clusters are defined as “homogeneous” or “heterogeneous”. In “homogeneous” clusters members share the laryngeal feature. Thus, they are voiced, voiceless or glottalized. Most scholars (Akhvlediani 1949; Vogt 1961; Melikishvili 1997) consider regressivity and homogeneity as the basic features characterizing Georgian consonant clustering.

There are phonemes and phoneme groups in Modern Georgian whose status is difficult to establish without referring to their historical development. This concerns the status of harmonic groups, labialised consonants (the status of the sonorant /v/) and the sonorant /r/. All of them are discussed in the following sections.

4.1 Harmonic clusters

Clusters of [-dorsal] [+dorsal] obstruents are called harmonic in Georgian because they share a laryngeal specification. There are two types of harmonic clusters: referred to as A and B. Type A refers to the combinations of stops, affricates and fricatives with the velar stops /g, k, k'/ and type B refers to

the combinations of stops, affricates and fricatives with the fricatives /ɣ, x, χ'/. In some studies the combinations of fricatives with dorsal obstruents are also treated as harmonic clusters (Marr 1925; Gamkrelidze and Machavariani 1965). I incorporate these additions and present the clusters as follows:

(9)	Type A	Type B
	<i>bg pk p'k'</i>	<i>bɣ px p'χ'</i>
	<i>dg tk t'k'</i>	<i>dɣ tx t'χ'</i>
	<i>jg ck c'k'</i>	<i>jɣ cx c'χ'</i>
	<i>ǰg čk č'k'</i>	<i>ǰɣ čx č'χ'</i>
	<i>zg sk</i>	<i>zɣ sx</i>
		<i>žɣ šx</i>

Below I give a list of phonological arguments supporting the claim of analysing the harmonic groups as complex segments.

Harmonic clusters always syllabify together intervocalically, while other obstruent clusters are never tautosyllabic, e. g. *sit'χ'va* 'word' [si.t'χ'va], *cecxi* 'fire' [ce.cxli] (Akhvlediani 1949; Jgenti 1956).

In reduplicated forms, harmonic clusters retain their complexity, e. g. *čkar-čkara* 'quickly'; *cxel-cxeli* 'hot'. Other types of clusters do not usually participate in reduplicated forms. Exceptions are clusters containing the phoneme /s/, e. g. *sc'or-sc'ori* 'right', *svel-sveli* 'wet', etc.

Harmonic clusters were present in Old Georgian and have even been reconstructed for Proto-Kartvelian. There are many examples of correspondences illustrating harmonic clusters in all Kartvelian languages, e. g.

(10)	Georgian	Megrelian	Laz	Svan	Gloss
	<i>mat'χ'l-</i>	<i>mont'χ'or-</i>	<i>mont'k'or</i>		'wool'
	<i>txra</i>	<i>txorua</i>	<i>ontxoru</i>	<i>lištxri</i>	'to dig'

The harmonic clusters are considered as complex segments in numerous studies (Vogt 1961; Gamkrelidze and Machavariani 1965; Bush 1997; Nepveu 1994; Cho and King 1997; etc.). The analysis of harmonic clusters as complex segments has been substantiated by perceptual experiments as well (see Butskhrikidze and Heuven 2001).

4.2 Sonorants

Sonorants are obligatory constituents of long consonant sequences. Establishing their status therefore plays an important role in a proper analysis of the clusters. There is some historical evidence for the status of the sonorant /v/ as a defective segment. According to some studies (e. g. Gamkrelidze and Machavariani 1965), there was a distinction between syllabic and non-syllabic /u/ in Proto-Kartvelian, which was maintained in all Kartvelian languages. The mutation of the non-syllabic /u/ to sonorant /v/ happened in a later period of Georgian language development. The Modern Georgian sonorant /v/ corresponds to the Old Georgian non-syllabic /u/. The correspondences are illustrated by the examples given below:

(11)

Old Georgian	Modern Georgian	Phonetic representation	Gloss
<i>varsk'ulavi</i>	<i>varsk'vlavi</i>	[vars ^k ʷlavi]	'star'
<i>čuen</i>	<i>čven</i>	[čʷen]	'we'
<i>k'uali</i>	<i>k'vali</i>	[kʷali]	'trace'

In these cases the orthographic *v* is probably cognate to the Old Georgian non-syllabic /u/ and even though in the written form we see consonant clustering (e. g. /k'vd/ in *sik'vdili*; /kvs/ in *ekvsi*, etc.), /v/ can be seen as a secondary articulation on the previous consonant (Nepveu 1994; Bush 1997).

The important consequence of considering the defective nature of the sonorant /v/ in the analysis of the clusters is that the combination of C+/v/ can be presented and analyzed not as a true cluster, but as a sequence of a consonant followed by the glide /w/ (C^w).

The most sonorous sonorant /r/ also has interesting patterns. I argue that in Modern Georgian /r/ is phonetically syllabic, as shown by the following: The sonorant /r/ can precede and follow any consonant or vowel. In some studies, /r/ is described as a "fleeting" sonorant (Vogt 1961). In certain words it can appear optionally, e. g.

(12)

<i>prta</i> ~ <i>pta</i>	'wing'
<i>grdemli</i> ~ <i>gdemli</i>	'anvil'
<i>brjeni</i> ~ <i>bjeni</i>	'wise'

However, there are cases when the presence of the sonorant /r/ is obligatory, e. g. in complexes like /jrc'/, /jrc/, /grk'/, etc. As we can see in the examples given in (12) consonants in sequences like /prt/, /grd/, /brj/ share

a laryngeal specification: they are either voiced or voiceless and without /r/ well-formed clusters. That is not the case with clusters /jrc'/, /jrc/, /grk'/. Since /*jc', *jc, *gk'/ are ill-formed clusters, the presence of the sonorant /r/ is obligatory, and this very fact may be evidence of the syllabic nature of the sonorant /r/.

Another interesting characteristic of the sonorant /r/ is that it is the only sonorant, which can appear between identical consonants; geminates are not otherwise allowed in Modern Georgian. Thus /grg, trt, žrz/ are well-attested sequences for Georgian. This appears to provide more evidence for the syllabicity of Modern Georgian sonorant /r/.

There are some phonotactic restrictions which support the claims made so far about C+/v/ combinations (treated as one segment) and about the syllabicity of the sonorant /r/.

The co-occurrence of sonorant /v/ with the other sonorants seems sensitive to number of obstruents in a cluster. Two observations can be made:

- (13) a. In a cluster CS₁S₂ (obstruent+sonorant+sonorant), S₁ is always the sonorant [v], while S₂ is either [n], [l] or [r], e. g. /k'vn/ in *k'vnesa* 'to moan', /kvn/ in *kvnet'a* 'to bite', /gvr/ in *gvrit'i* 'turtle-dove'.
- b. In clusters with two obstruents (e. g. C₁S₁C₂S₂) or more, the order of sonorants is different. The most sonorous sonorant /r/ takes the first position and the sonorant /v/ can be only the second sonorant of a cluster, e. g. in combinations /drt'v/ in *drt'vinva* 'to grumble', /grgv/ in *grgvinva* 'thunder', /brč'χ'v/ in *brč'χ'viali* 'sparkling'.

With regard to these generalizations, the cases of combinations where C₁ filled with a harmonic cluster or /s/+obstruent cluster are interesting. Both combinations count as a single obstruent for these generalizations, e. g. *pxvn* in *pxvnili* 'powder', *sk'vn* in *sk'vna* 'tie up' are in line with the generalization formulated in (13a), thus, these clusters could be formalized as CS₁S₂.

From a specific case of sonorant distribution that is generalized in (13b), one could argue that the cluster of the type C₁S₁C₂S₂ cannot be analysed as one whole, but instead is divisible into sub-structures C₁S₁/C₂S₂. For the examples given in (13b) the division will be as follows: *dr/t'vinva*, *gr/gvinva*, *br/č'χ'viali*.

To summarize, both diachronic patterns and synchronic distributional generalizations of harmonic groups and the sonorants /v/ and /r/ strengthen three claims:¹

- (14) a. Harmonic clusters can be analysed as complex segments;
- b. Consonant +/v/ combinations can be treated as one segment, assuming that /v/ is a secondary articulation on a preceding consonant;
- c. The sonorant /r/ can be analysed as syllabic.

These three claims facilitate greatly the analyses of the Georgian consonant clusters to be given in Section 5.

5 The Gradual Consonant Analysis

Many linguists have noticed that long consonant sequences occur predominantly in verbal rather than in nominal forms in Georgian (Vogt 1961 and Ertelishvili 1970, among others). This could be related to the fact that Georgian verbs have a much more complex morphological structure than nouns. Thus, verbal forms take up a larger part of the discussion, although analyses of nominal forms are also considered.

This section contains two parts. In the first part (5.1), consonantal verbal stems are examined. Comparative and historical evidence suggests that Modern Georgian consonantal stems containing up to four consonants are historically derived from stems that have a vowel between the consonants. The second part (5.2) introduces the Gradual Consonant Analysis. Its application is demonstrated on verbal and nominal forms containing long consonant sequences.

5.1 Consonantal verbal stems

The data considered are meant to prepare the reader for the analysis, the Gradual Consonant Analysis, which follows. A close study of the verbal forms suggests that consonant sequences in Modern Georgian are historically derived from much simpler forms. They are the result of processes of vowel deletion and complex segment formation.

Ertelishvili (1970), in his study of the phonotactic structure of verbal stems, suggests that consonant sequences, their history and characteristics are

directly related to the phonotactic patterns of stems. He illustrates this by examining consonantal verbal stems. Modern Georgian stems that consist only of consonants are derived from Old Georgian stems that have a vowel between the consonants. Sometimes the corresponding stems with a vowel between the consonants are also attested in paradigmatically related forms and in dialectal forms.

Consider the following verbal stems, consisting of two, three and four consonants. The data are taken from Ertelishvili (1970). According to Ertelishvili (1970), there are 45 biconsonantal verbal stems in Georgian. The number of consonants is counted as they appear in surface forms. I consider some of the types of biconsonantal verbal stems and illustrate that in most cases consonant sequences are derived after stem-vowel deletion has applied.

(15) Stems with two consonants

Modern Georgian			Old Georgian	
Stems	CC		CVC	
<i>bn / ben</i>	<i>e-u-bn-eb-a</i>	'somebody says to somebody'	<i>v-e-u-ben-i-t</i>	'we spoke'
<i>sm / sem</i>	<i>e-sm-i-s</i>	'somebody hears'	<i>še-g-e-sem-in</i>	'you will hear'

In the examples in (15), the stem vowel deletes when the stem is preceded or followed by a grammatical affix containing a vowel.

(16) $(V)_{affix}\text{-}CVC_{stem}\text{-(}V)_{affix} > (V)_{affix}\text{-}CC_{stem}\text{-(}V)_{affix}$

When the second consonant of the stem of the CC type is /v/ in Modern Georgian, the corresponding form in Old Georgian has the vowel /u/ instead of /v/.

(17) Stems of the C/v/ type

Modern Georgian		Old Georgian	
Stems	C/v/	CV	
<i>sv / su</i>	<i>še-sv-a</i>	<i>da-su-a</i>	'somebody made sit'
<i>c'v / c'u</i>	<i>da-c'v-a</i>	<i>da-c'u-a</i>	'somebody burned'

Notably, it seems to be the case that when a stem of the CC type contains a harmonic cluster, relatively little diachronic change takes place and harmonic

clusters remain unaffected. The clusters behave as complex segments. Consider the following correspondences:

(18) Stems of the Cc type

Modern Georgian			Old Georgian	
Stems	Cc		Cc	
<i>cx</i>	<i>a-cx-ob-s</i>	‘somebody bakes’	<i>še-a-cx-o-t</i>	‘you to bake’
<i>c’χ’</i>	<i>da-a-c’χ’-o</i>	‘somebody had put’	<i>še-u-c’χ’-ev</i>	‘put you’

There are also biconsonantal stems in Modern Georgian which do not correspond to Old Georgian stems of the CVC type; however, comparison of such consonantal stems with corresponding forms in other Kartvelian languages, e. g. Svan and Megrelian, provides some evidence for their derived nature. One of these stem types is the sequence C/r/, which has a corresponding stem of the CV/r/ type in other Kartvelian languages.

(19) Stems with /r/

Modern Georgian		Svan	Megrelian
Stems	Cr	CVr	CVr
<i>br / bVr</i>	<i>br-ial-i</i>	<i>bər-bər</i>	<i>bar-bač</i> ‘sparkle’
<i>gr / gVr</i>	<i>gr-ial-i</i>	<i>gər-gən</i>	<i>gur-gun</i> ‘rumble’
<i>zr / zVr</i>	<i>zr-ial-i</i>	<i>zər-zən</i>	<i>zir-zol</i> ‘humble’
<i>pr / pVr</i>	<i>pr-ial-i</i>	<i>pər-pən</i>	<i>par-pal</i> ‘to fly’

As shown in the examples in (19), Modern Georgian stems of the C/r/ type correspond to stems in other Kartvelian languages which have a vowel between the consonants, i. e. CV/r/. In Svan, the stem vowel is a schwa, and in Megrelian, the vowel can be /a/, /u/ or /i/.

There are 42 stems in Georgian containing three consonants only. As illustrated below, these consonant sequences emerge as the result of stem-vowel deletion. The process, as already suggested in (16), occurs in the following context: (V)_{affix}-CVC_{stem}-(V)_{affix}.

(20) Stems with three consonants

	Modern Georgian	Old Georgian
Stems	CCC	CVCC
<i>drk' / derk'</i>	<i>drk'-eb-a</i> 'somebody bows'	<i>mo-v-derk'</i> 'I bowed'
<i>k'rt / k'ert</i>	<i>k'rt-om-a</i> 'to tremble'	<i>gan-h-k'ert</i> 'you trembled'

Once again, as in the case of the biconsonantal stems in (17), Modern Georgian /v/ corresponds to the vowel /u/ in Old Georgian.

(21) Stems with /v/

	Modern Georgian	Old Georgian
Stems	C/v/C	CuC
<i>xvd / xud</i>	<i>še-xvd-a</i> 'he met somebody'	<i>še-m-xud-a</i> 'somebody met me'
	CC/v/	Ccu
<i>tkv / tku</i>	<i>tkv-a</i> 'somebody said'	<i>tku</i> 'said'

Similar to the examples in (18), harmonic clusters behave as complex segments.

(22) Stems with Cc

	Modern Georgian	Old Georgian
Stems	CCc	CVCc
<i>rtx / ratx</i>	<i>gan-i-rtx-a</i> 'was spread'	<i>gan-i-ratx</i> 'spread you'
<i>rt'χ' / rat'χ'</i>	<i>še-rt'χ'-m-a</i> 'to bound'	<i>še-i-ratχ'</i> 'somebody bounded'

According to Ertelishvili (1970), stems with the sonorant /r/ between identical consonants are always derived from reduplicated forms.

(23) Stems with /r/

Literary Georgian			Dialectal form	
Stems	$C_i r C_i$		CVrCVC	
<i>trt</i>	<i>trt-ol-a</i>	'to tremble'	<i>tar-tal-eb-s</i>	'tremble'
Literary Georgian			Related form	
<i>prp</i>	<i>prp-en-a</i>	'to adore'	<i>par-pat'-i</i>	'to fly around'
			<i>pr-en-a</i>	'to fly'

Note that the forms of the Modern Georgian C/r/ stems given in (19) are also derived from reduplicated forms.

As was argued in Butskhrikidze (2002), the sonorant /r/ is syllabic in two contexts: (i) when preceded and followed by consonants which do not share laryngeal specification; and (ii) when preceded and followed by identical consonants. The examples in (20) and (23), where the Modern Georgian forms have corresponding forms (in (20) Old Georgian and in (23) Svan and Megrelian) with a vowel before the sonorant /r/, also substantiate the claim that Modern Georgian /r/ is syllabic in those contexts.

Stems containing four consonants only are very rare, and similarly to the previous cases, it is always possible to reconstruct a vowel between the consonants by taking into consideration paradigmatically related forms, historical and comparative evidence. The following alternations are attested:

- (24) Stems with four consonants: *crcv* / *jarcu* 'to fade'
crcn / *crcen* / *cercen* 'to peel off'
rc'χ'v / *c'χ'v* / *c'χ'ev* 'to deplete'

As shown by the examples in (15), (19), (20) and (23), the comparative, historical and related forms demonstrate that Modern Georgian consonantal verbal stems are derived from stems of the CVC type. The deletion of a stem vowel is triggered by the addition of a vowel-initial affix to the stem. Thus, deletion occurs in the following context:

- (25) The context of stem-vowel deletion: $(V)_{affix}$ -CVC_{stem}-(V)_{affix}

The presence of either a prefix or a suffix is necessary for stem-vowel deletion to apply.

As shown by the examples in (15), (19) and (20), the stem vowel which is deleted is usually either /a/ or /e/. In general these are the vowels which undergo deletion in Modern Georgian (Butskhrikidze 2002).

To conclude, the data considered in this section suggest that consonant sequences in Modern Georgian are historically derived from much simpler forms. They are the result of processes of vowel deletion and complex segment formation.

This short excursion into the history and constituency of the consonantal verbal stems is intended as an introduction to the types of evidence (e. g. historical, comparative, paradigmatic and syntagmatic) used in the analysis of consonant sequences, called the Gradual Consonant Analysis, and to the idea that in order to study and analyse the long consonant sequences, various aspects of the language should be taken into account. In the following pages I spell out what “various” means. The Gradual Consonant Analysis assumes the prosodic hierarchy (segment – stem + (affix)-word).

5.2 The Gradual Consonant Analysis applied

Before addressing the procedural aspects of the Gradual Consonant Analysis, I should explain why the analysis is called “gradual”. The adjective *gradual* implies proceeding or happening step by step or by degree. The term “gradual”, as it is used here, refers to the type of procedural mechanism, which applies step by step to the consonant sequences. The procedure continues in the case of the presence of relevant evidence, eventually affecting the degree of complexity of consonant sequences. Each step in the procedure yields a less complex structure, eventually arriving at the simplest possible form. Without proper evidence the procedure does not proceed further. Thus, the existence of proper evidence is crucial to the mechanism of the analysis. What are the types of evidence considered in the Gradual Consonant Analysis?

(26) Types of evidence in the Gradual Consonant Analysis:

- (a) Paradigmatic and syntagmatic
- (b) Historical
- (c) Phonetic
- (d) Comparative

I will briefly summarise what each type of evidence refers to, and then demonstrate the application of the Gradual Consonant Analysis to the Georgian data.

Paradigmatic and syntagmatic evidence encompasses the following:

- (a) The ability of consonants to participate in minimal pairs, i. e. their distinctive function
- (b) The appearance of consonants in paradigmatically related forms
- (c) Distributional patterns, i. e. combinatorics of consonants

These patterns establish the status of consonants in isolation and in sequence. For instance, this type of evidence is used to establish the status of harmonic clusters as complex segments.

Historical evidence, showing the regular changes between the two stages of a language's development, establishes the status of consonants both in isolation and in sequence. For instance, this type of evidence has been used to establish the status of the sonorant /v/ as a secondary articulation on the preceding consonant.

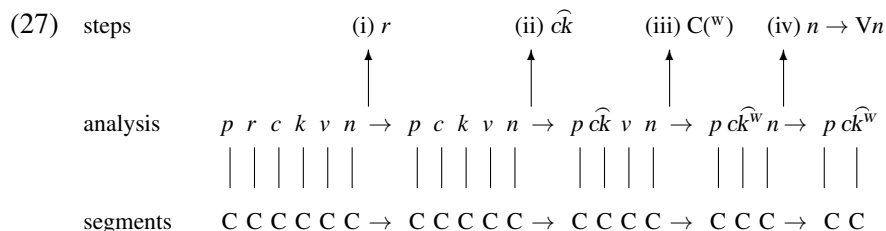
Phonetic evidence is based on articulatory, acoustic or perceptual study. Such studies can reveal the status of consonants in isolation and in sequence, and, eventually, the result of a phonetic experiment can be incorporated in the analysis. For instance, this type of evidence has been used to establish the status of harmonic clusters as complex segments.

Comparative evidence is based on the direct correspondences between the literary form and the corresponding dialectal form, and between the language form and its corresponding form in genetically affiliated languages.

I now consider the application of the Gradual Consonant Analysis to Georgian consonant sequences.

Consider, for instance the longest possible cluster /prckvn/ in *prckvna* 'to peel'. Several generalisations and assumptions should be taken into account. For expository purposes the analysis is presented in a stepwise fashion, the succession of the steps being fairly random. In the first step (i), the optional appearance of the sonorant /r/ is taken into account. This leaves us with the five-member sequence /pckvn/. In the second step (ii), the harmonic cluster /ck/ is treated as one segment. This leaves us with the four-member sequence /pckvn/. In the third step (iii), the C + /v/ combinations are regarded as one segment. This yields the three-member sequence /pck^wn/. In the fourth step (iv), a paradigmatically related form with a vowel before /n/ is taken

into account, e. g. /ga-v-pcken-i/ 'I peeled'. Thus, we are left with the two-member cluster /pck^w/. The analysis is depicted in (27).



As shown in (27), the analysis of the longest cluster, /prckvn/, yields the two-member cluster /pck^w/. One might say that even this two-member cluster can be analysed as a CVC sequence, using comparative evidence. In Megrelian, the corresponding form has a vowel between the consonants /p/ and /ck/; thus, the word /purckon-u-a/ 'to peel' is attested. This observation leads to the assumption that even the longest Georgian clusters can be analysed as having a CVCV... template. The word /prckvna/ is a derived word and can be analysed as: [[pVck^w]_{stem}-[en]_{affix}[a]_{affix}]_{word}.

Thus, the surface six-consonant sequence is the result of vowel deletion, which occurs twice: the first vowel which is deleted is the stem vowel, denoted by the unspecified V, and the second one is a vowel of the affix -en, which is deleted after the addition of another infinitival affix, /-a/. Surface complexity is also related to the optional appearance of the sonorant /r/ and complex segment formation (e. g. the harmonic cluster /ck^w). In the metathesis process discussed in Butskhrikidze (2002) and extensively in Butskhrikidze and van de Weijer (2001), it is shown that in verbal forms the morphological /v/ tends to become a secondary articulation on preceding consonants, especially the velars. It is possible that in the word /pck^wna/, the secondary articulation on the harmonic cluster /ck/ is derived, i. e. has a morphological origin. It is also possible that it is the result of the deletion of the vowel /o/ (Butskhrikidze 2002). Perhaps the latter suggestion is more plausible, because the corresponding Megrelian form is /purckon-u-a/, with a rounded bilabial vowel /o/ after the harmonic group /ck/. Thus, the word /prckvna/ can be analysed as: [[pVck^w]_{stem}-[en]_{affix}[a]_{affix}]_{word}, i. e.: [[[[CVC]_{stem}VC_{affix}]V]_{affix}]_{word}.

Let us now analyse long consonant sequences where the occurrence of the sonorant /r/ is obligatory, for instance the sequence /grgv/ in the word *grgvina* 'thunder'. The application of the analysis to the sequence /grgv/ is as follows: in the first step (i), the phonetic syllabicity of the sonorant /r/ is

taken into account. In the second step (ii), the C + /v/ combination is regarded as one segment. This yields the sequence /gVrg^w/. The procedure is depicted in (28):

(28)	steps		(i) r		(ii) C^w
	analysis	$grgv$	\rightarrow	$gVrgv$	\rightarrow $gVrg^w$
	segments	CCCC		CVCCC	CVCC ^w

According to Ertelishvili (1970), the form *grgvinva* is historically derived from the reduplicated form **gur-gun*. The form *gur-guli/gur-guri* ‘thunder’ is attested (the sonorants /n/, /l/ and /r/ alternate quite frequently in reduplicated forms; see Ertelishvili 1970). Thus, the verbal form /grgvinva/ seems to be derived from the noun form by addition of the thematic suffix /-av/ (compare for example the form /grgvin-av-s/ ‘thunders’) and the infinitival suffix /-a/. The occurrence of the /v/ seems to compensate for the deletion of the bilabial vowel /u/. The /v/ serves as a secondary articulation on the dorsal consonant /g/. Clusters of the /grg/ type, such as /trt/ in *trtola* ‘to tremble’ and *crc* in *crcena* ‘to fade’ are attested in infinitival verbal forms, and, as Ertelishvili (1970) suggests, they are derived from the reduplicated forms of the C/v//r/CV sonorant type. These reduplicated forms are usually onomatopoeic words. Thus, the derived word /grgvinva/ can be represented as: [[[gVr]_{stem}[g^win]_{stem}av]_{affix}a]_{affix}]word.

The words /crcena/ and /trtola/ have the same type of representation as the word /grgvinva/.

Now let us consider the noun *msxverp’li* ‘sacrifice’, with four consonants in word-initial position. In the analysis of this form, several considerations should be taken into account: (i) the sonorant /m/ is a nominaliser in this form, i. e. a prefix, (ii) the sequence /sx/ is a complex segment and (iii) the sonorant /v/ is a secondary articulation on /s^w/. This leaves us with /sx^w/. In Old Georgian, the sonorant /v/ had a corresponding vowel /u/. Besides the word-initial consonants this noun form also has word-medial consonant sequences. This word seems to be a compound, i. e. it is morphologically complex.²

Consider furthermore the adjective *mjlavri* ‘powerful’. As was the case with /msxverp’li/, /m/ is a nominaliser in this form. The consonant sequence /jl/ belongs to the stem. Actually the word *mjlavri* is derived from *jala* ‘power’. Thus, the sequence /jl/ is the result of the stem-vowel deletion, which occurs after the affixation of /-av/ to the stem /jala/.

An interesting fact observed in the analysis of forms with long consonant sequences is that words which at the first sight seem morphologically undervived are in fact derived. Thus, the formation of consonant sequences and morphological derivation are related processes. This relates directly to the above-mentioned observation that consonant sequences are predominantly found in verbal forms which are characterised by morphologically complex structure.

As has been shown, consonant clustering is primarily the result of stem-vowel deletion, which generally occurs because of the addition of a vowel-initial affix to the stem. Additionally, the optional appearance of the sonorant /r/ and complex cluster formation (e. g. harmonic clusters and labialised consonants) make consonant sequences even more complex. The Gradual Consonant Analysis, by considering different types of evidence, offers in-depth analysis of consonant sequences and predicts where, when and which (in terms of constituency) types of consonant sequences are expected to emerge during language change.

6 Conclusions

In this paper we have shown that due to their length and internal structure, the consonant sequences of Georgian are a challenge to linguistic theory.

There are universally two requirements for complex onsets and codas: 1. The number of consonants in onsets and codas should not exceed more than two or three; and 2. Onsets and codas should obey the Sonority Sequencing Principle.

Georgian challenges both these two purportedly universal requirements: 1. Syllable onsets in Georgian may contain as many as six consonants; and 2. The Sonority Sequencing Principle is violated.

Two possibilities have been considered: either to accept the existence of the special phonotactic patterns in Georgian as a random fact of linguistic variation or to find new principles to explain the constituency and distributional characteristics of the Georgian consonant sequences. This paper took the latter approach.

Examination of consonantal stems of Georgian and the application of the Gradual Consonant Analysis has shown that a combination of language-external and language-internal evidence provides a useful basis for exploring “complex” structures of a language. Several criteria have been used to compare the Gradual Consonant Analysis to previous analyses as found in

the literature: (i) simplicity or economy, (ii) generality or abstractness and (iii) adequacy (empirical and explanatory). The evaluation demonstrated that the Gradual Consonant Analysis provides a better understanding of Georgian consonant sequences as it provides a direct and clear link between empirical structures and theoretical constructs, and explains why consonants form complex structures in Georgian.

In most cases, the long consonant sequences of Georgian have been explained as the result of a productive vowel-reduction process combined with a process of complex segment formation. Georgian consonant sequences are generally derived from CVC stems with added vowel-initial affixes. Thus, the phonological complexity is primarily due to morphological complexity. The structures of the CCC type in Georgian can be explained to be the result of reductions of structures of the CVCVCV type.

An interesting fact observed in the analysis of forms with long consonant sequences is that words which at the first sight seem morphologically unde-
rived are in fact derived. As has been shown, the comparative evidence from genetically related languages and dialectal forms shed some light to the origin of the Modern Georgian forms with complex clusters. Further study needs to be done in this direction to uncover the true nature of some problematic cases as e. g. is the case with *msxverp'li* which seems to be morphologically complex one but it is difficult to decompose because of the lack of relevant evidence.

Acknowledgments

I am very grateful to a number of people with whom I have discussed various issues raised in this paper at various times. Among these are: Rusudan Asatiani, Colin Ewen, Thomas Gamkrelidze, Vincent van Heuven, Harry van der Hulst, Larry Hyman, Claartje Levelt, Irine Melikishvili, Miklós Törkenczy, Tedo Uturgaidze and Jeroen van de Weijer. None of these people, of course, should be held accountable for my views or mistakes.

Notes

1. Besides, both generalizations (13a, 13b) suggest the treatment of the harmonic clusters and /s/+obstruent clusters as a complex segment.
2. For the time being, I am unable to trace its origin and discuss the association of the consonants /r/, /p'/ and /l/ to any structural constituent.

References

- Akhvlediani, Giorgi
 1949 *Zogadi ponetikis sapudzvlebi*. [Foundations of general phonetics]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.
- Anderson, Stephen R.
 1981 *Phonology in the twentieth century: theories of rules and theories of representations*. Chicago: University of Chicago Press.
- Blevins, Juliette
 1995 The syllable in phonological theory. In *The handbook of phonological theory*, edited by J. A. Goldsmith, 206–244. Oxford: Blackwell.
- Bush, Rayn
 1997 *Georgian syllable structure*. MA thesis. University of California Santa Cruz.
- Butskhrikidze, Marika
 2002 *The consonant phonotactics of Georgian*. Doctoral Dissertation, Leiden University. Utrecht: Landelijke Onderzoekschool Taalwetenschap. (LOT; 63).
- Butskhrikidze, Marika and Vincent J. van Heuven
 2001 Georgian harmonic clusters as complex segments. A perceptual experiment. In *Linguistics in the Netherlands* 18, T. van der Wouden and H. Broekhuis (eds.), 31–44. Amsterdam: Benjamins.
- Butskhrikidze, Marika and Jeroen M. van de Weijer
 2001 On v-metathesis in Modern Georgian. In *Surface syllable structure and segment sequencing*, E. V. Hume, N. S. H. Smith and J. M. van de Weijer (eds.), (HIL Occasional Papers; 4), 91–101. Leiden: Holland Institute of Generative Linguistics.
- Cairns, Charles C. and Mark H. Feinstein
 1982 Markedness and theory of syllable structure. *Linguistic Inquiry* 13: 193–226.
- Cho, Young-mee Yu and Tracy King
 1997 Semi-Syllables and Georgian Clusters. Handout presented at 10th Non-Slavic languages Conference. University of Chicago.
- Chomsky, Noam and Morris Halle
 1968 *The sound pattern of English*. New York: Harper & Row.
- Ertelishvili, Parnaoz
 1970 *Zmnuri pudzeebis ponemat'uri st'rukt'urisa da ist'oriis sak'itkhebi kartulshi*. [Phonematic structure and historic aspects of verbal roots in Georgian]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.
- Fischer-Jørgensen, Eli
 1975 *Trends in Phonological Theory. A historical introduction*. Copenhagen: Akademisk Forlag.
- Gamkrelidze, Thomas and Giorgi Machavariani
 1965 *Sonant'ta sistema da ablauti kartvelur enebshi*. [The system of sonants and ablaut in Kartvelian languages]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.
- Jespersen, Otto
 1904 *Lehrbuch der Phonetik*. Leipzig: Teubner.

- Jgenti, Sergi
 1956 *Kartuli enis ponetika*. [Phonetics of the Georgian language]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.
- Kenstowicz, Michael
 1994 *Phonology in generative grammar*. Oxford: Blackwell.
- Marr, Niko
 1925 *Grammatika drevneliteraturnogo gruzinskogo jazyka*. [Grammar of Old Georgian]. Leningrad.
- McCarthy, John and Alan Prince
 1995 Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*. Jill Beckman, Suzanne Urbanczyk and Laura Walsh Dickey (eds.), 249–384.
- Melikishvili, Irine
 1997 *Saerto-kartveluri jiri strukturuli da tipologiuri tvalsazrisit*. [Structural and typological aspects of the common Kartvelian roots]. Doctoral dissertation. Tbilisi.
- Nepveu, Denis
 1994 *Georgian and Bella Coola: Headless Syllables and Syllabic Obstruents*. MA thesis. University of California Santa Cruz.
- Selkirk, Elisabeth
 1982 The syllable. In *The structure of phonological representations (part II)*, H. van der Hulst and N. Smith (eds.), 337–383. Dordrecht: Foris.
- Uturgaidze, Tedo
 1976 *Kartuli enis ponematuri struktura*. [Phonematic structure of Georgian]. Tbilisi: Mecniereba.
- Vogt, Hans
 1961 *Kartuli enis ponematuri struktura*. [Phonematic structure of Georgian]. Tbilisi: Tbilisis Sakhelmc'ipo Universit'et'is Gamomcemloba.

When typological *rara* generate *rarissima*: analogical extension of verbal agreement in Dutch dialects

Gunther De Vogelaer & Johan van der Auwera

1 Introduction

In a recent paper, Siewierska (1999) points out the typological rarity of so-called “grammatical agreement” of verbs, which is defined as the type of agreement in which an inflectional person marker is accompanied obligatorily by another person marker carrying a larger referential load, usually an anaphoric element.¹ Grammatical subject agreement is found in 2 languages in the sample ($n = 272$). The rarity of the phenomenon is assumed to be due to a tendency to develop new agreement markers when the “old” ones become non-referential. Although typologically rare, grammatical subject agreement commonly occurs in the Germanic languages, such as in Dutch. Significantly, in some Dutch dialects the subject agreement morphology extends to word classes other than verbs, viz. complementizers such as *da(t)* ‘that’, *a(s)* ‘if’ etc., and to the answer particles *ja* ‘yes’ and *neen* ‘no’.

In this paper, we will discuss grammatical subject agreement of non-verbal constituents, a typological *rarissimum*. The discussion focuses on Dutch, with occasional references to the other Continental West Germanic languages where similar phenomena are found (Frisian and German). In Section 2, it will be argued that the extension of grammatical subject agreement is a (relatively) recent phenomenon. Section 3 provides a diachronic account of the pathway that has led to this extension. In Section 4, cross-linguistic parallels will be discussed for each of the phenomena that have played a role in the rise of grammatical subject agreement of non-verbal constituents. The existence of cross-linguistic parallels for each of the processes involved in the emergence of non-verbal agreement indicates that the rarity of the phenomenon results from the rarity of some of the conditions that trigger its emergence.

2 From *rarum* to *rarissimum*: agreement in Dutch dialects

In (1), instances are given of grammatical agreement in Dutch. These instances, as most data in this paper, are drawn from the Syntactic Atlas of

Dutch Dialects (henceforth: SAND), a large-scale fieldwork project carried out at several institutions in the Netherlands and Belgium. The project aimed at taking stock of syntactic variation in the Dutch language area, which consists of the Netherlands, the northern half of Belgium, and French Flanders, i. e. a small area in the north of France where some older people still speak a dialect of Dutch (see Cornips and Jongenburger 2001a, 2001b, and the SAND-atlases itself for more information on the project). In (1a), a 3PL-inflectional *-n* is obligatorily attached to the verb. As the 3PL-pronoun *ze* ‘they’ is also obligatory, the *-n* in (1a) is a relevant example of grammatical agreement. This pattern is found in Standard Dutch, and also generally in the Dutch dialects (although some dialects use a different ending than *-n*, e. g. *-ne*, *-t* or *-nt*). Example (1b–1c) shows data from the Flemish dialects, which are spoken in French Flanders and the Belgian provinces West and East Flanders. In some of these Flemish dialects, the same agreement-*n* attaches to the complementizer *da* ‘that’ (1b) or to the answer particle *ja* ‘yes’ (1c). In (1b), the complementizer *da* ‘that’ carries a 3PL agreement-*n* which is also found on the verb *gaan*. In (1c), the answer particle *ja* is followed by a 3PL *-n* and the clitic *s* ‘they’.

- (1) Grammatical agreement of complementizers and answer particles in Dutch (Flemish dialects)
 - a. *Gaa-n=ze morgen naar Gent?*
 go-3PL=they tomorrow to Ghent
 ‘Are they going to Ghent tomorrow?’
 - b. *Ze zegg-en da-n=ze naar Brussel gaa-n.*
 They say-3PL that-3PL=they to Brussels go-3PL
 ‘They say that they are going to Brussels.’
 - c. *Jaa-n=s.*
 yes-3PL=they
 ‘Yes, they do.’

Grammatical agreement is cross-linguistically rare. In Siewierska’s (1999) sample of 272 languages only two examples are found (cf. also Dahl 1990; Dryer 2005 on non-pro-drop). The phenomenon is typically found in (Western) Europe (grammatical agreement is found consistently in Dutch, English, Faroese, Frisian, French, German, Icelandic, French, and some Rhaeto-Romance dialects). Siewierska (2004: 268) mentions some non-European instances as well: four Papuan languages (Au, Ekari, Koiari and Vanimo), and three Oceanic ones (Anejom, Fehan and Labu).

Among the Western European languages with grammatical agreement, a limited number of dialects of Continental West Germanic (Dutch, Frisian and German) show complementizer agreement. Agreement of the answer particle *ja* ‘yes’ seems to be found only in Dutch.² Hence the phenomena under investigation are *rarissima*. Given that grammatical agreement as such is already a *rarum*, this is hardly surprising. It is furthermore interesting that even in contemporary varieties of Dutch, Frisian, and German, grammatical agreement of non-verbal constituents is by no means found in all of the dialects. For Dutch and Frisian, SAND (see esp. map 21) provides 155 dialects with some form of complementizer agreement, in a representative sample of 267 dialects. In addition, non-verbal grammatical agreement hardly occurs in the historical varieties of the West Germanic languages (Goeman 1997a; De Vogelaer 2005: 99–103; De Vogelaer, Devos, and van der Auwera 2006), and it is not found in historical varieties of the other Germanic languages either. This indicates that non-verbal grammatical agreement is a relatively recent extension rather than a relic from earlier stages of the (West) Germanic languages. Complementizer agreement and agreement of answer particles are therefore particularly interesting phenomena also from a diachronic point of view. Accordingly, our explanation for the existence of non-verbal agreement will focus on the diachrony of the phenomenon, more precisely on the conditions in which non-verbal agreement emerges.

3 Extension of a *rarum*: a diachronic account

3.1 The rise of complementizer agreement

The Syntactic Atlas of Dutch Dialects (SAND) provides 155 dialects with some form of complementizer agreement ($n = 267$). Example (2) repeats the data from (1b), and adds an example with a lexical subject. In both cases, the complementizer *da* ‘that’ is followed by a third person plural subject that causes an agreement-*n* to appear following *da* ‘that’.

(2) Complementizer agreement in Flemish dialects

- a. *Ze zegg-en da-n=ze naar Brussel gaa-n.*
 They say-3PL that-3PL=they to Brussels go-3PL
 ‘They say that they are going to Brussels.’
- b. *Ze zegg-en da-n die mannen ook gaa-n.*
 They say-3PL that-3PL those men also go-3PL
 ‘They say that those men are going as well.’

In the dialects in which complementizer agreement is found, it is often optional: in most dialects, the inflectional *-n* following *da* ‘that’ can be dropped. In addition, there tend to be restrictions on the subjects that cause the agreement markers to occur. For instance, complementizers are often inflected only when they are followed by clitics, as in (2a); patterns like (2b) occur in fewer dialects than (2a), and, when they occur, the agreement marker is left out more frequently (see, e. g., Hoekstra and Smits 1997: 20–21). In the majority of the complementizer agreement dialects, the phenomenon is only found for a number of grammatical persons (e. g., only for third person plural or second person singular; see SAND, map 21). The first restriction can be illustrated with a map. Figure 1 on the next page shows the data extracted from SAND-map 21, and adds whether the relevant dialects only show agreement of complementizers followed by pronouns (cf. (2a)), or also of complementizers followed by nouns (as in (2b)). Also depicted is whether these dialects have *ja*-agreement (1c). Agreement in a pronominal environment is found in many distinct regions in the Dutch-speaking area. A very restricted subset of these dialects, i. e. 31 SAND-dialects, also have complementizer agreement in clauses with nominal subjects. Agreement of the answer particle *ja* ‘yes’ is even more rare: only 20 SAND-dialects have it, all of which are dialects from the south west of the Dutch language area (where Flemish dialects are spoken).

Complementizer agreement has received a good deal of attention in the literature. In general, two types of explanations have been proposed for its existence (see De Vogelaer, Devos, and van der Auwera 2006: 215–217 for discussion). First, it has been suggested that agreement in Dutch is bound to a structural position in the clause rather than to inflected verbs (see most generative accounts, e. g. Hoekstra and Marác 1989; Haegeman 1992; Carstens 2003, Van Koppen 2005). In contrast, the inflectional *-n* following *da* has also been explained as an analogical extension from the enclitic position to the verb to other, syntactically similar environments (e. g. Kathol 2001; Zwart 2006; see also Goeman 1997b for accounts that are similar yet do not use the term analogy). The structural accounts and the analogical ones differ as to their implications for the systematicity of complementizer agreement: a structural account seems to imply that complementizers systematically behave like inflected verbs, whereas an analogical account simply states that complementizers can (but must not) take over some (but not all) of the characteristics of inflected verbs, leaving much more room for differences in the way complementizers and inflected verbs behave. Hence, the fact that com-

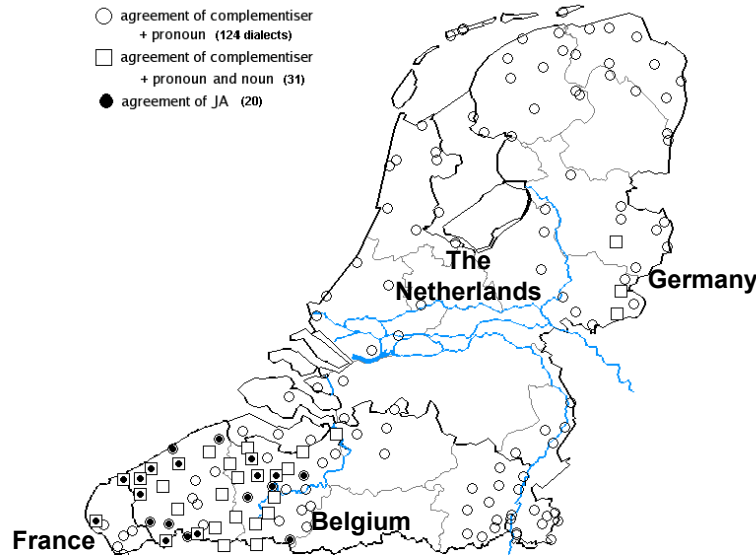


Figure 1. Non-verbal grammatical agreement in Dutch dialects (data: SAND)

plementizer agreement shows up quite sporadically in Dutch seems to provide a strong argument in favour of the analogical accounts (De Vogelaer, Devos, and van der Auwera 2006; Zwart 2006: 67–70).³

Analogical accounts differ from the structural ones in that analogical extension is not obligatory, unlike mechanisms such as the “spell-out of agreement features”. Neither is there a principled reason why the analogy would be restricted to complementizers (cf. the rise of *ja*-agreement, see Section 3.2). But like the structural accounts, analogical accounts rest on the classical observation that both complementizers and finite verbs occupy what is called the “left brace” of the brace construction. Significantly, the brace construction refers to a pattern characterising the word order of the Continental West Germanic languages, i. e. the languages in which complementizer agreement occurs. As shown in (3a–3b) on the next page, for the main clause the left brace is the finite verb and the right brace is non-finite verbal material. In main clauses, the subject is either placed in front of the left brace (SV order, see (3a)), or it appears to the right of the left brace (XVS order, as in (3b)). In clauses such as (3b), the position of the subject vis-à-vis the finite verb is

identical to its position vis-à-vis the complementizer in subordinate clauses (3c), in which the left brace is the complementizer and the right brace is verbal material (both finite and non-finite).

(3) The Dutch brace construction

		left brace					right brace	
a.	<i>Hij</i>	<i>zal</i>		<i>het</i>	<i>boek</i>	<i>morgen</i>	<i>lezen.</i>	
	He	will		the	book	tomorrow	read	
b.	<i>Morgen</i>	<i>zal</i>	<i>hij</i>	<i>het</i>	<i>boek</i>		<i>lezen.</i>	
	Tomorrow	will	he	the	book		read	
c.	<i>Ik geloof</i>	<i>dat</i>	<i>hij</i>	<i>het</i>	<i>boek</i>	<i>morgen</i>	<i>zal lezen.</i>	
	I believe	that	he	the	book	tomorrow	will read	

‘(I believe that) he will read the book tomorrow.’

The pattern in (3b) occurs very frequently in all varieties of Dutch: both De Schutter (1976: 172) for Standard Dutch and De Meersman (1985: 128) for East Flemish dialects found that around 40% of their corpus examples of declarative main clauses show XVS-word order, i. e. neither the subject nor the inflected verb occupies the first position (see Nübling 1992 for similar data for German).⁴ Furthermore, polar questions (with VS-order) and non-polar non-subject questions (with XVS-order) have postverbal subjects. This constellation, we claim, makes it possible for patterns normally appearing in clauses like (3b) to extend to clauses like (3c). More precisely, forms which are, historically speaking, exclusively found enclitic to the inflected verb in (3b) may also be adopted in the position enclitic to the complementizer. Thus several present-day pronouns are formed in the enclitic position to the verb, and subsequently extended to other positions. Example (4) shows data for the 1PL pronoun *me*, a pronoun found in southwestern dialects. It originates in the position enclitic to the verb, through fusion of the older pronoun *we* and the inflectional *-n*. Next, *me* is introduced in other syntactic environments, as shown in (4b–4c). Here too, the syntactic similarities between complementizers and inflected verbs make the complementizers the most plausible target for analogical extensions from the enclitic position to the verb. Thus, in (4b), the pronoun *me* which had emerged in clauses like (4a) is also used following complementizers instead of the etymological form *we*. (4c) shows that analogical extensions are also found elsewhere, such as in sentence-initial posi-

tion. Only a subset of the dialects in which (4b) is found also show (4c), suggesting that the use of *me* in sentence-initial position is a younger development than the use of *me* following complementizers.

- (4) Analogical extensions in Dutch: first person plural pronouns in Flemish dialects
- a. *Naar Brussel ga=me!* with $ga=me < gaa-n=we$
 to Brussels go=we go=we go-1PL=we
 ‘To Brussels we go!’
 - b. *Ik zeg da=me naar Brussel gaa-n!* historically: $da=we$
 I say that=we to Brussels go-1PL that=we
 ‘To Brussels we go!’
 - c. *Me=gaa-n liever naar Brussel.* historically: $we=gaa-n$
 we=go-1PL rather to Brussels we=go-1PL
 ‘We rather go to Brussels.’

Example (4) also suggests an explanation for the observation that complementizer agreement occurs far more often in clauses with clitic subjects than in clauses with strong pronouns or lexical elements as subjects. The fact that the 1PL inflectional ending is easily deleted when it is followed by a clitic, or fuses with the clitic, makes the 1PL form of the complementizer very similar to the 1PL form of the verbs most closely resembling the complementizer, in this case the present tense of monosyllabic verbs such as *gaan* ‘to go’ or *doen* ‘to do’. The situation in the first person plural is to a large extent representative for the entire verbal paradigm: in many dialects, verbal endings are deleted when the verb is followed by a clitic in several grammatical persons. This is illustrated with data from the dialect of Geraardsbergen, in the first two columns of (5) on the following page. The first column (*clitic* + *gaan* ‘to go’) shows the verb *gaan* ‘to go’ as it appears in clauses with SV word order (and sentence-finally in subordinate clauses): 7 of the 8 verb forms in that column carry an overt inflectional ending. The second column (*gaan* ‘to go’ + *clitic*) shows that this ending is often dropped when the verb is followed by a subject clitic. There, only 2 of the 8 verb forms carry an overt inflectional ending. Significantly, this deletion of overt inflectional endings renders the *gaan* ‘to go’ + *clitic* paradigm very similar to the combination “complementizer + clitic”, which is shown in the third column in (5). Thus, apart from structural similarities between verbs and complementizers of the type illustrated in (3), many dialects show a very clear morphological similarity between the combinations “verb + clitic” and “complementizer + clitic”, which

may very well have facilitated analogical extensions from the “verb + clitic” paradigm to the complementizer paradigm. More precisely, the insertion of the 3SG *-t* and the 3PL *-n* on the complementizer fully levels the differences between the two paradigms. Due to this complete identity with the “verb + clitic” paradigm, the complementizer paradigm from the Flemish dialect of Geraardsbergen in (5) is considered a complete paradigm, even though it only shows overt agreement in two forms (3SG.MASC and 3PL).

- (5) A complementizer agreement paradigm in the dialect of Geraardsbergen (sources: SAND and Goeman 1980: 295)

	<i>clitic + gaan</i> ‘to go’	<i>gaan</i> ‘to go’ + <i>clitic</i>	<i>da</i> ‘that’ + <i>clitic</i>
1SG	<i>k=ga</i> ‘I go’ I=go	<i>ga=ek</i> ‘go I’ go=I	<i>da=ek</i> ‘that I’ that=I
2SG	<i>ge=gaa-t</i> you=go-2	<i>ga=je</i> go=you	<i>da=je</i> that=you
3SG.MASC	<i>hij=gaa-t</i> he=go-3SG	<i>ga-t=en</i> go-3SG=he	<i>da-t=en</i> that-3SG=he
3SG.FEM	<i>ze=gaa-t</i> she=go-3SG	<i>gaa=se</i> go=she	<i>da=se</i> that=she
3SG.NEUT	<i>t=gaa-t</i> it=go-3SG	<i>gaa=t</i> go=it	<i>da=t</i> that=it
1PL	<i>me=gaa-n</i> we=go-1PL	<i>ga=me</i> go=we	<i>da=me</i> that=we
2PL	<i>ge=gaa-t</i> you=go-2	<i>ga=je</i> go=you	<i>da=je</i> that=you
3PL	<i>ze=gaa-n</i> they=go-3PL	<i>gaa-n=ze</i> go-3PL=they	<i>da-n=ze</i> that-3PL=they

In general, these paradigmatic zeros are caused by cliticization (cf. (4), where 1PL *me* is the product of a merger of 1PL *-n* and the clitic *we*). Since the occurrence of many paradigmatic zeros in the complementizer paradigm is quite typical for the dialects displaying the phenomenon, it appears to be the case that cliticization is an important factor in the emergence of complementizer agreement.

An additional and in fact stronger indication for the fact that cliticization is an important factor in the emergence of complementizer agreement is that the phenomenon typically occurs in clauses with clitic subjects. Indeed the dialects in which complementizer agreement is found in clauses with non-clitic

subjects constitute a subset of the dialects with complementizer agreement in clauses with clitic subjects (cf. Figure 1). This observation indicates that complementizer agreement in clauses with clitic subjects constitutes a necessary condition for complementizer agreement with non-clitic subjects to emerge. In a way, the emergence of complementizer agreement in clauses with clitic subjects creates a double analogical pressure to the combination “complementizer + noun”, viz. pressure from the “verb + noun”-pattern (e. g., *gaan-n dieven* ‘go thieves’) on the one hand, and from the “complementizer + clitic”-pattern (e. g., *da-n ze* ‘that they’) on the other. Diachronically, then, complementizer agreement with clitic subjects seems to be the older phenomenon. Example (6) shows the entire pathway:

(6) The emergence of complementizer agreement: a pathway

starting point:		step 1:	step 2:	
cliticization		transfer of inflection to clitic contexts	transfer to non-clitic contexts	
<i>da zij ...</i>	>	<i>da=ze ...</i>	>	<i>da-n=ze ...</i>
that they		that=they		that-3PL=they
<i>da dieven</i>		<i>da dieven</i>		<i>da-n dieven</i>
that thieves		that thieves		that-3PL thieves

The pathway in (6) reflects the diffusion of the different developments in the contemporary dialects: as one moves from the left to the right in (6), fewer dialects are found in which the relevant development is witnessed: the starting point, cliticization to complementizers, is witnessed in all Dutch dialects. The first analogical extension (step 1) is quite widespread; the second one (step 2) is rare, and is observed only in dialects in which the first extension has taken place. This can be illustrated with additional SAND-data: SAND mentions 28 dialects with a 3PL *-n* before non-clitic subjects, on a total of 66 dialects with *-n*-inflection preceding 3PL *ze*. In addition, all 28 dialects inflect complementizers when they are followed by clitics as well.

The data in (2), (4), (5) and (6) all come from the Flemish dialects, but the differences between complementizer agreement in clauses with clitic subjects and clauses with non-clitic subjects are also observed in other dialects with complementizer agreement. For instance, for the area to the north of the river Maas (in the Netherlands), SAND provides 22 dialects with complementizer agreement in the third person plural, of which only 3 show complementizer agreement in clauses with lexical subjects. Hence the pathway in (6) seems to be valid for other dialects, too.

3.2 The rise of agreement of answer particles

Focusing on agreement of the answer particle *ja* ‘yes’, Figure 1 on page 51 provides instances of overt agreement markers from 20 SAND-dialects, all dialects from French Flanders and the Belgian provinces of West and East Flanders.

(7) *ja*-agreement in Flemish dialects

Jaa-n=s
 yes-3PL=they
 ‘Yes, they do.’

In the literature, more instances of *ja*-agreement can be found from the Hollandic dialects, which are spoken in the Dutch provinces North and South Holland, at the west coast of the Netherlands (Van Ginneken 1954: 11; see also example (10) on page 58). Apart from agreement of *ja* ‘yes’, agreement of *neen* ‘no’ is found. Outside the Hollandic dialects, clear instances of *neen*-agreement are extremely rare, since the fact that *neen* ends in an /n/ in most dialects inhibits the insertion of an agreement-*n*. Like complementizer agreement, *ja*-agreement seems to originate in clauses with clitic subjects: in all dialects of Dutch, both in the contemporary and the historical sources, not a single instance of *ja*- or *neen*-agreement is found without a clitic being present as well (e.g., 3PL *jaa-n=s* exists, but *jaa-n* ‘yes-3PL’ does not). Hence the use of the clitics seems to be a necessary condition for the appearance of agreement markers, giving rise to the hypothesis that the former phenomenon predates the latter. As for the use of clitics following *ja* ‘yes’ itself, this is believed to originate from elliptical sentences in which the combination of *ja* and the clitic was followed by a verb, typically *doen* ‘to do’ or another auxiliary (see Devos 1986 for arguments). The hypothesized source construction, a pattern resembling short answers in English, is still found in some present-day Dutch dialects (see Van Craenenbroeck 2004: 225). The complete pathway for the rise of agreement markers following *ja* ‘yes’ would thus be like (8). In (8), two steps are distinguished: first, short answers develop a tendency to lose their verb, with the combination of *ja* ‘yes’ and the clitic as a result. Second, an agreement marker (in this case a 3PL *-n*) is inserted in between *ja* ‘yes’ and the clitic.

(8) The emergence of agreement of *ja* ‘yes’: a pathway

ja ze (doen) > *jaa=s(e)* > *jaa-n=s*
 yes they do yes=they yes-3PL=they

The relevant mechanism is, again, analogy. Like complementizers, answer particles which are followed by clitics are thus possible targets for analogical extensions from the enclitic position to the verb. However, answer particles are typically used in an isolated way, i.e. the clitic is not followed by any other constituents. Hence, the syntactic behaviour of answer particles differs substantially from the behaviour of verbs, and they are thus weaker targets for analogy than complementizers. This explains why *ja*-agreement is rarer than complementizer agreement. The fact that *ja*-agreement is restricted to dialects in which complementizer agreement occurs even suggests that the complementizer environment acts as a second source construction. Answer clitics' status as relatively weak targets for analogy explains why there are no dialects with a complete paradigm for *ja*-agreement. Example (9) shows the paradigms of *gaan* 'to go' + *clitic* and of *ja* 'yes' for the West Flemish dialect of Ieper. The *ja*-paradigm provides one clear instance of grammatical agreement: in the third person plural an inflectional *-n* is found. The paradigm also shows one form in which *ja*-agreement is clearly lacking, viz. *jaa=n* (lit. 'yes-he'): given that verbs receive a *-t* when they are followed by a 3SG masculine pronoun, one would expect *ja* 'yes' to show the same ending (i.e. a form like *jaa-t=n* 'yes he [does]', on the analogy of *gaa-t=n* 'goes he').

(9) Agreement of *ja* 'yes' in the West Flemish dialect of Ieper

	<i>gaan</i> 'to go' + <i>clitic</i>		<i>ja</i> 'yes'	
1SG	<i>gaa=k</i>	go=I 'I go'	<i>jaa=k</i>	yes=I 'yes I do'
2SG	<i>ga=je</i>	go=you	<i>jaa=j</i>	yes=you
3SG.MASC	<i>gaa-t=n</i>	go-3SG=he	<i>jaa=n</i>	yes=he
3SG.FEM	<i>gaa=se</i>	go-3SG=he	<i>jaa=s</i>	yes=she
3SG.NEUT	<i>gaa=t</i>	go=it	<i>jaa=t</i>	yes=it
1PL	<i>ga=me</i>	go=we	<i>jaa=m</i>	yes=we
2PL	<i>ga=je</i>	go=you	<i>jaa=j</i>	yes=you
3PL	<i>gaa-n=ze</i>	go-3PL=they	<i>jaa-n=s</i>	yes-3PL=they

(source: SAND)

Like in the previous section (cf. (5)), it should also be noted that both the paradigm of the monosyllabic verb *gaan* 'to go' + *clitic* and the paradigm of *ja* 'yes' show very few overt inflectional endings. In that sense, even without the extension of grammatical agreement, both paradigms are already very similar.

The data that were shown in (7)–(9) are exclusively Flemish, as the Flemish dialects are the only ones in which *ja*-agreement is found in present-day dialects. But the generalizations hold for other regions as well. The data in (10) shows a Hollandic example of *ja*-agreement from the literature. In these dialects, most first and third person plural verbs receive an inflectional *e*, which is found following complementizers as well. Older sources indicate that the very same *e* also attached to *ja* ‘yes’, but a /d/ appeared in between *ja* and the combination of the inflectional *e* and the clitic *we* or *ze*.

- (10) Grammatical agreement of answer particles: Hollandic dialects (Van Ginneken 1954: 11)

<i>Q:</i>	<i>Mog-e=we/ze morgen naar Gent?</i> may-1,3PL=we/they tomorrow to Ghent? ‘Can we/they go to Ghent tomorrow?’
<i>A:</i>	1PL: <i>Ja-d-e=we.</i> yes-/d/-1PL=we ‘Yes, we do.’
	3PL: <i>Jaa-d-e=ze.</i> yes-/d/-3PL=they ‘Yes, they do.’

Although the /e/ following *ja* ‘yes’ is, in diachronic terms, clearly an extension of an agreement marker, it is not clear how it should be analysed synchronically. It may be a fully functional inflectional element in its own right, like the *-n* in the Flemish 3PL-form *jaa-n=s* ‘yes they do’. But the use of a /d/ in Hollandic *jadewe* ‘yes we do’ and *jadeze* ‘yes they do’ suggests that the /e/ may also be part of an enclitic pronoun *ewe* or *eze*. /d/ is commonly used as a linking sound in the relevant dialects, but the linking /d/ remains absent in front of inflectional *e*’s which are attached to verbs and complementizers.

3.3 Summary: the diachrony of non-verbal grammatical agreement

The table in example (11) summarizes the diachrony of the two *rarissima* under investigation, *ja*-agreement and complementizer agreement. Both can be conceived of as analogical extensions of a *rarum*, grammatical verb agreement. The starting point of the development is the cliticization of pronouns

to a non-verbal constituent, resulting in a pattern which resembles the combination of the present tense of a verb followed by a clitic. Dialects differ as to the type of verbs that are taken as a model for the analogy: in the examples above, monosyllabic verbs such as *gaan* ‘to go’ are the relevant class, but there are dialects in which polysyllabic verbs act as the model. The actual development of complementizer agreement consists of two steps: first, the inflectional endings are extended from the enclitic position to the verb to other clitic contexts, through analogy. The data indicate that the transfer towards the position following the complementizer is much more likely than the transfer to the position following *ja* ‘yes’: the former extension gives rise to 3PL inflectional *-n*’s in 66 SAND-dialects; the latter in only 20, all of which show a 3PL *-n* on complementizers as well. This relates to the greater resemblance of the syntactic positions in which complementizers and verbs are placed (both appear in a clause, in the so-called ‘left brace’, whereas *ja* ‘yes’ is not followed by a clause). The second step is a further analogical extension of the use of the inflectional endings, which are introduced in non-clitic contexts. Whereas endings after step 1 are only used when they are followed by clitics, a smaller number of dialects have extended them to clauses with lexical subjects and strong pronominal subjects (“step 2”).

(11) The emergence of grammatical agreement of non-verbs: summary

	Starting point: – cliticization – elliptical <i>ja</i> -tags	Step 1: analogy 1 transfer of inflection to clitic contexts	Step 2: analogy 2 transfer to non-clitic contexts
<i>gaa-n ze</i>	> <i>gaa-n=ze</i>	> <i>gaa-n=ze</i>	> <i>gaa-n=ze</i>
<i>da ze</i>	> <i>da=ze</i>	> <i>da-n=ze</i>	> <i>da-n=ze</i>
<i>dat dieven</i>	<i>dat dieven</i>	<i>dat dieven</i>	<i>da-n dieven</i>
<i>ja ze [doen]</i>	> <i>ja=ze</i>	> <i>jaa-n=z(e)</i>	> [<i>jaa-n=z(e)</i>]

The diachronic scenario in (11) is, of course, a language-specific one, and hence it provides no answer to the question why the relevant phenomena are *rarissima*. At least two answers are conceivable to that question: it may be that the relevant diachronic processes are rare. But it may also be the case that the processes are themselves not rare, but that the conditions allowing them are only seldom met. In the next section some cross-linguistic parallels for the processes will be discussed, and we will end up arguing for the second of the two possible answers.

4 Cross-linguistic parallels

4.1 Cliticization in non-verbal environments

The starting point in the diachronic pathway for the rise of both complementizer agreement and *ja*-agreement is cliticization. The syntactic environment in which the cliticization process takes place, differs in the two cases. The clitics following complementizers can be considered second-position clitics, a type which is typologically well-attested (cf. Klavans 1985, Halpern and Zwicky 1996). The clitics following *ja* 'yes' and *neen* 'no' derive from a pattern in which *ja* 'yes' and *neen* 'no' were followed by a pronoun and a verb. Hence, from a diachronic point of view, the cliticization may have been caused by the verb that followed, although the clitics have attached phonologically to the preceding constituent, in this case the answer particles *ja* 'yes' and *neen* 'no'. Cross-linguistically, such processes are indeed attested (Klavans 1985: 104–105; cf. Bybee 2002; Cysouw 2005).⁵ A very close parallel to the Dutch development is observed in the development of the French answer particle *oui* 'yes' (which did, however, not develop agreement). French *oui* 'yes' derives from the Latin phrase *hoc ille* (lit. 'yes that'), which, in turn, is an elliptical version of *hoc ille fecit* (lit. 'yes that does'), via the intermediate form *oïl* (= *o* + 3SG *il*; von Wartburg 1922). In older varieties of French, other combinations of *o* and pronouns were found than *o-il*, such as *o-je* 'yes I' and *o-tu* 'yes you' (Robert 1985). Hence, French *oui* 'yes' developed along a similar pathway as did the combination of *ja* 'yes' and clitics found in Dutch dialects. The main difference is that French only retained *oïl* (> *oui*), i. e. the form with 3SG *il*, whereas Dutch *ja* 'yes' still combines with non-3SG clitics as well, giving rise to complete paradigms of the type exemplified in (9). It is tempting to suppose that the explanation of this similarity has an areal component: the use of *hoc ille* 'yes that' was situated in the so-called *langue d'oïl*-area, i. e. more or less the northern half of France, an area neighbouring the part of the Dutch language area where clitics following *ja* 'yes' and *neen* 'no' are found. Interestingly, the area in which the phenomenon occurred may have been much larger: clitics following answer particles were also found in Middle High German (Simon 2006: 15–16).

Apart from elliptical phrases, other source constructions are conceivable for answer particles to which clitics may attach. One does not have to look far to find cases in which the answer particle derives from a merger of sev-

eral distinct elements, i. e. a “univerbation”. A relevant example is English *yes*, which derives from Old English *gea* ‘so’ + *si* (= ‘be it’; *si* is the 3rd person imperative of *beon* ‘to be’; cf. Bosworth and Toller 1898: 438). Similarly, German *nein* ‘no’ derives from *ne* ‘not’ + *ein* ‘one’ (Köbler 1995). Neither of these patterns contains a clitic, however. In Slovene, object clitics are found in answers, which may be used in their bare form or attach to other clitics, negative particles, ... (Dvořák 2003).⁶ An instance of subject clitics following answer particles is found in Arowak, an Arawakan language mainly spoken in Suriname: in Arowak, pronouns are used in almost every *no*-answer. The answers in (12) consist of the morpheme *m(a)* and a pronoun.⁷

- (12) Arowak (Van Baarle, Sabajo, and van der Stap 1989): unmarked *no*-answers to yes / no-questions

<i>manda</i>	= ‘not-I’
<i>manba</i>	= ‘not-you’
<i>manla</i>	= ‘not-he’
<i>mantha</i>	= ‘not-she’
<i>mangwa</i>	= ‘not-we’
<i>manha</i>	= ‘not-you (plural)’
<i>mangna</i>	= ‘not-they’

Both the elliptical combinations of an answer particle and a clitic, and univerbations in which clitics are attached to non-verbal hosts constitute possible source constructions for the development of non-verbal agreement, provided that the conditions are met in which verbal endings are analogically extended to other clitic contexts than the pre- or postverbal position.

4.2 Analogical extension of agreement markers

4.2.1 Extension of agreement markers

In Section 4.1 it was stated that pronouns can cliticize to non-verbal constituents. In principle, this is a likely source for non-verbal grammatical agreement: as items that are used together tend to fuse, there is no reason why these clitics cannot turn into agreement markers in the long run (cf. the Linear Fusion Hypothesis by Bybee 2002; see Krug 1998 for a relevant case study). However, this is not the source of non-verbal grammatical agreement

in Dutch and Frisian dialects. Rather, the grammatical agreement markers on complementizers and answer particles are extensions from the position enclitic to the verb. This type of extension, too, seems to be found in other languages. First, analogical extension is necessarily involved in almost every grammaticalization from pronoun to agreement marker, as in grammaticalization in general (Hopper and Traugott 2003: 63–69). Secondly, even fully grammaticalized agreement markers can extend their use to other syntactic environments.

As for analogical extension during grammaticalization, Auger (2003) provides an example from the Picardian dialect spoken in the French Vimeu region. Some French dialects show cliticization of subject pronouns and subject doubling. In some of these varieties, all clauses contain a preverbal subject clitic, to which a semantically richer subject can be added, as shown in (13a). Example (13b) shows that the use of the preverbal clitic is extended from main clauses to clauses in which subject personal pronouns do not occur, such as subject relative clauses. This exemplifies that the clitic in (13b) can no longer be considered a mere weak variant of a strong pronoun (i. e. it has become a “special clitic” in Zwicky’s (1977) terms, or even an agreement marker).

- (13) Extension of preverbal clitics to relative clauses in Picardian dialects (Auger 2003: 6, 12)

- a. *Et pis Adof i riouot.*
and then Adolphe he_{clitic} laugh.IMP.3SG
‘And then Adolphe laughed.’
- b. *éch diabe qu’il allouot inl’veu deux piots éfants.*
the devil who he_{clitic} aller.IMP.3SG seize two little children
‘The devil who was going to seize two little children.’

Extensions such as shown in (13) form a crucial part of the grammaticalization process observed in French clitics, since they are indicative of, in Lehmann’s (1995: 164) terms, an ongoing “obligatorification”. The grammaticalization process fuses the clitic with the verb; analogy levels the differences between the syntactic environments in which the grammaticalization process is observed, such as the Picardian main clauses with pronominal subjects (13a), and the environments in which personal pronouns originally did not occur, such as Picardian relative clauses (13b).

An example of analogical extension of fully grammaticalized agreement markers comes from Russian (Dobrushina and Goussev 2005). Russian has a

1PL imperative, as in (14a), which translates as ‘let’s go’. In contrast to (14a), (14b) is marked twice for plurality: (14b) not only specifies that the subject is a first person plural, but also that the imperative is directed towards more than one addressee (or towards one addressee, expressing an honorific meaning). Thus, it also specifies the number (or honorific status) of the addressee(s). This feature is very unusual in Russian: the *pojĭ-ëm-te* construction type is actually the only one in Russian that has double plural marking. In addition, the construction is lexically very restricted. In present-day Russian it is only used with the verb *idti* ‘go’, either the bare form or some of its prefixed forms (Podlesskaya 2006: 279). The *te*-suffix is originally the 2PL imperative suffix, which has extended its use from 2PL-imperatives to give rise to a new type of imperative. Similar extensions are observed in other languages, e. g. in Turkic languages and probably in Bantu as well (see van der Auwera, Schalley, and De Vogelaer to appear).

- (14) Extension of agreement markers to imperatives in Russian (Dobrushina and Goussev 2005: 193–194)

- a. *Pojĭ-ëm!*
go.IND.PERF.FUT-1PL
‘Let’s go, me and you (singular)!’
- b. *Pojĭ-ëm-te!*
go.IND.PERF.FUT-1PL-2PL
‘Let’s go, me and you (plural / honorific)!’

Weiss (1998) provides Bavarian examples of analogical extension of agreement markers. In Bavarian, the particle *gell* ‘right’ can be inflected, as shown in (15). In (15a), the agreement marker *-ns* adds a honorific meaning to the particle; in (15b), *-ts* indicates that the use of *gell* ‘right’ is directed towards more than one addressee.

- (15) Agreement of the particle *gell* ‘right’ in Bavarian dialects (Weiss 1998; our glosses)

- a. *gell-ns*
right-2/HON
‘Right, you (honorific).’
- b. *gell-ts*
right-2PL
‘Right, you (plural).’

Agreement of *gell* ‘right’ is also discussed by Simon (2003: 188–189). According to Simon, the agreement markers following *gell* ‘right’ are identical to the ones that are used in the Bavarian imperative paradigm, i. e. paradigmatic zero expresses the second person singular, *-ns* expresses honorificity, and *-ts* expresses the second person plural. In addition, particles such as *gell* ‘right’ share with the imperative the characteristic that they are typically directed towards one or more addressees, which may or may not deserve respect. Hence, the agreement markers are likely to be taken over from the imperative. Historically, the agreement markers *-ns* and *-ts* consist of an “old” agreement marker (3PL *-n* and 2PL *-t*) and a clitic (the *-s* derives from 3PL *sie* and the Bavarian 2PL-pronoun *eß*; cf. Simon 2003: 187–188), which have fused and grammaticalized to become new agreement markers. Also, the particle *gell* ‘right’ probably derives from the verb *gelten* ‘to be worth’ (Schmeller 1872–1877). In principle, the presence of inflection following *gell* ‘right’ may be a relic from its older use as a verb rather than an extension. However, the literature on Bavarian provides other particles which do not have a verbal origin and which can also be followed by an agreement marker (e. g. *há* ‘hey’, *mei* ‘my (god)’; see Simon 2003: 188–189).

4.2.2 *Motives for analogy*

Although the Picardian dialects, Russian, and the Bavarian dialects indeed show examples of analogical extensions of agreement, the motivation for the analogy differs quite radically in each of the cases. The analogical extension of the Russian 2PL-marker *-te* is semantically motivated: in the Russian verb system, *-te* marks plurality (or honorificity) in the addressees. This is a distinction that is historically absent in inclusive first person plural imperatives like *pojd-ëm* ‘let’s go’. Adding *-te* to *pojd-ëm* creates the possibility to distinguish between one or two addressees, or between an addressee with / without an honorific status, as in the second person imperatives, which appears to be useful in imperative systems in the languages of the world (see van der Auwera, Schalley, and De Vogelaer to appear for examples). Hence, the extension of *-te* must basically be seen as the extension of a semantic distinction to the inclusive imperative, and the analogical principle at work can be described as “similar meaning seeks similar form”. In Bavarian, the same principle seems to be operating: like in the Russian example, the extension of agreement from imperatives to particles creates the possibility to distinguish

between one or two addressees, or between addressees with or without an honorific status.

In the Picardian and the Dutch examples, no semantic distinctions are added to the language. Rather, the analogical extensions reduce formal differences between the guises in which a certain linguistic element appears in different syntactic environments, by generalising one form at the expense of other ones: in Picardian dialects, the insertion of *i/il* in relative clauses mirrors the combination of the inflected verb and the clitic in main clauses; in Dutch, the insertion of agreement markers in between answer particles / complementizers and enclitics mirrors the use of agreement markers in between verbs and enclitics. Hence, in Picardian and in Dutch, the analogy seems formally driven: a similar forms seeks an even more similar form. In both cases, the guise of a linguistic element is modelled after a very frequent pattern that is used in a similar syntactic environment. The Dutch examples furthermore suggest that the more a target construction resembles the donor construction, the easier the analogy will take effect: the combinations of agreement markers and clitics first spread towards the enclitic position following the complementizer, which occupies the first brace in the Dutch clause, as does the inflected verb. Only in a few dialects with complementizer agreement does the agreement marker also show up following *ja* 'yes' and *neen* 'no'. Apart from these similarities, the Picardian and the Dutch examples differ crucially. In Picardian, the analogical extension of the "clitic + verb"-pattern to relative clauses is part of a grammaticalization process that turns pronouns such as 3SG.MASC *il* into clitics and agreement markers. In Dutch, cliticization is observed as well (and, hence, grammaticalization), but the analogy spreads elements which have already grammaticalized a long time ago, viz. the "old" agreement endings, rather than the innovative person markers, the clitics.⁸

Cross-linguistic parallels for the formally-driven analogical extension of grammatical agreement endings seem to be rare. But within the Continental West Germanic dialect continuum, even the rarest types of complementizer agreement emerge independently at different places. For the Dutch language area, Figure 1 on page 51 shows that there are two quite distinct regions with complementizer agreement in clauses with lexical subjects, viz. Flanders in the southwest of the language area, and the east of the Dutch provinces of Gelderland and Overijssel in the easternmost part of the area. This geographical pattern probably indicates that the phenomenon has originated independently in both regions. In addition, complementizer agree-

ment in clauses with lexical subjects is also found in the Bavarian dialects, see (16).

- (16) Complementizer agreement in Bavarian (Schiepek 1899, quoted in Weiss 2005):

... *dá-n d' Láit häi-en.*
 ... that-3PL the people hear-3PL
 '... that the people hear.'

The observation that non-verbal grammatical agreement originates independently in several Continental West Germanic dialects suggests that the phenomenon can hardly be considered an isolated historical accident. Rather, it looks as if the rarity of non-verbal agreement results from the rarity of some of the conditions that trigger its emergence. First, and quite trivially perhaps, a language must have grammatical agreement markers before it can extend them. This precondition is already quite rare. Second, the language must show cliticization of subject pronouns. Third, there must be similarities between the donor construction and the target construction. Among the Germanic languages with grammatical agreement, Continental West Germanic shows a stronger tendency towards cliticization than other languages, and, only in Continental West Germanic do verbs and complementizers show a comparable syntax, in that both occupy a similar position in the clause (viz. the “left brace”), and both can be followed by clitics. In addition, there is morphological similarity, as complementizers closely resemble monosyllabic verbs.

5 Conclusion

Continental West Germanic dialects have grammatical agreement, which is a *rarum* in itself. In some varieties, this *rarum* extends to other classes than verbs, viz. complementizers and the answer particles *ja* (‘yes’) and *neen* (‘no’). Grammatical agreement of such non-verbal constituents, both of complementizers and answer particles, is a *rarissimum*, which is, in addition, relatively rare in Continental West Germanic as well.

We have proposed a pathway for the emergence of non-verbal grammatical agreement, which correlates the phenomenon with cliticization. The rise of post-verbal clitics in Continental West Germanic facilitates a number of analogical extensions: first, the use of agreement markers is extended to other clitic contexts, viz. the enclitic positions following complementizers and an-

swer particles. Second, the use of agreement markers following complementizers is extended even further, to clauses with strong pronominal and lexical subjects.

In Section 4, the question was raised why the relevant phenomena are *rarissima*: it may be that the relevant diachronic processes are rare, but the rarity of non-verbal agreement may also result from the rarity of some of the conditions that trigger its emergence. We have argued for the second of the two possible answers. Indeed cross-linguistic parallels can be found for each of the processes that have played a role in the rise of grammatical agreement of non-verbal constituents, including analogical extension of agreement markers. It appears that there are different circumstances that may motivate such an analogy. In Continental West Germanic, the analogy is formally driven rather than semantically driven.

Abbreviations

1 = first person; 2 = second person; 3 = third person; FEM = feminine; FUT = future; HON = honorific; IMP = imperfective; IND = indicative; MASC = masculine; NEUT = neutral; PERF = perfective; PL = plural; SG = singular

Acknowledgments

We would like to thank the audience at the *Rara & Rarissima* Conference in Leipzig for their helpful comments, as well as Magda Devos and Horst Simon for discussing some of the ideas in this paper with us. Our gratitude also goes to everyone involved in the gathering and the transcription of the SAND-data, on which this paper is based (see also Section 2).

Notes

1. Note that the terminology in Siewierska (2004) differs from Siewierska (1999). In Siewierska (1999), the term “grammatical agreement” refers exclusively to the use of agreement markers which are obligatorily accompanied by an overt controller, whereas the notion has a broader meaning in Siewierska (2004). There, all co-occurrences of an agreement marker with an overt local controller are labelled “grammatical agreement”. Agreement markers which cannot occur without an overt controller are called “syntactic agreement markers” (cf. Siewierska 2004: 126). Hence, clauses such as Spanish *tú tienes los ojos verdes* ‘you have green eyes’, in which the subject pronoun *tú* is present but can

be dropped, represent grammatical agreement according to Siewierska (2004), but not according to Siewierska (1999). We will adopt the terminology from Siewierska (1999).

2. But Bavarian shows a similar phenomenon, viz. agreement of particles such as *gell* 'right'. See Section 4.2.
3. In most structural accounts of complementiser agreement, the differences between complementisers and inflected verbs are not even mentioned, let alone explained. One exception is Van Koppen (2005), who tries to account for the fact that in some complementiser agreement dialects the phenomenon only occurs in some grammatical persons. There is no attempt to explain the different behaviour of nouns vis-à-vis pronouns, however.
4. Neither De Schutter (1976) nor De Meersman (1985) distinguish between clauses with a nominal subject and clauses with a pronominal subject, let alone between clauses with a strong and a weak pronominal subject. For our purpose, the most relevant figure would be the proportion of main clauses (declarative, interrogative, etc.) with an enclitic subject pronoun vis-à-vis the proportion of main clauses with a proclitic subject.
5. Cysouw (2005: 32) even discusses a number of examples of languages where a situation as in Dutch, in which simple clitics incidentally lean to the "wrong" side of their host, has given rise to so-called "ditropic clitics" (Embick and Noyer 1999: 291). Ditropic clitics are clitics depending structurally and functionally of a host on their one side, but attaching phonologically to the constituent on their other side. Truly ditropic clitics, i.e. special clitics in the sense of Zwicky (1977) showing no regularity at all in their choice of a (phonological) host, appear to be rare (Cysouw 2005). But in comparison to Cysouw's (2005) examples the Dutch clitics show only minor peculiarities in the type of phonological hosts that may be selected, viz. their occurrence following the answer particles *ja* 'yes' and *neen* 'no'. Hence the Dutch clitics belong to a class of elements of which the ditropic clitics are only a (probably very small) subset. Other, comparable examples are the French articles *le* 'the (SG.MASC)' and *les* 'the (plural)' which cliticise to prepositions (e.g. *à* 'to' + *le* > *au*; *de* 'from' + *le* > *du*) and English auxiliaries cliticising to subject pronouns rather than to main verbs (e.g. *I'll go*).
6. Dvořák (2003) also discusses examples of languages in which the unmarked ways to express the meaning 'yes' or 'no' is basically an inflected verb. For instance, in Portuguese and in Welsh, (inflected) verbs are used rather than particles in answering polar questions. Apart from object clitics, Slovene also commonly uses inflected verbs. Thus, it appears to be quite natural that answers to polar questions contain person markers.
7. However, *m(a)* has a wider productivity than the meaning 'no'; for instance, it can be used as a so-called 'privative morpheme', by and large meaning 'without'. This privative meaning might also be its origin. De Goeje (1928) provides the example of *m-akosi* 'to be blind', which is derived from *m(a)* and *akosi* 'eyes'.
8. With respect to the innovative person markers, some dialects of Dutch resemble Picardian in that they also allow subject doubling. In Flemish dialects, for instance, phrases are found such as *ze zijn zij thuis* (lit. 'they are they at home') or *zijn ze zij thuis* (lit. 'are they they at home'), in which both the 3PL clitic *ze* and the strong pronoun *zij* are used. Outside the Flemish dialects, subject doubling is rare or even non-existent. See De Vogelaer and Neuckermans (2002), and De Vogelaer (2005, 2008) for further discussion.

References

- Auger, Julie
2003 Les pronoms clitiques sujets en picard: une analyse au confluent de la phonologie, de la morphologie et de la syntaxe. *Journal of French Language Studies* 13: 1–22.
- Bosworth, J. and T. N. Toller
1898 *An Anglo-Saxon Dictionary*. Oxford: Clarendon.
- Bybee, Joan
2002 Sequentiality as the basis of constituent structure. In *The Evolution of Language out of Pre-language*; T. Givón and B. Malle (eds), 109–134. Amsterdam/Philadelphia: Benjamins.
- Carstens, Vicki
2003 Rethinking complementizer agreement: Agree with a case-checked goal. *Linguistic Inquiry* 34: 393–412.
- Cornips, Leonie and Willy Jongenburger
2001a Het design en de methodologie van het SAND project. *Nederlandse Taalkunde* 16: 215–232.
- Cornips, Leonie and Willy Jongenburger
2001b Elicitation techniques in a Dutch syntactic dialect atlas project. In *Linguistics in The Netherlands 18*; Hans Broekhuis and Ton van der Wouden (eds.), 57–69. Amsterdam/Philadelphia: Benjamins.
- Cysouw, Michael
2005 Morphology in the wrong place: A survey of preposed enclitics. In *Morphology and its Demarcations*; Wolfgang Dressler, Dieter Kastovsky, Oskar Pfeiffer and Franz Rainer (eds.), 17–37. Amsterdam/Philadelphia: Benjamins.
- Dahl, Östen
1990 Standard Average European as an exotic language. In *Toward a Typology of European Languages*, Johannes Bechert, Giuliano Bernini and Claude Buri-dant (eds.), 3–8. Berlin/New York: Mouton de Gruyter.
- De Goeje, C. H.
1928 *The Arawak Language of Guiana*. Amsterdam: Koninklijke Akademie van Wetenschappen.
- De Meersman, Alfons
1985 Een VSO-maneuver in de Zuidnederlandse dialecten. In *Hulde-album Prof. dr. Marcel Hoebeke*; Hugo Ryckeboer, Johan Tældeman and Valère F. Vanacker (eds.), 123–131. Gent: R. U. G.-Seminarie voor Nederlandse taalkunde.
- De Schutter, Georges
1976 De bouw van de Nederlandse zin. Beschrijving en voorstel tot beregeling. In *Verslagen en Mededelingen van de Koninklijke Vlaamse Academie voor Nederlandse Taal- en Letterkunde* 1976: 165–282.
- De Vogelaer, Gunther
2005 *Subjectmarkering in de Nederlandse en Friese dialecten*. Doctoral dissertation, Ghent University.

- De Vogelaer, Gunther
 2008 (De)grammaticalisation as a source for new constructions: the case of subject doubling in Dutch. In *Constructions and Language Change*; Alexander Bergs and Gabriele Diewald (eds.), 229–257. Berlin/New York: Mouton de Gruyter.
- De Vogelaer, Gunther, Magda Devos, and Johan van der Auwera
 2006 Voegwoordvervoeging: morfologisch of syntactisch verschijnsel? *Taal and Tongval*, Theme issue 19 (Het morfologische landschap van het Nederlands): 212–230.
- De Vogelaer, Gunther and Annemie Neuckermans
 2002 Subject doubling in Dutch: a dialect phenomenon in typological perspective. *Sprachtypologie und Universalienforschung* 55: 234–258.
- Devos, Magda
 1986 Het persoonlijk voornaamwoord 2e pers. enk. in het Westvlaams: geografie en historiek. In *Vruchten van zijn akker: Opstellen van (oud-) medewerkers en oud-studenten voor Prof. dr. V.F. Vanacker*. Magda Devos and Johan Taeldeman (eds.), 167–191. Gent: Seminarie voor Nederlandse Taalkunde en Vlaamse Dialectologie.
- Dobrushina, Nina and Valentin Goussev
 2005 Inclusive imperative. In *Clusivity. Typology and Case Studies of the Inclusive-Exclusive Distinction*; Elena Filimonova (ed.), 179–211. Amsterdam/Philadelphia: Benjamins.
- Dryer, Matthew S.
 2005 Expression of pronominal subjects. In *The World Atlas of Language Structures*; Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 410–415. Oxford: Oxford University Press.
- Dvořák, Boštjan
 2003 Elliptische Prädikatisierung enklitischer Personalpronomina im Slowenischen. *Philologie im Netz* 26: 37–61.
- Embick, David and Rolf Noyer
 1999 Locality in post-syntactic operations. In: *Papers on Morphology and Syntax, Cycle One*; Karlos Arregi, Benjamin Bruening, Cornelia Krause, and Vivian Lin (eds.), 265–316. (MIT Working Papers in Linguistics 33.) Cambridge, Mass.: MIT Department of Linguistics and Philosophy.
- Goeman, Ton
 1980 COMP-agreement? In: *Linguistics in the Netherlands 1979*; Wim Zonneveld and Fred Weerman (eds.), 291–306. Dordrecht: Foris.
- Goeman, Ton
 1997a De zeldzaamheid van Comp-agreement in taaltypologisch en historisch opzicht: Voorkomen buiten de Germania en datering van voegwoordcongruentie-vormen in het Nederlands. In *Vervoegde voegwoorden*; Eric Hoekstra and Caroline Smits (eds.), 87–111. Amsterdam: P. J. Meertensinstituut.
- Goeman, Ton
 1997b Historiografie van het onderzoek naar voegwoordvervoeging: een bibliografisch overzicht (1821–1996). In *Vervoegde voegwoorden*; Eric Hoekstra and Caroline Smits (eds.), 112–145. Amsterdam: P. J. Meertensinstituut.

- Haegeman, Liliane
 1992 *Theory and Description in Generative Syntax: A Case Study in West Flemish*. Cambridge: Cambridge University Press.
- Halpern, Aaron and Arnold Zwicky
 1996 *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford: CSLI Press.
- Hoekstra, Eric and Caroline Smits
 1997 Vervoege voegwoorden in de Nederlandse dialecten: een aantal generalisaties. In: *Vervoege voegwoorden*; Eric Hoekstra and Caroline Smits (eds.), 6–30. Amsterdam: P. J. Meertensinstituut.
- Hoekstra, Jarich and Laci Maracz
 1989 Some implications of I-to-C-movement in Frisian. In *Linguistics in the Netherlands 1989*; Hans Bennis and Ans van Kemenade (eds.), 81–90. Dordrecht: Foris.
- Hopper, Paul and Elizabeth Traugott
 2003 *Grammaticalization*. 2nd edition. Cambridge: Cambridge University Press.
- Kathol, Andreas
 2001 Syntactic categories and positional shape alternations. *Journal of Comparative Germanic Linguistics* 3: 59–96.
- Klavans, Judith L.
 1985 The independence of syntax and phonology in cliticization. *Language* 61: 95–120.
- Köbler, Gerhard
 1995 *Deutsches Etymologisches Wörterbuch*. <http://homepage.uibk.ac.at/~c30310/derwbhin.html>
- Krug, Manfred
 1998 String frequency: a cognitive motivating factor in coalescence, language processing and linguistic change. *Journal of English Linguistics* 26: 286–320.
- Lehmann, Christian
 1995 *Thoughts on Grammaticalization*. Munich: LINCOM Europa.
- Nübling, Damaris
 1992 *Klitika im Deutschen. Schriftsprache, Umgangssprache, alemannische Dialekte*. Tübingen: Narr.
- Podlesskaya, Vera I.
 2006 Auxiliation of ‘give’ verbs in Russian. Discourse evidence for grammaticalization. In *Voice and Grammatical Relations. In Honor of Masayoshi Shibatani*; Tasaku Tsunoda and Taro Kageyama (eds.), 271–298. Amsterdam/Philadelphia: Benjamins.
- Robert, Paul
 1985 *Le Grand Robert de la langue française. Dictionnaire alphabétique et analogique de la langue française*. (second edition, revised by Alain Rey). Paris: Le Robert.

SAND

- 2005 Barbiers, Sjef, Hans Bennis, Gunther De Vogelaer, Magda Devos, and Margreet van der Ham: *Syntactic Atlas of Dutch Dialects. Volume 1: Pronouns, Agreement and Dependencies (maps + commentary)*. Amsterdam: Amsterdam University Press.

Schiepek, Josef

- [1899] 1908 *Der Satzbau der Egerländer Mundart (2 Teile)*. Prag: Verlag des Vereines für Geschichte der Deutschen in Böhmen.

Schmeller, Johann Andreas

- 1872–1877 *Bayerisches Wörterbuch*. München: Oldenbourg.

Siewierska, Anna

- 1999 From anaphoric pronoun to grammatical agreement marker: why objects don't make it. *Folia Linguistica* 33: 225–251.

Siewierska, Anna

- 2004 *Person*. Cambridge: Cambridge University Press.

Simon, Horst J.

- 2003 *Für eine grammatische Kategorie "Respekt" im Deutschen. Synchronie, Diachronie und Typologie der Deutschen Anredepronomina*. Tübingen: Niemeyer.

Simon, Horst J.

- 2006 Reconstructing historical orality in German — what sources should we use? In *Dialogic Language Use. Dimensions du Dialogisme. Dialogischer Sprachgebrauch*; Irma Taavitsainen, Juhani Härmä und Jarmo Korhoen (eds.), 7–26. Helsinki: Société Néophilologique.

Van Baarle, P., M. A. Sabajo, and G. van der Stap

- 1989 *Arhwaka Lokonong Djang: Arowakse Taalkursus en Woordenboek*. Amsterdam: Instituut voor Algemene Taalwetenschap; Haarlem: Sociaal-culturele vereniging Ikyoshie.

Van Craenenbroeck, Jeroen

- 2004 *Ellipsis in Dutch dialects*. Utrecht: LOT.

van der Auwera, Johan, Ewa Schalley and Gunther De Vogelaer

- to appear Analogie und die Verbreitung der verbalen Kongruenz bei Imperativen, Konjunktionen und Antwortpartikeln. In *Verstärkung*; Rudiger Harnisch (ed.). Berlin: Walter de Gruyter.

Van Ginneken, Jac.

- 1954 *Drie Waterlandse dialecten. Deel I: Grammatica, Phonologie, Klankleer* (ed. by A. Weynen). Alphen a/d Rijn: Samson.

Van Haeringen, Coenraad B.

- 1939 Congruerende voegwoorden. *Tijdschrift voor Nederlandse taal- en letterkunde* 58: 161–176.

Van Koppen, Marjo

- 2005 *One Probe — Two Goals: Aspects of Agreement in Dutch Dialects*. Utrecht: LOT.

Wartburg, Walther von

- 1922 *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes*. Bonn: Schroeder.

- Weiss, Helmut
1998 *Syntax des Bairischen. Studien zur Grammatik einer natürlichen Sprache*. Tübingen: Niemeyer.
- Weiss, Helmut
2005 Inflected complementizers in Continental West Germanic dialects. *Zeitschrift für Dialektologie und Linguistik* 72: 148–166.
- Zwart, Jan-Wouter
2006 Complementizer agreement and dependency marking typology. *Leiden Working Papers in Linguistics* 3.2: 53–72.
- Zwicky, Arnold
1977 *On Clitics*. Bloomington: Indiana University Linguistics Club.

Social-deixis classifiers in Weining Ahmao

Matthias Gerner & Walter Bisang

1 Introduction

The findings presented on Ahmao classifiers at the “*Rara & Rarissima*” conference in Leipzig in 2006 generated a stream of successive works on different aspects of the system: pragmatic (see Gerner and Bisang 2008), historical (see Gerner and Bisang 2009) and documentary (see Gerner and Bisang 2010). In this paper for the proceedings we wish to recapitulate the main points.

Weining Ahmao belongs to the Miao-Yao language family whose speakers are spread over several countries in East and Southeast Asia. As an analytic language, Ahmao has developed a highly uncommon classifier system that is unattested even by its closest genetic relatives. Whereas classifiers across the region generally constitute a one-form word category with occasional sandhi-derivations, in Ahmao each of its about 50 classifiers can be inflected in a 12-form paradigm built on three semantic categories: number [singular, plural], definiteness [definite, indefinite] and size / importance [augmentative, medial, diminutive]. In addition to the wealth of intrinsic meanings they provide, Ahmao classifiers also exhibit a rare form of social deixis in that they are gender-age indices of the speaker. The augmentative form prototypically points to male speakers, the medial to female speakers and the diminutive to child speakers. Variations from this pattern are allowed, but produce certain pragmatic implicatures.

Quite generally, classifier systems in East and Southeast Asia establish a “classification of experience” (Lucy 2000: 326), a categorization of linguistic forms that mirrors a classification of referents. A classification of experience is arranged in languages with classifier systems along three parameters: *animacy*, *physical property* and *functionality* (cf. Aikhenvald 2000; Denny 1976; Adams and Conklin 1973; Allan 1977; Dixon 1982; Croft 1994; Bisang 1999).

In addition to categorization, classifiers assume the function of individuation in that they highlight spatial-temporal boundaries of the concept expressed by the noun (Bisang 1999, 2002). As was pointed out by Greenberg

(1974), nouns are neutral with regard to number in languages with numeral classifiers. Since this property generally applies to nouns in East and mainland Southeast Asian languages, most of them must use classifiers in numeral constructions for individuating the concept expressed by the noun that has to be counted. Categorization and individuation are the core, but not the only, functions assumed by classifiers. We typically find classifiers in (in)definite, relative and possessive constructions, although languages may vary in the degree to which they require classifiers for these grammatical functions. Table 1 on the facing page provides an overview of the non-core functions of classifiers, as well as examples of the East Asian languages in which these functions are attested.

The Ahmao classifiers are involved in expressing each of these functions. Additionally, they convey an array of intrinsic and social-deictic meanings not attested in any of the languages of East Asia. The Ahmao classifier system reflects social aspects of the speaker in terms of gender and age, but it should not be confounded with a gender system. Even though gender and classifier systems both have the function of noun categorization, gender systems differ from other classifier systems by the following two properties (Aikhenvald 2000: 4; Allan 1977: 286; Corbett 1991: 136–137; Dixon 1982: 212–218):

- gender systems are agreement systems in which nouns agree morphologically with their collocating modifiers or predicates / verbs, whereas classifiers are independent morphemes (often derived from bleached nouns);
- the number of genders in a language is generally below 20, whereas the number of classifiers may be more than 30.

The reason for linking *gender* to agreement systems is historically conditioned, as some of the languages in which this phenomenon was first scrutinized classify experience predominantly in terms of gender: e. g., male / female or male / female / neuter (e. g. de la Grassière 1898).

The classifier system of Ahmao is unique in that it exhibits a certain type of social deixis that is determined by the speaker's identity in terms of gender [male vs. female] and age [child vs. adult]. *Social deixis* refers to the lexicalized or grammaticalized means a language deploys to display the social status of speech participants (e. g. Levinson 1983: 89). Social deixis has been explored in linguistics within theories of politeness, although it should be differentiated from the notion of *volitional politeness*, which reflects the decisions the speaker makes to attenuate the performance of face-threatening

Table 1. Functional range of classifiers in the isolating languages of East Asia

Concept	Function of Classifier	Structure	Languages
Indefiniteness	Classifier functions as or contributes to the formation of an indefinite article	(i) NUM:1+CL+N	Mandarin (Li and Thompson 1981), Vietnamese (Daley 1998), Kam-Tai (Gerner 2006), many Tibeto-Burman
		(ii) CL+N	
		(iii) N+CL	
Definiteness	Classifier functions as or contributes to the formation of a definite article	(i) N+CL+NOM	Nuosu (TB) (Gerner 2003: 992), Vietnamese (Daley 1998), Lolo (TB) (Gerner 2003: 993–994)
		(ii) N+CL	
		(iii) N+GCL	
Deixis and anaphora	Classifier collocates with demonstrative pronouns	(i) DEM+CL+N	Mandarin (Li and Thompson 1981), Kam-Tai (Conklin 1981; Gerner 2006), Miao (Bisang 1993; Gerner 2006), Vietnamese (Daley 1998), many Tibeto-Burman
		(ii) CL+N+DEM	
		(iii) N+DEM+CL	
Nominalization	Classifier collocates with verb phrases	CL(+N)+VP+DEM	Kam-Tai (Gerner 2006)

acts. Social deixis operates independently of any communication goal on the part of the speaker. Authors define social deixis as “social warrants” (Kasper 1990: 196) or “social entitlements” (Kochman 1984: 202) to which individuals lay claim in social interaction. It may pertain to inherent social characteristics (such as gender, age or kinship positions) or to acquired social features (such as rank or title); cf. Shils (1968).

The Ahmao classifiers appeal to inherent social characteristics. Historically, this system is derived from the merger of earlier classifier versions with the two size prefixes *a⁵⁵nie⁵³* ‘augmentative / female’ and *ɲoa¹¹* ‘small’. This type of merger occurs in no other genetically related language, although cognate augmentative and diminutive prefixes are attested in other Miao languages.

In Section 2, we present a short synchronic sketch of the Ahmao classifier system (cf. Gerner and Bisang 2008, 2010), and in Section 3, we hypothesize about the historical factors that were involved in the shift from a one-form to a multi-form classifier system (cf. Gerner and Bisang 2009).

2 Traits of Ahmao classifiers

2.1 Inflectional traits

Each of the 48 Ahmao classifiers can be inflected in a 12-form paradigm comprising six singular and six plural forms. For both singular and plural classifiers, there are three definite and three indefinite forms, respectively, thus resulting in four triples: singular-definite, singular-indefinite, plural-definite, plural-indefinite. In each of these subsets, three classifier versions related to size values may be distinguished: augmentative, medial and diminutive. At the same time, these size attributes correlate with the gender / age of the speaker. Augmentative classifiers are pragmatically unmarked when employed by men, medial classifiers are pragmatically unmarked when used by women, and diminutive classifiers are mainly reserved for children. When the size value and the type of speaker differs from this pattern, some kind of pragmatic implicature is triggered, as described in Section 2.3. For each of the singular and plural classifiers, the inflectional paradigm may be outlined as follows. The plural forms of each classifier are shaped by means of the plural / mass quantifier *ti⁵⁵*, which may also be declined in the same type of paradigm (although there are only six forms instead of twelve):¹

Table 2. Paradigm of plural morpheme

Gender / Age Register of Speaker	Size	Plural	
		Definite	Indefinite
Male	Augmentative	<i>ti</i> ⁵⁵	<i>di</i> ³¹
Female	Medial	<i>tiai</i> ⁵⁵	<i>diai</i> ²¹³
Child	Diminutive	<i>tia</i> ⁵⁵	<i>dia</i> ⁵⁵

Table 3. General form of classifier paradigms

Gender / Age Register	Size	Singular		Plural	
		Definite	Indefinite	Definite	Indefinite
Male	Augmentative	CVT	C*VT	<i>ti</i> ⁵⁵ <i>a</i> ¹¹ CV ¹¹	<i>di</i> ³¹ <i>a</i> ¹¹ C*V ¹¹
Female	Medial	<i>Cai</i> ⁵⁵	C* <i>ai</i> ²¹³	<i>tiai</i> ⁵⁵ <i>a</i> ¹¹ CV ¹¹	<i>diai</i> ²¹³ <i>a</i> ¹¹ C*V ¹¹
Child	Diminutive	<i>Ca</i> ⁵³	C* <i>a</i> ³⁵	<i>tia</i> ⁵⁵ <i>a</i> ¹¹ CV ¹¹	<i>dia</i> ⁵⁵ <i>a</i> ¹¹ C*V ¹¹

Remarks:

C means “consonant” (simple, double, affricated, etc.)

V means “vowel” (monophthong, diphthong)

T means “tone” using numbers 1–5 to indicate the pitch contours

***** means “suprasegmental phenomenon” (e. g., breathy voicing), but possibly also the absence of a sound change

Other classifiers prefix a bleached version of the quantifier *ti*⁵⁵ together with the morpheme *a*¹¹ before each plural version. The general sound structure of Ahmao inflected classifiers is outlined in Table 3. We write the singular-augmentative-definite form as CVT, which is taken as the basic form of the system. All medial / female forms employ the rhyme [ai] and tend to exhibit the [55] tone for the definite, and the [213] tone for the indefinite, classifier versions. Diminutive / child forms display the rhyme [a] together with the [53] tone for definite and the [35] tone for indefinite forms. Additionally, suprasegmental processes such as breathy voicing, voicing / devoicing or tone changes may accompany the formation of the indefinite versions. In Table 3, these processes are marked by an asterisk.

The paradigm of the quantifier *ti*⁵⁵ given in Table 3 derives indefinite from definite classifiers through a voicing process. The same suprasegmental process also underlies the animate classifier *tu*⁴⁴, whose declined forms are displayed in Table 4 on the next page.

Table 4. Paradigm of animate classifier

Gender / Age Register	Size	Singular		Plural	
		Definite	Indefinite	Definite	Indefinite
Male	AUG	<i>tu</i> ⁴⁴	<i>du</i> ³¹	<i>ti</i> ⁵⁵ <i>a</i> ¹¹ <i>tu</i> ⁴⁴	<i>di</i> ³¹ <i>a</i> ¹¹ <i>tu</i> ⁴⁴
Female	MED	<i>tai</i> ⁴⁴	<i>dai</i> ²¹³	<i>tiai</i> ⁵⁵ <i>a</i> ¹¹ <i>tu</i> ⁴⁴	<i>diai</i> ²¹³ <i>a</i> ¹¹ <i>tu</i> ⁴⁴
Child	DIM	<i>ta</i> ⁴⁴	<i>da</i> ³⁵	<i>tia</i> ⁵⁵ <i>a</i> ¹¹ <i>tu</i> ⁴⁴	<i>dia</i> ⁵⁵ <i>a</i> ¹¹ <i>tu</i> ⁴⁴

Table 5. Paradigm of classifier for plants

Gender / Age Register	Size	Singular		Plural	
		Definite	Indefinite	Definite	Indefinite
Male	AUG	<i>dzfiau</i> ³⁵	<i>dzau</i> ⁴⁴	<i>ti</i> ⁵⁵ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵	<i>di</i> ³¹ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵
Female	MED	<i>dzfai</i> ²¹³	<i>dzai</i> ²¹³	<i>tiai</i> ⁵⁵ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵	<i>diai</i> ²¹³ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵
Child	DIM	<i>dzfia</i> ³⁵	<i>dza</i> ³⁵	<i>tia</i> ⁵⁵ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵	<i>dia</i> ⁵⁵ <i>a</i> ⁴⁴ <i>dzfiau</i> ³⁵

Table 6. Paradigm of classifier for clothes

Gender / Age Register	Size	Plural		Plural	
		Definite	Indefinite	Definite	Indefinite
Male	AUG	<i>pho</i> ⁵⁵	<i>pho</i> ¹¹	<i>ti</i> ⁵⁵ <i>a</i> ⁴⁴ <i>pho</i> ⁵⁵	<i>di</i> ³¹ <i>a</i> ⁵⁵ <i>pho</i> ⁵⁵
Female	MED	<i>phai</i> ⁵⁵	<i>phai</i> ²¹³	<i>ti</i> ⁵⁵ <i>a</i> ⁴⁴ <i>pho</i> ⁵⁵	<i>diai</i> ²¹³ <i>a</i> ⁵⁵ <i>pho</i> ⁵⁵
Child	DIM	<i>pha</i> ³⁵	<i>pha</i> ³⁵	<i>ti</i> ⁵⁵ <i>a</i> ⁴⁴ <i>pho</i> ⁵⁵	<i>dia</i> ⁵⁵ <i>a</i> ⁵⁵ <i>pho</i> ⁵⁵

A number of classifiers with breathy voicing for the definite classifiers involve non-breathy voicing for the indefinite-singular forms. This is the case, for example, for the plant classifier *dzfiau*³⁵ ‘bunch’, although its plural forms do not represent this sound change, cf. Table 5

For a certain number of classifiers, a tone change alone distinguishes indefinite from definite classifiers. The classifier *pho*⁵⁵ for clothes, for example, exhibits a high tone for the augmentative-definite and a low tone for the augmentative-indefinite classifier version. All forms of this classifier are aspirated, cf. Table 6.

2.2 Functional traits

The classifiers in Weining Ahmao cover a wide range of functions — more than in other East Asian languages. The attested functions comprise those

mentioned in Section 1, namely, individuation (numeral / quantifier constructions), deixis (demonstratives), (in)definiteness, relativization, and possession. Classifiers are obligatory in numeral constructions, but numerals may co-occur only with the indefinite classifier versions as in (1b), not with the definite forms exemplified in (1a).

(1) Numeral Constructions

- a. * i^{55} **dlɿ**³⁵ ndzao⁴⁴
 NUM:1 CL:AUG:SG:DEF rice
 ‘the (sole) rice grain’
- b. i^{55} **dlɿ**³⁵ ndzao⁴⁴
 NUM:1 CL:AUG:SG:INDEF rice
 ‘one rice grain’

With demonstrative pronouns, the definite classifier version alone may be involved, as in (2a), rather than the indefinite classifier form illustrated in (2b).

(2) Demonstrative Pronouns

- a. **dlɿ**⁵³ mau⁴⁴ jɿ⁵⁵
 CL:AUG:SG:DEF letter DEM:PROX
 ‘this letter’
- b. ***dlɿ**⁵³ mau⁴⁴ jɿ⁵⁵
 CL:AUG:SG:INDEF letter DEM:PROX
 ‘this letter’

Definite classifier forms that modify a head noun alone are interpreted as definite articles, cf. (3a), and indefinite forms as indefinite articles, cf. (3b).

(3) a. Definiteness

ŋkey⁵³ tie⁵⁵
 CL:AUG:SG:DEF skirt
 ‘the skirt’

b. Indefiniteness

ŋgey³¹ tie⁵⁵
 CL:AUG:SG:INDEF skirt
 ‘a skirt’

Relative clauses, too, involve classifiers, as illustrated in the following example in which the modifying clause is framed by a classifier and an optional demonstrative pronoun.

(4) Relativization (Wang 1986: 71)

$\eta u^{31} ntsa^{55} tsa^{55} pfi^{11} [_{REL} ts\dot{r}^{55} a^{55} v\partial^{55} ku^{11} tau^{33}$
 NAME regret 3P.SG [_{REL} CL:AUG:SG:DEF precious_stone COP obtain
 $lfi^{35} saw^{33} \dot{r}^{55}]$
 lose drop DEM:S&A:FAM]
 ‘Ngucha felt sorry about the precious stone she lost.’

Like other Miao languages, in Ahmao a classifier can be used to mark the possessee in a possessive noun phrase. However, Ahmao differs from other Miao languages in that it does not systematically require a classifier in these instances. In (5), the classifier can be omitted with little change in meaning.

(5) Possession / Association

$ku^{55} ts\dot{r}^{55} a^{11} zfiy^{11}$
 1P.SG CL:AUG:SG:DEF grandfather
 ‘my grandfather’

2.3 Semantic and deictic traits

In the previous section, we listed an array of morphosyntactic constructions that require the presence of classifiers. Each time classifier forms are projected in a construction, they contribute a size-related meaning to the description of the referent and index the speaker by gender and age.

In this section, we provide various examples of this, both elicited sentences and clauses retrieved from folk stories.

The first three sentences – examples (6) on the facing page through (8) on page 84 – are elicited and illustrate now for different speaker roles the size versions of a classifier qualify the referent of the head noun. For each classifier version, there is one speaker role that is canonically associated with it in the sense that if uttered by the relevant speaker then the classifier phrase is pragmatically unmarked. If uttered by an atypical speaker, then a pragmatic implicature is triggered. In (6a) on the facing page, the augmentative-definite classifier qualifies the head noun ‘hill’ as being large in size, and it appears pragmatically unmarked only if it is uttered by a male speaker. In (6b), the medial-definite classifier conveys the medial size, and it is employed in a pragmatically neutral way by women. The version in (6c) describes the referent with a reduced size, and it is used only by child speakers in a pragmatically neutral way.

- (6) a. lu^{55} tau^{55} pi^{55} ku^{11} lu^{55} tau^{55} zau^{44} .
 CL:AUG:SG:DEF hill DEM:PROX COP CL:AUG:SG:DEF hill nice
 Male speaker: 'This big hill is nice.' [Pragmatically neutral]
 Female speaker: 'This big hill is nice.' [Audacious, daring or boyish tone]
 Child speaker: 'This big hill is nice.' [Pragmatically odd]
- b. la^{55} tau^{55} pi^{55} ku^{11} lu^{55} tau^{55} zau^{44} .
 CL:MED:SG:DEF hill DEM:PROX COP CL:AUG:SG:DEF hill nice
 Male speaker: 'This hill is nice.' [Modest tone]
 Female speaker: 'This hill is nice.' [Pragmatically neutral]
 Child speaker: 'This hill is nice.' [Pragmatically neutral]
- c. la^{53} tau^{55} pi^{55} ku^{11} lu^{55} tau^{55} zau^{44} .
 CL:DIM:SG:DEF hill DEM:PROX COP CL:AUG:SG:DEF hill nice
 Male speaker: 'This small hill is nice.' [Imitating children]
 Female speaker: 'This small hill is nice.' [Pragmatically almost neutral]
 Child speaker: 'This (small) hill is nice.' [Pragmatically neutral]

If the speaker describes a personal attribute of another person by employing an irrelevant form of the classifier, then this may display a disrespectful attitude. An adult male speaker who refers to someone's character using a diminutive classifier version, as in (7c), reveals his poor view of that person. As noted in Section 2.2, the classifier modifying the possessee is not obligatory and would be omitted in most circumstances.

- (7) a. $ɲfiy^{11}$ tu^{44} pi^{55} lau^{53}
 3P:SG CL:AUG:SG:DEF character
 Male speaker: 'His character' [Pragmatically neutral]
 Female speaker: 'His character' [Respectful and also slightly boyish]
 Child speaker: 'His character' [Pragmatically odd]
- b. $ɲfiy^{11}$ ta^{44} pi^{55} lau^{53}
 3P:SG CL:MED:SG:DEF character
 Male speaker: 'His character' [Loose way of talking]
 Female speaker: 'His character' [Appropriate tone]
 Child speaker: 'His character' [Pragmatically almost appropriate]
- c. $ɲfiy^{11}$ ta^{44} pi^{55} lau^{53}
 3P:SG CL:DIM:SG:DEF character
 Male speaker: 'His character' [Impolite way of talking]
 Female speaker: 'His character' [Pragmatically almost appropriate]
 Child speaker: 'His character' [Appropriate tone]

Similar size-related connotations and pragmatic nuances exist for indefinite versions of a classifier. In the following examples, the human classifier lu^{55} has indefinite forms whose tones change relative to the definite forms. (Note also that the human classifier has medial and diminutive forms that

overlap with those of the general inanimate classifier *lu*⁵⁵ for both the definite and indefinite form).

- (8) a. *saur*⁵⁵ *bfi*³¹ *tu*³³ *hi*³³ *mfi*³⁵ *lu*⁴⁴ *tur*⁵⁵ *nur*⁵⁵.
 side DEM:MED:HIGH all NEG have CL:AUG:SG:INDEF person
 Male speaker: ‘In that place there isn’t anybody.’ [Pragmatically neutral]
 Female speaker: ‘In that place there isn’t anybody.’ [Imitating a boyish or mocking tone]
 Child speaker: ‘In that place there isn’t anybody.’ [Pragmatically odd]
- b. *saur*⁵⁵ *bfi*³¹ *tu*³³ *hi*³³ *mfi*³⁵ *la*²¹³ *tur*⁵⁵ *nur*⁵⁵.
 side DEM:MED:HIGH all NEG have CL:MED:SG:INDEF person
 Male speaker: ‘In that place there isn’t anybody.’ [Modest tone]
 Female speaker: ‘In that place there isn’t anybody.’ [Pragmatically neutral]
 Child speaker: ‘In that place there isn’t anybody.’ [Pragmatically neutral]
- c. *saur*⁵⁵ *bfi*³¹ *tu*³³ *hi*³³ *mfi*³⁵ *la*³⁵ *tur*⁵⁵ *nur*⁵⁵.
 side DEM:MED:HIGH all NEG have CL:DIM:SG:INDEF person
 Male speaker: ‘In that place there isn’t anybody.’ [Pragmatically odd]
 Female speaker: ‘In that place there isn’t anybody.’ [Pragmatically acceptable]
 Child speaker: ‘In that place there isn’t anybody.’ [Pragmatically acceptable]

If a speaker wants to convey a size value for some referent that does not correspond to his or her gender/age register (e.g., an adult male speaker wishing to feature a small-sized entity), then he or she may either choose a classifier with the relevant size value, thereby generating a pragmatic effect, or add an adjective of size, thus erasing the size information included in the classifier.

In narratives, various forms of classifiers are attested. Examples (9)–(12) stem from Wang’s lengthy folk story “Ngucha and Daushenau” which is transcribed in the International Phonetic Alphabet (Wang 1986). Because the speaker’s situation is disconnected from that of the story, the classifiers do not necessarily reflect the narrator’s gender and age. However, if direct discourse is used in the narrative, then the classifiers index the gender and age of the speaker who is a participant. Moreover, the classifier forms also do not seem to agree with the gender/age of the story participants, at least not in any statistically significant way (cf. Gerner and Bisang 2010). The two following examples employ augmentative and diminutive versions of the general classifier for round entities.

- (9) ... *sey*⁵⁵ *vfi*³¹ *za*⁵³ *zo*³³ ***lu*⁵⁵** *gfiau*³¹ *a*³³ *dlfi*³⁵
 stand outside home neighbor CL:AUG:SG:DEF outer door
 '[A man] stood outside the neighbor's house in front of the outer door.'
- (10) *si*⁵⁵ *fo*³³ *qaur*³¹ *to*⁵⁵ *lho*¹¹ *k^hey*¹¹ *ŋu*³¹ *ntsa*⁵⁵ *vai*¹¹ *taur*³³
 turn face run return come COV:take NAME hide be_at
***la*³⁵** *ŋgfia*³⁵ *tɕy*⁵⁵ *ly*⁵⁵
 CL:DIM:SG:DEF house straw
 '[Daushenau] turned his face, ran back and hid Ngucha in the straw house.'

Example (11) is uttered by a female story participant (Ngucha) and involves the medial version of the classifier of animacy. We have a correlation of speaker register and size parameter of the classifier system.

- (11) *ŋu*³¹ *ntsa*⁵⁵ *sey*⁵⁵ *ntla*⁵⁵ *xu*⁵⁵: “*tsha*³³ *ŋtshai*³³ *za*³¹, ***da*²¹³**
 NAME rise call shout terrible EXCL CL:MED:SG:INDEF
*pi*⁵⁵ *tlaur*⁵⁵ *qa*³³ *sa*³³ *dfi*³⁵ *dfir*³⁵ *dau*¹¹ *za*³¹.”
 ghost INT:which come opposite_side DP EXCL
 ‘Ngucha stood up and cried: “Terrible! What ghost is coming from the opposite side?!”’

A noun phrase that is modified by the universal distributive quantifier *tsha*³³ ‘every’ is always interpreted as a specific indefinite and hence requires an indefinite classifier form. Example (12) employs the augmentative indefinite form of the human classifier whose tone is lowered. The augmentative form correlates with the gender and age of the zero-anaphora head noun ‘soldiers’.

- (12) *tsha*³³ *lu*⁴⁴ *hi*³³ *tau*³³ *ki*⁵⁵ *a*³³ *dzu*³³ ...
 every CL:AUG:SG:INDEF NEG obtain how
 ‘Each [soldier] had no choice, but ...’

This tour of elicited examples and sentences in context shows that speaker registers are attached in canonical ways to each of the 12-classifier forms. Pragmatic implicatures are triggered whenever the alignment is non-canonical.

3 The historical development of the Ahmao inflectional classifiers

3.1 The influence of societal factors

In Gerner and Bisang (2009), we provide a detailed review of the history of inter-ethnic relations in Western Guizhou. The Ahmao, who several centuries ago fled from the progressing Han populations toward the country owned by Yi landlords in Western Guizhou, became serfs of the Yi (the Yi are another ancient people and constitute one of the 56 officially recognized nationalities in the People's Republic of China). This double pressure from the Han and Yi populations reportedly affected the ethnopsyché of the Ahmao in a more adversative way than that of other Miao groupings² (see Gerner and Bisang 2009 for bibliographic references).

Wang Deguang,³ who served as an informant for this paper, prompted an interesting ethno-social theory. He proposed that gender relations in Ahmao society were upset by the inter-ethnic oppression the Ahmao were exposed to. Women would have had to endure the greater burden of this imbalance, as they were the targets of male aggression. Wang suggested that the female register for classifiers developed as an articulation of self-protection against male forcefulness.⁴

Wang's suggestion has a certain appeal, although no independent evidence in the form of ethnological data is available. Even if contemporary ethnological information was accessible, the data would probably be of little help in measuring the exact impact society had on the formation of the Ahmao classifier paradigm, because the social order in which the Ahmao sound changes are embedded was eradicated in 1949 when the People's Republic was founded. (The actual changes first occurred at least 100–150 years ago.) In our reconstruction of these sound changes, we therefore have to rely, above all, on the principles of historical linguistics, which allow us to redraw rather securely the grammaticalization path of the set of Ahmao classifiers.

3.2 How the Ahmao classifiers are grammaticalized

About no other isolating language in East or Southeast Asia, let alone any other Miao language, is it reported that classifiers can be declined in a paradigm of forms. There are various indicators showing that the Ahmao classifier system was generated in two waves of innovation, first through interaction with the two now frozen and unproductive noun prefixes *a*⁵⁵ *jie*⁵³

Table 7. The Ahmao prefixes $a^{55}pie^{53}$ (augmentative) and ηa^{11} (diminutive)

Noun	Augmentative prefix $a^{55}pie^{53}$	Diminutive prefix ηa^{11}
ηfiu^{35} ‘ox, bull’	$a^{55}pie^{53}\eta fiu^{35}$ ‘cow’	$\eta a^{11}\eta fiu^{35}$ ‘calf’
mpa^{44} ‘pig, hog’	$a^{55}pie^{53}mpa^{44}$ ‘sow’	$\eta a^{11}mpa^{44}$ ‘piglet’
nau^{31} ‘bird’	$a^{55}pie^{53}nau^{31}$ ‘female bird’	$\eta a^{11}nau^{31}$ ‘bird poult’
$d\eta i^{11}$ ‘hand’	$a^{55}pie^{53}d\eta i^{11}$ ‘thumb’	$\eta a^{11}d\eta i^{11}$ ‘pinkie, little finger’
tca^{44} ‘wind’	$a^{55}pie^{53}tca^{44}$ ‘storm’	$\eta a^{11}tca^{44}$ ‘breeze of wind’
$dl\eta i^{35}$ ‘river’	$a^{55}pie^{53}dl\eta i^{35}$ ‘big river’	$\eta a^{11}dl\eta i^{35}$ ‘brook’

‘female’ and ηa^{11} ‘small’ (cf. Section 3.2.1), and second through an array of suprasegmental processes, such as devoicing or breathy voicing, whereby indefinite classifier forms were derived from definite forms. It is probable that these suprasegmental processes occurred at a later stage in time (cf. Section 3.2.2).

3.2.1 The split into augmentative, medial and diminutive forms

In synchronic Ahmao, there are two frozen size prefixes, the augmentative $a^{55}pie^{53}$ and the diminutive ηa^{11} . As in many languages in East and Southeast Asia (cf. Matisoff, 1991), these size prefixes are derived from the terms for ‘mother’ and ‘child’: the augmentative is related to pie^{53} ‘mother’ and the diminutive is retrieved from $\eta a^{55}zau^{11}$ ‘child’. Both strings may collocate with a number of mainly domestic animal nouns to indicate gender or age (although human nouns may not be modified by these prefixes). In addition to domestic animals, a few inanimate nouns may be attached to these prefixes, thereby contributing the meanings of ‘big’ ($a^{55}pie^{53}$) and ‘small’ (ηa^{11}). The chart in Table 7 is a short excerpt of a longer list provided in Gerner and Bisang (2009).

The Ahmao prefixes have cognate forms in other related Miao languages (see the Appendix in Section 3.2.2 for more genetic information). In Hekou Hmong, for example, two similar prefixes, na^{21} ‘mother’ and mi^{35} ‘child’ exist. The Hekou Hmong prefix na^{21} is obviously cognate with the second syllable of $a^{55}pie^{53}$ (Ahmao), whereas the prefix mi^{35} (Hmong) reveals affinities with ηa^{11} (Ahmao). The scope of these two prefixes is similar: both may be attached to domestic animal nouns and certain inanimate nouns (see Table 8 on the following page).

Table 8. Cognate prefixes in Hekou Hmong

Noun	Augmentative prefix <i>na</i> ²¹	Diminutive prefix <i>mi</i> ³⁵
<i>gai</i> ⁵⁴ ‘cock’	<i>na</i> ²¹ <i>gai</i> ⁵⁴ ‘hen’	<i>mi</i> ³⁵ <i>gai</i> ⁵⁴ ‘chicken’
<i>mbua</i> ⁴⁴ ‘pig, boar’	<i>na</i> ²¹ <i>mbua</i> ⁴⁴ ‘sow’	<i>mi</i> ³⁵ <i>mbua</i> ⁴⁴ ‘piglet’
<i>loj</i> ³³ ‘city’	<i>na</i> ²¹ <i>loj</i> ³³ ‘big city’	<i>mi</i> ³⁵ <i>loj</i> ³³ ‘small city’

The classifiers in Hekou Hmong are unique forms that have not merged with other morphemes as in Weining Ahmao. In Ahmao, a complex process of sound and meaning changes has amalgamated the two prefixes *a*⁵⁵ *pie*⁵³ ‘female’ and *ŋa*¹¹ ‘small’ with various classifiers through the following procedure (cf. Gerner and Bisang 2009 for more details).

Phase 1. Metanalysis and phonological changes

The grammaticalization process started with a reinterpretation of the morphemes’ boundaries (metanalysis). The size prefixes *a*⁵⁵ *pie*⁵³ and *ŋa*¹¹ were no longer viewed as forming a unit with the noun that followed, but were rebracketed together with the preceding classifier. The rearranged units then underwent rapid sound changes, i. e. the classifiers and size prefixes merged through multi-layered processes of syncope, apocope and aphaeresis.

Phase 2. Semantic and pragmatic reanalysis

When the various forms came into existence, their meaning was reinterpreted. The [ai]-forms were reanalyzed as female and augmentative classifiers, and the [a]-forms as child and diminutive classifiers. For the very few classifiers that can modify both animal and inanimate nouns, the female and augmentative senses on the one hand and the child and diminutive senses on the other hand started to merge. Before the newly acquired senses expanded to other classifiers, another round of pragmatic reanalysis of these merged morphemes occurred. As the gender and child senses are blocked for inanimate nouns, the classifier forms shifted from expressing the gender / child roles of the noun referents to encoding the gender / child roles of the speaker. The female domain probably first came into existence as a linguistic refuge in the face of male aggression (see Section 3.1). The child domain arose primarily through a process of imitation, as children associate more easily with their mother at an early age.

Phase 3. Analogical spread

These phonological, semantic and pragmatic changes then spread to all other classifiers, even to those that may not collocate with nouns prefixed by augmentative and diminutive morphemes.

Phase 4. Semantic and pragmatic reanalysis

Initially, the original unchanged classifier form conveyed a neutral meaning, but when the [ai]- and [a]-forms, as well as their speaker roles appeared, male speakers naturally took up the unmerged forms as their default. Through this process, the original forms became associated with male aggressiveness. As a consequence, the size values of the various forms were redistributed. The augmentative [ai]-form was downgraded and given the value of *medial*. Eventually, a three-value system emerged: *augmentative* (male speaker domain), *medial* (female speaker domain) and *diminutive* (child speaker domain).

3.2.2 The split into definite and indefinite forms

The second major split in the classifier system was the separation of definite and indefinite forms. This split probably occurred at a later point in time. For all three size values, there are definite and indefinite classifier versions, whereas in other Miao languages bare classifier-noun compounds are always interpreted as definite articles. Preposing the numeral for ‘one’ generates an indefinite article reading in unfocused contexts (and in focused contexts implying the value ‘one’). The following examples are from three additional Miao languages. Comparative data from Weining Ahmao are displayed first.

(13) Weining Ahmao

- | | |
|--|---|
| a. tu⁴⁴ pfu³⁵
CL:AUG:SG:DEF ox
‘the ox’ | b. i⁵⁵ du³¹ pfu³⁵
NUM:1 CL:AUG:SG:INDEF ox
‘one ox’ |
| c. du³¹ pfu³⁵
CL:AUG:SG:INDEF ox
‘an ox’ | |

(14) Hekou Hmong

- | | |
|---|--|
| a. to²¹ po³¹
CL ox
‘the ox’ | b. i⁵⁵ to²¹ po³¹
NUM:1 CL ox
‘an / one ox’ |
|---|--|

(15) Kaili Qanao

a. **te**¹¹ **lio**³⁵
 CL ox
 ‘the ox’

b. **i**⁵⁵ **te**¹¹ **lio**³⁵
 NUM:1 CL ox
 ‘an / one ox’

(16) Huayuan Qoxong

a. **ɲoŋ**¹¹ **zu**¹¹
 CL:AUG:SG:DEF ox
 ‘the ox’

b. **i**⁵⁵ **ɲoŋ**¹¹ **zu**¹¹
 NUM:1 CL:AUG:SG:INDEF ox
 ‘an / one ox’

Because classifier-noun compounds generally convey the meaning of definite articles, it is plausible that in Weining Ahmao the underlying classifier forms had been definite versions out of which in Miao languages grew the indefinite versions by means of such suprasegmental processes as voicing, aspiration or tone changes. These processes were triggered by the numeral **i**⁵⁵ through a process called *the change in glottal articulation* in the literature (Lehmann 1992: 193). The glottal stop [ʔ] intervened between the numeral and the classifier in the following way. The Proto-Miao-Yao numeral ‘one’ seems to have displayed a final alveolar stop and can be reconstructed as **iet*. Similar to most Miao languages, this numeral was engaged in a long-term process of lenition whereby its final alveolar stop weakened into a glottal stop. It was with the glottal stop that the numeral **iʔ*⁵⁵ ‘one’ started its morphophonological interaction with the set of Ahmao classifiers. It then underwent morphological reanalysis, resulting in the rebracketing of the glottal stop [ʔ] with the classifiers. Scholars have acknowledged cases in which glottal articulation is a trigger of such suprasegmental phenomena as voicing, aspiration, etc. (e. g. Lehmann 1992: 193; Matisoff 1973: 32–34; Gerner 2007). In support of this claim is the fact that in Ahmao the numeral **i**⁵⁵ ‘one’ may not collocate with the definite classifier, which suggests that **i**⁵⁵ together with the indefinite classifier versions has been decoupled from the definite classifier versions. More details are given in Gerner and Bisang (2009).

Abbreviations

1P.SG = first person singular; 3P.SG = third person singular; AUG = augmentative; C = consonant; CL = classifier; COP = copular; COV = coverb; COV:take = coverb with verbal origin; DEF = definite; DEM = demonstrative; DEM:S&A:FAM = demonstrative: familiar to speaker and addressee; DEM:MED-HIGH = demonstrative: medial distance to, higher altitude than speaker; DEM:PROX = demonstrative:

proximal distance to speaker; DIM = diminutive; DP = dynamic perfect particle; EXCL = exclamation; INDEF = indefinite; INT = interrogative; INT:which = interrogative with gloss; MED = medial; N = noun; NEG = negation; NOM = nominalization; NUM = numeral; NUM:9 = numeral with its value; PL = plural; SG = singular; T = tone; TB = Tibeto-Burman; V = vowel; VP = verb phrase

Acknowledgments

The research undertaken for this paper was partly supported by two Strategic Research Grants from City University of Hong Kong (Project Nos. 7001921 and 7002188). We wish to express our special gratitude to Wang Deguang, a native Ahmao speaker who served as our principal informant (during 2005–2007). Warm regards also go to Noel Johnston for his comments on an earlier version of the manuscript.

Appendix: Information on Miao languages

Internationally, the better known appellation for the Miao group is Hmong or Hmongic. Western countries' acquaintance with this Southeast Asian people is primarily due to the Indo-Chinese wars in the aftermath of which more than 100,000 Hmong fled to the United States, France and Australia. The homeland of this group is Southwest China, from which migration started in the 18th century to neighboring Thailand, Laos, Vietnam and Burma. In the 2000 census, more than 8 million Miao were reckoned to be scattered over Southwest China. In the Chinese linguistic tradition, three 'dialects' of 'the Miao language' have been distinguished: Western, Central and Eastern (Wang

Table 9. Background information on the Miao languages cited

Language	Branch	Location	Number of speakers
Hekou Hmong	Western	China: Yunnan / Honghe (Hekou ...)	ca. 500,000
Kaili Hmu-Qanao	Central	China: Guizhou / Kaili (Kaili ...)	total ca. 1,800,000
Huayuan Qoxong	Eastern	China: Hunan / Xiangxi (Huayuan ...)	ca. 50,000
Weining Ahmao	Between Central and Western	China: Guizhou / Bijie (Weining ...)	ca. 350,000

Fushi 1985). These dialects would correspond to the concept of language in the Western tradition if the vague notion of intelligibility was taken as criterion. In fact, taking intelligibility alone as the criterion we would estimate that there are perhaps 100 Miao languages in Southwest China. In this paper, comparative classifier data from four representative languages of the Miao group have been presented.

Notes

1. The numbers ⁵⁵, ¹³, etc. are tone markers and indicate relative pitch on a scale from ¹ (lowest) to ⁵ (highest). The first number represents the beginning and the second number the end of the tonal contour. The transcription of sounds in this paper follows the International Phonetic Alphabet without shortcuts. For the interlinear abbreviations used in the examples, refer to the section on abbreviations.
2. The Miao living in other areas of Southwest China fought back as is well documented in Chinese and Western sources (cf. Jenks 1994).
3. He is the author of several papers (Wang 1986, 1987) and represents the intellectual flagship of the small Ahmao diaspora in Beijing.
4. In Gerner and Bisang (2009), we describe it as a kind of self-politeness strategy against face-threatening acts. The notion of self-politeness was coined by Chen (2001).

References

- Adams, Karen L. and Nancy F. Conklin
 1973 Towards a theory of natural classification. *Papers from the Annual Regional Meeting of the Chicago Linguistic Society* 9: 1–10.
- Aikhenvald, Alexandra Y.
 2000 *Classifiers. A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Keith
 1977 Classifiers. *Language* 53: 285–311.
- Bisang, Walter
 1993 Classifiers, quantifiers and class nouns in Hmong. *Studies in Language* 17: 1–51.
- Bisang, Walter
 1999 Classifiers in East and Southeast Asian languages: Counting and beyond. In *Numeral types and changes worldwide*, J. Gvozdanovic (ed.), 113–185. Berlin: Mouton de Gruyter.
- Bisang, Walter
 2002 Classification and the evolution of grammatical structures: A universal perspective. In *Language evolution in a typological perspective*, T. Leuschner (ed.), Thematic volume of *Sprachtypologie und Universalienforschung* 55: 289–308.

- Chen, Rong
2001 Self-politeness: A proposal. *Journal of Pragmatics* 33: 87–106.
- Conklin, Nancy Faires
1981 *The semantics and syntax of numeral classification in Tai and Austronesian*. Doctoral dissertation. Chicago: University of Michigan.
- Corbett, Greville
1991 *Gender*. Cambridge: Cambridge University Press.
- Croft, William
1994 Semantic universals in classifier systems. *Word* 45: 145–171.
- Daley, Karen Ann
1998 *Vietnamese classifiers in narrative texts*. (Publication 125). Arlington: Summer Institute of Linguistics and University of Texas.
- de la Grassière, R.
1998 La catégorie psychologique de la classification révélée par le langage. *Revue philosophique de la France et de l'étranger* 45: 594–624.
- Denny, J. Peter
1976 What are noun classifiers good for? *Papers from the Annual Regional Meeting of the Chicago Linguistic Society* 12: 122–132.
- Dixon, Robert M. W.
1982 *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton.
- Gerner, Matthias
2003 Demonstratives, articles and topic markers in the Yi group. *Journal of Pragmatics* 35: 947–998.
- Gerner, Matthias
2006 Noun classifiers in Kam and Chinese Kam-Tai languages: Their morphosyntax, semantics and history. *Journal of Chinese Linguistics* 34: 237–305.
- Gerner, Matthias
2007 The lexicalization of causative verbs in the Yi group. *Folia Linguistica Historica* 28: 145–185.
- Gerner, Matthias and Walter Bisang
2008 Inflectional speaker-role classifiers in Weining Ahmao. *Journal of Pragmatics* 40: 719–732.
- Gerner, Matthias and Walter Bisang
2009 Inflectional classifiers in Weining Ahmao: Mirror of the history of a people. *Folia Linguistica Historica* 30: 183–218.
- Gerner, Matthias and Walter Bisang
2010 Classifier declinations in an isolating language: On a rarity in Weining Ahmao. *Language and Linguistics* 11(3): 576–623. Taipei: Academia Sinica.
- Greenberg, Joseph
1974 Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. In *Proceedings of the 11th International Congress of Linguists, Bologna/Florence, Aug.–Sept. 1972*, 17–37.
- Jenks, Robert
1994 *Insurgency and social disorder in Guizhou (The "Miao" Rebellion 1854–1873)*. Honolulu: University of Hawaii Press.

- Kasper, Gabriele
 1990 Linguistic politeness: Current research issues. *Journal of Pragmatics* 14: 193–218.
- Kochman, Thomas
 1984 The politics of politeness: Social warrants in mainstream American public etiquette. In *Meaning, form, and use in context: Linguistic applications*, (Georgetown University Round Table on Languages and Linguistics 1984), Deborah Schiffrin (ed.), 200–209. Washington, D. C.: Georgetown University Press.
- Lehmann, Winfred
 1992 *Historical Linguistics: An introduction*. London: Routledge.
- Levinson, Stephen C.
 1983 *Pragmatics*. Cambridge: Cambridge University Press.
- Li, Charles and Sandra Thompson
 1981 *Mandarin Chinese*. Berkeley: University of California Press.
- Lucy, John
 2000 Systems of nominal classification: a concluding discussion. In *Systems of nominal classification*, Gunter Senft (ed.), 326–341. Cambridge: Cambridge University Press.
- Matisoff, James
 1973 *The grammar of Lahu*. (= Publications in Linguistics; 75). Berkeley: University of California.
- Matisoff, James
 1991 The mother of all morphemes: Augmentatives and diminutives in areal and universal perspective. In *Papers from the first annual meeting of the Southeast Asian Linguistics society*, M. Ratliff and E. Schiller (eds.), 293–349. Tempe: Arizona State University.
- Shils, Edward
 1968 Deference. In *Social stratification*, J. A. Jackson (ed.), 104–132. Cambridge: Cambridge University Press.
- Wang, Deguang
 1986 Weining Miaoyu huayu cailiao [Language material in the Weining dialect of the Miao language]. *Minzu Yuwen* 3: 69–80.
- Wang, Deguang
 1987 Guizhou Weining Miaoyu liangci shiyi [Additional remarks on the classifiers of the Miao language in Weining, Guizhou province]. *Minzu Yuwen* 5: 36–39.
- Wang, Fushi
 1985 *Miaoyu Jianzhi* [An outline grammar of Miao]. Beijing: Minzu Chubanshe [Nationalities Press].

“She kisses her late husband” = “She kissed her husband”: nominal tense in Movima

Katharina Haude

1 Introduction

Tense marking, a grammatical means to “locate situations in time” (Comrie 1994: 4558), is typically an inflectional category of verbs. In general, definitions of tense explicitly or implicitly claim that tense marking will be found in the verbal morphology of a language. In languages that do not have verbal tense markers, temporal reference is usually made through either aspectual or modal morphology (Smith 2008; Comrie 1985: 50–51), which also belongs to the verbal domain. The fact that temporal reference is typically carried out in the verbal domain is of course no coincidence. Verbs conventionally denote situations (events, processes, states) that are not time-stable, in contrast to nouns, which predominantly denote things, i. e. time-stable concepts.

In this paper I discuss the case of Movima, a linguistic isolate of the Bolivian Amazon area with still several hundred speakers. Movima has no verbal tense morphology. Unlike other tenseless languages, Movima does not recur to mood or aspect morphology, but instead uses a rich system of referential elements (articles, pronouns, demonstratives) to express temporal relations. This strategy can be described as “nominal tense marking” as defined by Nordlinger and Sadler (2004). However, Movima proves to be different from the languages that were investigated by Nordlinger and Sadler in their cross-linguistic study, because in Movima, the same markers that indicate temporal properties of a nominal referent, can also determine the temporal interpretation of a clause as a whole.¹ This conflation of “independent” and “propositional” nominal tense marking (Nordlinger and Sadler 2004) has so far not been identified in any other language and can be considered a cross-linguistic rarity.

The paper is structured as follows. In Section 2, Nordlinger and Sadler’s concept of nominal tense marking is introduced. In Section 3 it is shown that Movima has no verbal tense morphology (3.1), and that tense categories are optionally expressed by particles (3.2). Section 4 presents the system of ar-

ticles, which indicate spatial and temporal properties of nominal referents. Section 5 shows how the temporal-deictic function of the articles by implicature determines the temporal interpretation of the clause (5.1), and that the article can also serve as a marker of temporal reference in discourse (5.2). Section 6 concludes by presenting the possible pathway from independent to propositional tense marking.

The data on which the analysis is based were collected in Santa Ana del Yacuma, Bolivia, between 2001 and 2004. Most of the data presented here stem from spontaneous discourse; elicited examples are explicitly marked.

2 Nominal tense marking

In their study on nominal tense marking, Nordlinger and Sadler (2004: 778) give the following definition of nominal tense, aspect and mood (TAM) marking:

- i. Nouns (or other noun phrase/determiner phrase constituents) show a distinction in one or more of the categories of tense, aspect, and mood, where these categories are standardly defined as they would be for verbs (e. g. Crystal 1997).²
- ii. This tense-aspect-mood distinction is productive across the whole word class and not restricted to a small subset of forms (such as *ex-* in English).
- iii. The tense-aspect-mood distinction is not restricted to nominals functioning as predicates of verbless clauses but is encoded on arguments and/or adjunct noun phrases/determiner phrases in clauses headed by verbs.
- iv. The tense-aspect-mood marker is a morphological category of the nominal word class and cannot be treated as a syntactic clitic that merely attaches phonologically to the noun phrases/determiner phrases.

I will return to these criteria later when discussing the Movima data in Section 5.1.

A major observation of Nordlinger and Sadler is that the languages that display nominal tense marking do this in one of two ways: either the marker indicates temporal properties of the entity denoted by the nominal, or it determines the temporal interpretation of the entire proposition, similar to tense marking on verbs. In the first case, the marking of temporality on the nominal is independent from clausal tense, which is why Nordlinger and Sadler label

it “independent nominal tense marking”. In the second case, called “propositional nominal tense marking”, nominal marking either co-occurs with verbal tense marking or is the only marker of tense in the clause. Their definitions (extended to tense-aspect-mood in general) are as follows:

Independent nominal tense-aspect-mood “operates completely independently of the TAM of the clause and serves to locate the time at which the property denoted by the nominal holds of the referent or, in the case of possessive phrases, the time at which the possessive relation holds” (Nordlinger and Sadler 2004: 779).

Propositional nominal tense-aspect-mood “contributes TAM information relevant to the clause as a whole” (Nordlinger and Sadler 2004: 790).

Of Nordlinger and Sadler’s sample of approximately 22 languages, 12 languages display independent nominal tense and 10 show propositional tense. None of the investigated languages displays both phenomena, the message being that nominal tense marking has either one or the other function, but not both (see also Nordlinger and Sadler 2004: 778).

In the following sections I will show that Movima does not have verbal morphology for tense. Tense is expressed through particles, which are not grammatically obligatory. Temporal reference comparable to verbal tense marking in other languages is conveyed through the form of the article, an obligatory part of a referential phrase. I will show that this type of marking can be described in terms of Nordlinger and Sadler’s criteria (i)–(iv) above. However, in comparison to the languages studied there, Movima is unusual in that the same mechanism marks “independent” as well as “propositional” tense.

3 Time reference in Movima

3.1 Basic features of Movima clause structure

For a better understanding of the examples below I will provide here a brief account of Movima clause structure (see Haude 2006 for a detailed description). Movima clauses are generally predicate-initial with a preverbal slot for a topicalized pronoun:

(TOP) PRED=ARG (ARG)

An argument can be represented through a bound or free pronoun or through a referential phrase, which consists of an article and a content word. At any point, even inside a referential phrase, tense-aspect-mood and discourse particles can occur (see below). Example (1) illustrates a basic transitive clause, which consists of a predicate and two referential phrases representing the arguments.

- (1) *man<a>ye=is pa:ko os rulrul*
 encounter<DR>=ART.PL dog ART.N.PST jaguar
 ‘The dogs met the jaguar.’³

Non-core arguments are marked by an oblique case marker prefixed to the referential element (article or pronoun):⁴

- (2) *ilo:ni--y’ti n-os chapmo*
 walk--1PL OBL-ART.N.PST forest
 ‘We walked in the forest.’

The content word in a referential phrase does not necessarily have to be a morphological noun, as in the above examples. It can also be a verb, in which case the phrase refers to an event participant (see Haude 2009), as in (3).⁵ Still, for the sake of simplicity I will refer to referential phrases as noun phrases (NPs) from here on.

- (3) *is yey-na=n*
 ART.PL want-DR=2
 ‘the ones you want’

I will now turn to temporal reference in Movima. In 3.2 it is shown that there is no verbal tense marking morphology, 3.3 illustrates the function of tense particles, and in the subsequent sections I describe the deictic properties of the article.

3.2 The lack of verbal tense morphology

Movima has no verbal tense marking morphology. Depending on the context, the clause in (4) can have a past, present, or future interpretation.

- (4) *yolmoł--isne*
 stroll--3F.AB
 ‘She (absent) went / is going / will go for a walk.’

The only verbal affix that can have a temporal reading is the irrealis infix $\langle(k)a'\rangle$.⁶ In affirmative contexts, it indicates that the speaker believes that the situation will take place some time in the future:

- (5) *sal* $\langle a'\rangle$ *mo*
 return<IRR>
 ‘I’ll be back!’

Aspect affixes in Movima cover meanings such as “repeatedly” or “at once”. None of them locates the denoted situation temporally. On the whole, the major tense-aspect-mood categories are expressed through particles, most of which are not formally related to the verb. I will give an outline of the relevant particles in the following section.

3.3 Tense particles

Movima has a large array of particles and adverbs that express temporal, modal, and aspectual categories. Temporal relations are mainly established by the following ones:

Table 1. Movima tense particles

<i>kwil</i>	‘remote past’	(a long time ago)
<i>la’</i>	‘anterior’	(previous occurrence, before the day of speaking)
<i>kwey</i>	‘immediate past’	(same day, but before the time of speaking)
<i>loy</i>	‘intention’	(future, normally directly after the time of speaking)
<i>nokowa</i>	‘future’	(any future time)

As will become apparent below (Section 4.2), the major temporal categories in Movima are “past” (before the day of speaking), “immediate past” (on the same day, but before the time of speaking),⁷ and “nonpast”. Accordingly, the above particles can be grouped into two forms pertaining to past (*kwil* and *la’*), one that pertains to immediate past (*kwey*), and two that pertain to nonpast (*loy* and *nokowa*).

The following examples illustrate the way in which these particles convey temporal relations. Examples (6) and (7) illustrate the particles *kwil* (REM) and *la’* (ANT). They occur frequently inside a noun phrase (i. e., between the article and the noun), which indicates that they are not syntactically part of the verb phrase.

- (6) *n-os kwil baytim-a=y'ti rey mo: maj<a'><ni:~>ni=Ø*
 OBL-ART.N.PST REM field-LV=1PL MOD yet child<IRR><NMZ.N~>=1SG
 'In our field, long ago, I didn't have children yet, you see.'
- (7) *jo'yaj--us n-os la' walaylo*
 arrive--3M.AB OBL-ART.N.PST ANT afternoon
 'He arrived yesterday afternoon.' (lit.: "on the previous afternoon.")

The particle *kwey* (IMM) usually occurs before the predicate, as in (8), but it can also occur inside a noun phrase, as shown in (9):

- (8) *kwey joy-chet us pa:pa=kinos majni=Ø n-as ra:diyo*
 IMM go-R/R ART.M father_of=ART.F.AB child=1SG OBL-ART.N radio
 'The father of my daughter went to the radio (station) (earlier today).'
- (9) *no-kos kwey lu:vuy jayna kiro' joy-chet*
 OBL-ART.N.AB IMM twilight DSC DEM.PL.AB go-R/R
 '(Today) at dawn they left already.'

The occurrence of the particle *nokowa* (FUT) before the predicate and inside a referential phrase is illustrated by (10) and (11), respectively:

- (10) *n-as tojet-wa=as piyesta nokowa jiwa-te-na=y'ti rey*
 OBL-ART.N pass-NMZ=ART.N fiesta FUT come-APPL-DR=1PL MOD
 'When the *fiesta* is over we will bring (her) again.'
- (11) *dira tat koro' kos nokowa piyesta*
 still EV DEM.N.AB ART.N.AB FUT fiesta
 'There will still be a *fiesta*, they say.'

The tense-aspect-mood particle *loy* (ITN) is the only one that never occurs inside a referential phrase, but normally right before the predicate, as in example (12).⁸

- (12) *jayna loy rey t joy-chet ma'*
 DSC ITN MOD 1INTR go-R/R my_mother
 'I'm leaving again already, mother.'

None of these particles is grammatically obligatory. For example, the particle *la'*, which indicates the previous occurrence of an event, is not needed to indicate past tense in general. The phrase *nos walaylo* 'in the afternoon' in (13) also refers to an afternoon in the past (as indicated by the article *os*, see Section 4), but it is not specified as the previous afternoon, as in (7) above.

- (13) *lassinko os joyaj-wa=i, n-os walaylo*
 five_o'clock ART.N.PST arrive-NMZ=3PL OBL-ART.N.PST afternoon
 'At five o'clock (was) their arrival, in the afternoon.'

Likewise, the particle *kwey* is not obligatory in contexts that refer to the immediate past, as is the case in (14) (see Section 4.2 below for the "immediate past" interpretation of the Absential article *kos*):

- (14) *jayna kino' chi:~chi no-kos ima:yoj*
 DSC DEM.F.AB MD~go_out OBL-ART.N.AB morning
 'She has gone out (today) in the morning.'

The particles *loy* and *nokowa*, finally, are not obligatory for indicating non-past tense. The pair in (15) shows this for *loy*:

- (15) a. *n-as sot-tino:nak loy it sal-na=∅*
 OBL-ART.N other-year ITN 1 search-DR=1SG
 'Next year I will look for (you).'
 b. *n-as sot-tino:nak it sal-na=∅*
 OBL-ART.N other-year 1 search-DR=1SG
 'Next year I will look for (you).'

Example (16) is comparable in structure with (11) on the preceding page. Here, however, reference to nonpast (future) tense is made without the particle *nokowa*.⁹

- (16) *n-as vakasyon [...] it joy-chet n-as vayeł-wa=∅*
 OBL-ART.N holiday [...] 1INTR go-R/R OBL-ART.N see-NMZ=1SG
n-iy'bikweł
 OBL-PRO.2PL
 'In the holidays [...] I'll come to see you.'

In texts, the tense particles generally occur only to establish the time of the narrated events, typically at the beginning of a text or passage (see, for instance, (7) on the facing page). Once the discourse time has been established, the particle does usually not occur again. For example, in the discourse preceding the sentence in (14) above, the discourse time had been established as the previous day.

The exact status of the tense particles requires further research. Still, I hope to have shown that they serve to indicate temporal relations, but that they are not grammatical markers of tense.

4 Spatio-temporal reference

Despite the fact that there is no verbal tense morphology in Movima and that the tense particles are not grammatically obligatory, sentences like those in (13)–(16) above do contain information as to their temporal interpretation. In (13), it is asserted that the described situation has occurred in the past, i. e. before the day of speaking. Example (14) is unambiguously interpreted as describing a situation that has occurred on the same day, but previous to the moment of speaking. Example (15) and (16) clearly refer to a nonpast time. These meanings are brought about by the form of the article of one or more referential phrases in the clause, as will be shown in this and the following sections.

Movima has a rich inventory of referential elements: personal pronouns, articles, and demonstratives. They all have highly specific deictic properties, always indicating number, gender, and physical presence at or absence from the speech situation. For absent referents, articles (but not pronouns, and demonstratives only in restricted contexts) additionally indicate whether the referent is still in existence (Absential) or whether it has ceased to exist (Past). This property makes the article the main carrier of temporal information in Movima. Its forms are given in Table 2.

Table 2. The Movima article

	Masculine	Feminine	Neuter	Plural
Presential	<i>us</i>	<i>i'nes</i>	<i>as</i>	<i>is</i>
Absential	<i>kus</i>	<i>kinos</i>	<i>kos</i>	<i>kis</i>
Past	<i>us (usos)</i>	<i>isnos</i>	<i>os</i>	<i>is (isos)</i>

The article is defined by the fact that it is obligatorily followed by a content word, together with which it forms a noun phrase (see 3.1). Note that it does not indicate definiteness. For definite reference to an entity present at the speech situation, a demonstrative can be used.¹⁰ For definite reference to an entity absent from the speech situation, the deictic adverb *ney* ‘here’ is placed before the content word:

- (17) *ma'nes kis ney nonlo:-ba*
 tasty ART.PL.AB here milk-BR.round
 ‘Those milk cakes are tasty.’

The temporal-deictic function of the article will be elaborated on in the following sections.

4.1 Concrete referents: presence, absence, ceased existence

Let us first consider the function of the article in noun phrases that refer to concrete entities, illustrated in (18):

- (18) a. **as** *pa:ko*
 ART.N dog
 ‘the dog’
 b. **kos** *pa:ko*
 ART.N.AB dog
 ‘the / a dog’
 c. **os** *pa:ko*
 ART.N.PST dog
 ‘the / a dog’

The noun phrase in (18a), with the Presential neuter article *as*, is used for reference to a dog that is present near the speech situation, for example in sight or somewhere in the same house or compound. It can also have a generic meaning, referring to the species of dogs.¹¹ The noun phrase in (18b), with the Absential article *kos*, refers to a dog that exists somewhere, but is not near the speech situation. The noun phrase in (18c), which contains the Past article, is used to refer to a dog that has ceased to exist. The article thus distinguishes two temporal categories: ongoing and ceased existence of the referent.

It can be seen in Table 2 on the facing page that only the neuter and the feminine article distinguishes formally between presence, absence, and ceased existence, by having a Presential (*as*, *i'nes*), an Absential (*kos*, *kinos*), and also a distinct Past (*os*, *isnos*) form. By contrast, in the masculine and the plural paradigm, the Presential and Past categories are expressed by the same form (*us* and *is*, respectively); the Past forms given in brackets (*usos* and *isos*), are hardly ever used and, according to speakers, imply that the referent has ceased to exist a long time ago. The forms *us* (masculine Presential and Past) and *is* (plural Presential and Past) are best seen as each representing two homophonous morphemes with different meanings: when the referent is absent, *us* and *is* are automatically interpreted as Past articles, implying ceased existence of the referent, since otherwise, the Absential forms (*kus*, *kis*) would be used.

In order to apply the Past article, the referent must have ceased to exist completely. It is not possible, as in other languages that display similar phenomena (cf. Nordlinger and Sadler 2004), to use this form to refer to dead bodies or to objects that have simply lost their function, or that a possessive relation has ended while the possessed entity still exists. Hence, it is not equivalent to nominal temporality markers such as the English prefix *ex-*, which indicates a loss of function or termination of a possessive relation, but not that the referent itself has ceased to exist. In Movima, for the Past article to be applied, the referent must be physically “gone”; otherwise, the Presential or Absential article is used.

Example (19) illustrates this for a glass that is broken, but whose remainders are still around, and which is accordingly referred to by the Presential article (note, however, that the bound pronoun *=us* refers to an absent agent); Example (20) shows it for a person that has died, but whose body is still in existence (at the time of the actual utterance, it is being taken to the cemetery); since the body is absent from the speech situation, the Absential form is used.

- (19) *bay<a>cho=us as wa:so*
 break<DR>=3M.AB ART.N glass
 ‘He (absent) has broken the (present) glass.’ (elicited)

- (20) *kinos senyo:ra jala:yij n-os kayni-wa=sne*
 ART.F.AB madam angry OBL-ART.N.PST die-NMZ=3F.AB
 ‘The woman was angry when she died.’ (lit.: “... in her dying.”)

A further condition for the applicability of the Past article is that the referent must have ceased to exist before the day of speaking. When it has ceased to exist on the same day, then inevitably the Absential article is used, and the Past article would be ungrammatical:

- (21) a. *kwey it kay~kay no-kos jokme*
 IMM 1INTR MD~eat OBL-ART.N.AB bird
 ‘Today I ate a chicken.’
 b. **kwey it kay~kay n-os jokme*
 IMM 1INTR MD~eat OBL-ART.N.PST bird
 ‘Today I ate a chicken.’ (elicited)

While the Past article asserts that the referent has ceased to exist before the day of speaking, there is no similar marker for future existence, as in some other languages with nominal tense marking (e. g. Guaraní). For entities that

will come into existence in the future, either the Presential or the Absential article is used:

- (22) *loy it̪ ji:sa:-na=∅ kis lo:kwa*
 ITN 1 make-DR=1 SG ART.PL.AB stew
 ‘I’ll make stew.’

- (23) *loy it̪ ji:sa:-na=∅ is narasa:-mi*
 ITN 1 make-DR=1 SG ART.PL orange-BR.water
 ‘I’ll make orange juice.’

The choice of either the Presential or the Absential article for entities that are not yet in existence, as in (22) and (23), is not entirely obvious. It may have to do with modal implications of Absential marking. It will be shown in the following section, however, that noun phrases denoting future times and situations always contain the Presential article.

4.2 Times and situations: a three-way temporal distinction

The temporal function of the article is still more differentiated in noun phrases that refer to times (i. e., points in time or time spans) and situations (events, processes, states etc.). This type of noun phrase always contains the neuter article. Points in time are denoted by words such as *je:mes* ‘day’, *tino:nak* ‘year’ etc. Noun phrases referring to situations contain derived nouns. The suffix *wa* derives nouns from verbs; reduplication and the suffixation of *ni-wa* derive nouns from nonverbal predicates. Situation-denoting noun phrases are extremely common in Movima because they are used for subordination.

With time- and situation-denoting noun phrases, all three forms of the article (Presential, Absential, Past) receive a temporal interpretation. The Presential article is used in present (24), future (25), and habitual (26) contexts:

- (24) *jayna tojeɬ po:la as salmo-wa=nkweɬ*
 DSC very late ART.N return-NMZ=2PL
 ‘You (pl) are coming back very late.’ (lit.: ‘Your returning is very late.’) (Context: addressees are just coming back)
- (25) *n-as sot-tino:nak loy it̪ sal-na=∅*
 OBL-ART.N other-year ITN 1 search-DR=1 SG
 ‘Next year I’ll look for you.’

- (26) *jampa=∅ inla chot n-as kay-wa=∅*
 do_like=1SG PRO.1SG HAB OBL-ART.N eat-NMZ=1SG
 ‘I always do it when I eat.’ (lit.: “... at my eating.”)

The Absential article indicates that the time or situation has occurred earlier on the same day, as is the case in (27) (partly repeated from (14) above):

- (27) *jayna kino’ chi:~chi no-kos ima:yoj*
 DSC DEM.F.AB MD~go_out OBL-ART.N.AB morning
 ‘She (absent) has gone out (today) in the morning.’

The occurrence of the Absential article with a temporal function could also be observed in (10) on page 100.

The Past article indicates that the time or situation has occurred before the day of speaking, as in (28) (partly repeated from (13) on page 101) and (29).

- (28) *lassinko os joyaj-wa=i*
 five_o’clock ART.N.PST arrive-NMZ=3PL
 ‘At five o’clock (was) their arrival.’
- (29) *n-ot¹² dichi<ye:~>ye=∅ to! manes-pa:na=∅ as*
 OBL-ART.N.PST.1 child<NMZ.N~>=1SG very tasty-APPL-DR=1SG ART.N
tadoy-ni
 sweet-PRC
 ‘When I was a child, oh dear how much I liked sweets.’

The three-way temporal distinction with phrases referring to times and situations can be explained by the fact that these are abstract concepts that, unlike concrete entities, do not have a spatial location (see also Haude 2006: 169–170). Therefore, the Absential article can unambiguously be used for temporal instead of spatial deixis.¹³

There is a direct parallel with noun phrases denoting concrete entities, whose referent must have ceased to exist in order to be referred to with the Past article. In the case of times and situations, the Absential and Past forms can be used only when the situation is concluded. A situation that has started in the past but is going on in the present is always referred to with the Presential article:

- (30) *jayna rey kaw-tino:nak as rey ney-ni-wa=y’hi*
 DSC MOD much-year ART.N MOD here-PRC-NMZ=1PL
 ‘It’s been many years already that we’ve been here.’

Its function in time- and situation-denoting noun phrases shows clearly that the article distinguishes the same major tense categories that are distinguished by the tense particles (see Section 3.3): past (before the day of speaking), immediate past, and nonpast. A closer observation of the examples in Section 3.3 (in particular (6), (7), (9), and (10) on page 100) reveals that the form of the article in this type of noun phrase coincides with the tense particle that establishes the time of the context.

4.3 Absolute concepts: no spatio-temporal differentiation

Certain noun phrases do not participate in the three-way distinction of deictic reference, but generally contain the Presential article in all contexts. These noun phrases refer to concepts that are highly time-stable, do not change location, and are assumed to be generally known, and that can therefore be seen as targets of absolute reference. Geographical locations (31), institutional buildings (32), and personalities of religious worship (33) are concepts of this type:

- (31) *bo rey isne kino' n-as Tirinra*
 because MOD PRO.F.AB DEM.F.AB OBL-ART.N Trinidad
 'Because she (absent) is in Trinidad.'
- (32) *lek-ka-ye:cheŋ ena' nosdé n-as ele:siya*
 kick-MLT-BE.person-R/R DUR.std over_there OBL-ART.N church
 '(The children) kick each other over there at the church.' (speaker being at home, far from the church)
- (33) *uŋ pa:pa=n n-as ari:wa che i'neŋ ma:ma=n di'*
 ART.M.1 father_of=2 OBL-ART.N top and ART.F.1 mother_of=2 REL
ma:ma=u
 mother_of=3M
 'our (incl.) father up there and our (incl.) mother, who is his mother (i. e. Jesus and Mary)'

A similar case is generic reference, which is also carried out with the Presential form (see Section 4.1 above). Consider (34) (and cf. also the phrase *as tadoyni* 'sweets' in (29) on the facing page):

- (34) *kilmo is itila:kwa*
 wicked ART.PL man
 'Men are wicked!'

Noun phrases referring to habitually occurring times and situations, which also contain the Presential article (see (26) on page 106), can be considered as belonging to this class as well.

5 Independent and propositional nominal tense marking

5.1 Independent nominal tense marking with an implicature effect

The spatio-temporal marking on referential phrases in Movima can be described with Nordlinger and Sadler's term of "nominal tense marking" (Section 2): the article marks a temporal relation between the existence of the referent and the moment of speaking (cf. Nordlinger and Sadler's criterion (i)); the combination with different forms of the article is not restricted to particular types of content words (criterion (ii)), the few exceptions only involving reference to absolute concepts; the article occurs on dependent phrases and not on predicate nominals (criterion (iii)). As for criterion (iv), it was mentioned in that not only nouns, but also morphological verbs can occur inside a noun phrase; however, reference is functionally a nominal category (see Croft 2003), and hence the categories encoded by the article can be considered nominal.

The question remains whether we are dealing with the independent or the propositional type of nominal tense marking here. The previous sections have shown that the spatio-temporal deictic properties of the article indicate properties of the referent in relation to the time and place of the utterance. The deictic categories can vary, depending on the type of referent (concrete entity vs. time/situation). They are summed up in Table 3 on the facing page.

It can thus be concluded that Movima nominal tense marking is of the independent type: it locates the time at which the property denoted by the noun holds of the referent. This is independent of the contextual tense, as shown by (35)–(37). These examples all contain a noun phrase with a Past article, but they also contain a marker that rules out a clausal past-tense interpretation. In (35), this is the imperative suffix (*ti*), (36) contains a tense particle implying future reference (*loy*, cf. Section 3.3), and the referential elements in (37) indicate a difference in temporal properties of possessor and possessee (the Presential pronoun *a'ko* for the possessee, the Past article for the possessor). In all three examples, the Past article indicates ceased existence of the referent, but this is independent of the interpretation of the clause as a whole.

Table 3. The spatio-temporal categories encoded by the article

	Entity	Time / Situation
Presential	present absolute	present future habitual
Absential	absent but still in existence	concluded earlier on the same day
Past	absent and out of existence (before the day of speaking)	concluded before the day of speaking

These examples, therefore, clearly identify the Past article as a marker of independent nominal tense.

- (35) *ajlomaj-ti os naye-wa=n*
 tell_about-IMP.DR ART.N.PST marry-NMZ=2
 ‘Tell (her) about your (past) marriage!’
- (36) *loy it to'baycho-poj-chet n-isnos nonok=∅*
 ITN 1INTR remember-CAUS-R/R OBL-ART.F.PST grandparent=1SG
 ‘I’ll remember my late grandmother.’ (elicited)
- (37) *a'ko rey ɬaɬa<kwa~>kwa=os dokwe=∅*
 PRO.N MOD seam<INAL~>=ART.N.PST dress=1SG
 ‘This (a stitched strip of cloth) is the seam of my dress (which does not exist anymore).’ (elicited)

However, except in cases like these, which contain elements indicating nonpast tense, the Past article always determines the interpretation of the entire clause. Consider (38c) in comparison with (38a) and (38b):

- (38) a. *kay-a:poj=∅ as pa:ko*
 eat-DR-CAUS=1SG ART.N dog
 ‘I fed / am feeding / will feed the dog (present).’ (elicited)
- b. *kay-a:poj=∅ kos pa:ko*
 eat-DR-CAUS=1SG ART.N.AB dog
 ‘I fed / will feed the dog (absent).’ (elicited)
- c. *kay-a:poj=∅ os pa:ko*
 eat-DR-CAUS=1SG ART.N.PST dog
 ‘I fed the dog (that is now dead and gone).’ (elicited)

The past interpretation of a clause like (38c) can be understood easily, since it is most common to talk in a past context about entities that have ceased to exist. If the article only indicated a loss of function or possession, such as *ex-* or *former* in English, we would expect to find it used more frequently in nonpast contexts as well.

With phrases referring to times and situations (see Section 4.2), the link between the actual existence of the referent and the discourse time is even more immediate, as is illustrated by the effect of the situation-denoting phrase in (39a)–(39c) (these examples were offered spontaneously during elicitation):

- (39) a. *jayna it ba:lomaj n-as ji:sa-na:-wa=∅ as chakdi*
 DSC 1INTR finish OBL-ART.N make-DR-NMZ=1SG ART.N fence
 ‘I’ll finish making the fence.’ (I’m still building it)
- b. *jayna it ba:lomaj no-kos ji:sa-na:-wa=∅ as chakdi*
 DSC 1INTR finish OBL-ART.N.AB make-DR-NMZ=1SG ART.N fence
 ‘I’ve just finished making the fence.’ (today)
- c. *jayna it ba:lomaj n-os ji:sa-na:-wa=∅ as chakdi*
 DSC 1INTR finish OBL-ART.N.PST make-DR-NMZ=1SG ART.N fence
 ‘I’ve finished making the fence.’ (before today; fence is still there)

Thus, we are dealing here with independent nominal tense marking that has an effect on the temporal interpretation of the proposition. So far, this is an effect of implicature only, and not genuine tense marking (cf. Comrie 1985: 23–26).

However, even though implicational, the article provides temporal information consistently throughout the discourse. This is mainly due to the fact that in Movima, subordination is carried out through nominalization, which produces nouns that are highly sensitive for temporal location (see Section 4.2). Whereas other verbally tenseless languages use aspect and mood morphology for temporal reference, the lack of verbal tense marking in Movima is compensated by the deictic properties of the article.

Still, while the propositional tense-marking effect of the article is largely due to implicature, there are signs that the article also has the potential of indicating propositional tense directly, as the following section will show.

5.2 The article as a marker of propositional tense

In the previous sections I have shown how the article indicates the location and existence of its referent. However, this was a slight simplification of the facts. First of all, it has to be kept in mind that a speaker cannot always know for sure whether an absent discourse referent is still in existence or not. The choice of the article only reflects the speaker's assumptions to that effect.

Second, also for Movima speakers it does often not really matter whether an entity still exists or not. In fact, entities that are not relevant to the speaker at the time of speaking are generally referred to with the Past article in past-tense contexts, no matter whether they still exist or not. Typically, inanimate objects, like the car in (40), are referred to with the Past article in past-tense contexts, even though they may still be existing somewhere (the situation narrated in (40) had only taken place on the previous day):

- (40) *jayna lista da' n-os joyaj-wa=os awto jayna*
 DSC ready DUR.NSTD OBL-ART.N.PST arrive-NMZ=ART.N.PST car DSC
 'She was ready when the car arrived.'

Example (41) illustrates the use of the Past article with reference to a spider that has explicitly not been killed and may still be alive. However, as in (40) above, the Past article is used, because what is relevant to the speaker here is to maintain temporal reference in discourse, and not the possible ongoing existence of the spider:

- (41) *kas rey tikoy-na:-wa=0-as, os si:wa merek*
 NEG MOD kill-DR-NMZ=1SG--3N.AB ART.N.PST spider big
 'I didn't kill it, the big spider.'

In past contexts, even noun phrases referring to living human beings can contain the Past article. This is also the case when the person referred to is not relevant for the speaker at the time of speaking. In (42), for example, some kind of clerk is referred to:

- (42) *jayna baw<a>ra='ne os itila:kwa nokodé*
 DSC pay<DR>=3F ART.N.PST man over_there
 'She has already paid the man over there.'

In contrast, human referents that still play a role in the speaker's life are normally referred to by the Presential or Absential article (depending on their location), even if the narrated situations have occurred a long time ago:

- (43) *n-os* *to<chi~>chik-a='nes majni=∅ as-na=y'ti n-as*
 OBL-ART.N.PST little<NMZ.N~>-LV=ART.F child=1SG sit-LOC=1PL OBL-ART.N
Peru
 Peru
 'When my daughter was little (lit.: "at the (past) being small of my (present) daughter"), we lived in El Perú.'¹⁴

However, sometimes the Past article is used (by some speakers more often than by others) even when the nominal referent is present at the speech situation or when its existence is relevant for the speaker. For instance, in (44) the speaker refers to her own body part with the Past article, something which is not normally done. In (45), the speaker's parents are referred to with the Past article, even though shortly before, it had been pointed out that they are still alive. The article is chosen here not according to a deictic property of the referent, but according to the time of the context.

- (44) *jayna n-os* *imayni jayna tivijni os* *chodo:wi=∅*
 DSC OBL-ART.N.PST night DSC hurt ART.N.PST stomach=1SG
 'Then at night, my stomach hurt.'
- (45) *n-asko* *ela-na=us* *pa'* *isnos* *ma'*
 OBL-PRO.N.AB stay_behind-DR=ART.M my_father ART.F.PST my_mother
 'At that (time) my father left my mother.' (both absent, but alive)

These examples show that the choice of the article is not only determined by properties of the referent, thereby having an implicational effect on the temporal interpretation of the proposition. It can also serve as a device for directly marking temporal relations in discourse. In such cases the Movima article is clearly a marker of propositional nominal tense.

6 From independent to propositional nominal tense

I have shown that in principle, nominal tense marking in Movima is of the independent type, the choice of the article depending on deictic properties of the referent and having only an implicational effect on the temporal interpretation of the clause. However, the Past article can also be used to mark

discourse tense directly, thereby functioning as a marker of propositional nominal tense. There appears to be a gradual shift of nominal tense marking in Movima from purely independent towards propositional, as represented in Figure 1:

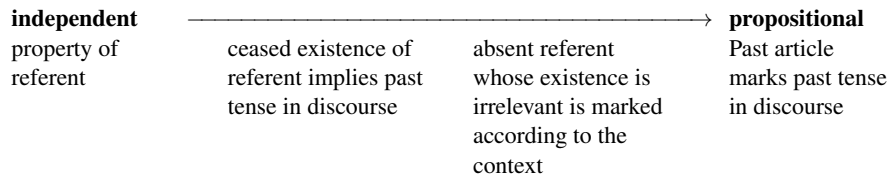


Figure 1. The shift from independent to propositional tense marking

In Figure 1, prototypical independent nominal tense marking is given on the left-hand side. Here, a temporal property of the referent is marked independently of the discourse tense (see (35)–(37) on page 109). The fact that this marking usually has an effect on the interpretation of the proposition as a whole (as shown in (38) and (39) on page 110) is a step towards propositional temporal marking. When entities are referred to by the past form, even though they are still in existence (as in (40)–(42) on page 111), and even present or relevant to the speaker (as in (43)–(45)), then the article does not indicate a temporal property of the referent anymore. It serves as a propositional nominal tense marker that provides temporal information of the clause as a whole.

To conclude, there are basically three factors that contribute to the unusual, perhaps even unique, way of marking temporal reference in Movima:

1. The lack of verbal tense marking is not compensated by modal or aspectual morphology on the verb, but by the temporal-deictic function of the article, an element inherent to nominal expressions.
2. The article, by indicating ongoing or ceased existence of the referent, can by implication indicate clausal tense.¹⁵
3. The article can also be chosen to indicate clausal tense directly. This function is comparable to verbal tense marking in other languages.

Thus, while primarily of the independent type, this nominal tense system also has a propositional tense-marking function. This conflation of independent and propositional nominal tense marking has so far not been described for any other language.

Abbreviations

1 = first person; 2 = second person; 3 = third person; AB = Absential; ANT = anterior; APPL = applicative; ART = article; BE = bound nominal element; BR = bound root; CAUS = causative; DEM = demonstrative; DR = direct; DUR = durative; DSC = discontinuous; EV = evidential; F = feminine; FUT = future; INAL = inalienable; INTR = intransitive; ITN = intentional; LOC = location; LV = linking vowel; M = masculine; MD = middle; MLT = multiple; MOD = modal; N = neuter; NMZ = nominalization; NMZ.N = nominalization of noun; NSTD = nonstanding; OBL = oblique; PL = plural; PRO = free pronoun; PST = past; REL = relativizer; REM = remote past; R/R = reflexive / reciprocal; SG = singular; STD = standing

Acknowledgments

I wish to thank the Movima speakers who taught me their language, in particular the following, who contributed the data presented here: Esaltación Amblo, Julia Malale, Balvina Almaquio, Eligardo Chirimani, Etelvina Gualusna, Griselda Cáumol, Herlan Rojas. I furthermore thank Loretta O'Connor and Michael Cysouw for their comments on an earlier draft of this paper.

Notes

1. I owe the characterization of Movima nominal tense marking in the title of this paper to Djoeko Veninga of the Netherlands radio station VPRO.
2. Note that Nordlinger and Sadler's definition is a bit vague, as it does not state in what way "tense" on nouns should be the same category as on verbs. See also the discussion in Tonhauser (2006).
3. The Movima article does not indicate definiteness; however, for the sake of consistency, it is here generally translated with the English definite article.
4. The symbol -- represents "external cliticization", whose only phonetic effect is a resyllabification with a host-final consonant; it contrasts with "internal cliticization", represented as =, which triggers stress shift and the appearance of an epenthetic /a/ on consonant-final hosts (Haude 2006: 97–101). The two processes, which are restricted to referential elements, mark the status of a nominal constituent as internal or external to the head.
5. While nouns and verbs in Movima have many properties in common, they can be distinguished by the fact that certain morphemes can occur only on nouns, but not on verbs. Adjectives form a subclass of nouns.
6. This morpheme is inserted after the first iambic foot of the base and appears with an initial /k/ after vowels (cf. Haude 2006: 78–83).
7. Bybee et al. (1994: 98) characterize these past-tense categories as "pre-hodiernal" and "hodiernal", respectively; see also Dahl and Velupillai (2005).

8. Inside a noun phrase *loy* is a marker of negation.
9. Depending on the context, this clause can also have a habitual meaning: ‘...I always come to see you’ (cf. Section 4.2).
10. In addition to gender and number, presential demonstratives indicate perceivability, position, and relative distance of the referent (see Haude 2006).
11. Because of the broader applicability of the Presential article (see also Section 4.3), it is not marked in the gloss as ‘Presential,’ but left unmarked.
12. The first-person marker *ɬ* is fused with the article in possessive noun phrases.
13. See Haude (2004: 87) on the possible analogy of the temporal and spatial categories in terms of “presence” and “accessibility”.
14. Note that in this example, also cited in Haude (2004: 84), it is not the case that the same person is referred to twice (cf. Tonhauser 2006: 347); rather, the referent of the first noun phrase is the state of being small, whereas the second noun phrase (*nes majni=∅*) refers to the possessor of that state (as shown by the literal translation).
15. However, the examples of propositional nominal tense marking in the Arawak language Chamicuro given in Nordlinger and Sadler (2004) and Parker (1999) hint at a possible conflation of the two types in this language as well: in all examples of propositional past-tense marking, the referent has either ceased to exist, or the examples do not hint at the contrary. More data would be needed to establish whether in Chamicuro, too, propositional nominal tense marking is an implicature effect of independent nominal tense marking.

References

- Bybee, Joan, Revere Perkins, and William Pagliuca
 1994 *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago/London: Chicago University Press.
- Comrie, Bernard
 1985 *Tense*. Cambridge: Cambridge University Press.
- Comrie, Bernard
 1994 Tense. In *The Encyclopedia of Language and Linguistics*, Ronald E. Asher and J. M. Y. Simpson (eds.), 4558–4563. Oxford etc.: Pergamon Press.
- Croft, William
 2003 *Typology and Universals*. Second edition. Cambridge: Cambridge University Press.
- Crystal, David
 1997 *A dictionary of linguistics and phonetics*. Oxford: Blackwell.
- Dahl, Östen and Viveka Velupillai
 2005 Tense and Aspect. In *The World Atlas of Language Structures*, Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 266–272. Oxford: Oxford University Press.
- Haude, Katharina
 2004 Nominal tense marking in Movima: nominal or clausal scope? In *Linguistics in the Netherlands 2004*, Leonie Cornips and Jenny Doetjes (eds.), 80–90. Amsterdam/Philadelphia: Benjamins.

Haude, Katharina

- 2006 *A grammar of Movima*. Doctoral dissertation, Radboud University Nijmegen.
http://webdoc.ubn.ru.nl/mono/h/haude_k/gramofmo.pdf

Haude, Katharina

- 2009 Reference and predication in Movima. In *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, Patience Epps and Alexandre Arkhipov (eds.), 323–342. (= Trends in Linguistics, Studies and Monographs; 217). Berlin / New York: Mouton de Gruyter.

Nordlinger, Rachel and Louisa Sadler

- 2004 Nominal tense marking in crosslinguistic perspective. *Language* 80.4: 776–806.

Parker, Steve

- 1999 On the behavior of definite articles in Chamicuro. *Language* 75.3: 552–562.

Smith, Carlota

- 2008 Time with and without tense. In *Time and Modality*, Jacqueline Guerón and Jacqueline Lacarme (eds.), 227–249. (= Studies in natural language and linguistic theory; 75). Dordrecht: Springer.
<http://uts.cc.utexas.edu/~carlota/papers/Paris article 8-18-06.pdf>

Tonhauser, Judith

- 2006 *The temporal semantics of noun phrases: evidence from Guaraní*. Doctoral dissertation, Stanford University.

The accentual system of Hocak

Johannes Helmbrecht

1 Introduction

Languages with so-called “fixed stress location” are defined in the following way by Goedemans and van der Hulst:

By [fixed stress location] we mean that stress is located on the same syllable in each word. The location is independent of the weight (usually quantity) of the syllables in the word and is determined with reference to a word edge only. (Goedemans and van der Hulst (2005a: 62)

In languages with fixed stress location primary stress is put either on the first, second, or third syllable counted from the left edge of the word, or on the antepenultimate, penultimate, or ultimate syllable counted from the right edge of the word. The cross-linguistic distribution of these possibilities is summarized in Table 1 on the following page. According to Goedemans and van der Hulst (2005a), more than half of the languages of a 500 languages sample belong to this type of stress placement; i. e. 281 languages out of 500 languages have a fixed stress location system. The majority of these languages places main stress either on one of the first two syllables of the word (108 languages), or on one of the last two syllables of the word (160 languages). There are only a few (12 languages) that place main stress on the antepenultimate syllable, and there is only one language that places main stress on the third syllable from the left edge of the word. This language is Hocak.

So, Hocak is cross-linguistically highly unusual with regard to stress assignment. However, the classification of Hocak as a language that stresses the third syllable from the left edge of the word is only half of the truth. There are many words and word forms that do not bear main stress on the third syllable, cf. the words and word forms in Table 2 on the next page. The words and word forms a.–d. in are not rare exceptions in the lexicon of Hocak, but systematic deviations from the proposed third syllable-from-the-left rule taken to be the main stress rule in Hocak by Goedemans and van der Hulst (2005a). The examples illustrate that primary stress can fall systematically on the sec-

Table 1. Typology of languages with fixed stress location (Goedemans and van der Hulst 2005a)

Linguistic types of fixed stress location	Number of languages
No fixed stress	219
Initial: stress is on the first syllable of the word	92
Second: stress is on the second syllable of the word	16
Third: stress is on the third syllable of the word	1
Antepenultimate: stress is on the third syllable from the right edge of the word	12
Penultimate: stress is on the second syllable from the right edge of the word	110
Ultimate: stress is on the last syllable of the word	50
Total	500

Table 2. Words with primary stress not on the third syllable

Pattern	Example	Gloss
a. CVV'CVCV	<i>booráxux</i>	'you break something into pieces'
b. CV'CVVCV	<i>hajáira híi</i>	'to do more of it'
c. CVCVCV'CV	<i>gikəŋəhé</i>	'to invite somebody'
d. CVCVCV'CVCV	<i>gipinipini</i>	'twirl'

ond (cf. a.–b.), and even on the fourth syllable (cf. c.–d.) from the left edge of the word.

In particular the examples a. and b. make it likely that syllable weight does play a role in Hocak. Both words have three syllables with primary stress on the second syllable, and both contain a heavy syllable, i. e. a long vowel in a. and a diphthong in the second syllable in b. Such a factor was explicitly excluded in the definition of fixed stress location systems in Goedemans and van der Hulst (2005a). Consequently, there is a separate typology by the same authors on so called “weight-sensitive stress” systems (cf. Goedemans and van der Hulst 2005b). I will show in this paper that syllable weight plays indeed a role in the determination of stress location of Hocak. When one of the first two syllables is a heavy syllable, then the main stress falls on the second syllable (cf. Susman 1943; Miner 1979, 1981, 1990; Hale 1985; Helmbrecht in prep.). This property, however, does not allow classifying Hocak alternatively

as language with weight-sensitive stress either, since it is not the heavy syllable that attracts primary stress *per se*.

So, Hocąk is a typological rarity with regard to main stress placement — it places main stress regularly on the third syllable of the word counted from the left edge under certain conditions. However, Hocąk does not fit easily into the typology set up by Goedemans and van der Hulst. As will be shown in this paper, Hocąk is indeed basically a fixed stress location language in which the primary stress location is determined from the left edge of the word, but syllable weight also plays a role in the determination of the primary stress location. Further, contrary to the classification of Hocąk as a language stressing the third syllable from the left edge, we find systematic deviations from this rule.

In this paper, a detailed description of the rules for the location of primary and secondary stress of Hocąk will be presented. This description allows a better understanding of a) the rarity of the accentual system of Hocąk, and b) why it could not be covered correctly by the typology set up by Goedemans and van der Hulst in their typology referred to above. In addition, I will provide some comparative evidence showing that Hocąk has undergone a stress shift one mora to the right at one point in its history, while its closest relative, Chiwere, did not. All other Siouan languages have main stress on the second syllable or second mora (there is some variation, but none has the rules found in Hocąk). It is this stress shift that is responsible for the typological highly unusual stress placement rules.

2 The Language

Historically, *Winnebago* was the name which was used for the language and the people of the Hocąk [hotʃaŋk] tribe by their Algonquian neighbors. This name was subsequently taken over by anthropologists and linguists and is still found in the literature. Since *Hocąk* is the self-denomination of the tribe, this name is used here and elsewhere to refer to the Hocąk people and their language.

Hocąk is a North American Indian language. It is a member of the Siouan language family (Mithun 1999: 501–510). Chiwere is the closest relative to Hocąk in the Mississippi Valley subgroup of Siouan and these two languages have some degree of mutual intelligibility. Chiwere and Hocąk share many similarities and innovations in comparison to the other Mississippi Valley Siouan languages. They are usually grouped together as one subgroup of

the Mississippi Valley branch of Siouan besides Dakota and Dhegiha (cf. Goddard 1996: 8). Hocak is still spoken by a small group of elders (about 200 people) in Wisconsin and Nebraska. Needless to say, Hocak is a highly endangered language.

Three sources of language data are employed in this paper. First, data are taken from previous publications on stress in Hocak, most importantly the dissertation by Susman (1943) and the articles by Miner (1979, 1990). The second source is the Hocak text corpus, and the Hocak/English dictionary database created by the DoBeS project “Documentation of the Hocak language” funded by the Volkswagen Foundation.¹ Currently, this project creates a documentation of the Hocak language on the basis of audio and video recordings, which are transcribed, glossed and translated. The third source of data are my own field notes which were collected over the last fifteen years during various field trips.

3 Basic rules of stress placement in Hocak

3.1 Primary stress

Hocak is a mora counting, but syllable accenting language. The rules that account for most cases of primary stress placement are shown in (1). The rules in (1a–c) apply to a single stress domain, which usually coincides with the inflected word form including derivations and compounds. Many problems in the description of stress placement in Hocak, as for example present in Susman (1943), originate from the wrong determination of stress domains. The words in Table 3 on the facing page illustrate the basic rules for the assignment of primary stress in Hocak.²

(1) Primary stress assignment in Hocak

- a. Primary stress falls on the third mora counted from the left boundary of the word.
- b. If disyllabic words have only two moras (i. e. two syllables with a short vowel as nucleus), primary stress falls on the second mora (i. e. the second syllable).
- c. Monosyllabic words always have a long vowel, i. e. two moras. Here, stress falls on the first mora.

Table 3. Stress placement patterns: primary stress

Patterns	Examples	English Gloss
1) one syllable, two moras	<i>zî</i> <i>šgáac</i>	‘yellow’ ‘to play’
2) two syllables, two moras	<i>hiwáx</i> <i>hošgác</i>	‘to ask’ ‘playground’
3) two syllables, three moras	<i>cîinák</i> <i>booká</i>	‘village’ ‘to knock over’
4) three syllables, three moras	<i>wanġġk</i> <i>gġiré</i>	‘bird’ ‘to help’
5) three syllables, four moras	<i>taanġzu</i> <i>bookéwe</i>	‘sugar’ ‘to fall down’
etc.		

3.2 Primary stress on diphthongs

Hocak distinguishes between long and short vowels phonologically and has a set of three nasalized vowels. Diphthongs do not play a significant role in the segmental phonology of Hocak. There are only a few instances of them in the lexicon. However, diphthongs often emerge because of the deletion of consonants in various morphophonological processes, or simply as a result of the inflection of verbs. For instance, the verb *woogícexi* ‘to be difficult for someone’ is conjugated for the first and second person by means of affixes which produce diphthongs in combination with the long vowel in the first syllable; compare the forms in (2).

(2) *wóigicèxi*

woo-í-gicèxi

ISC-U.1SG-be.difficult.for

‘It is difficult for me’, ‘I am afraid to do it.’

The first person singular undergoer pronominal affix *i-/hi-* is infixes between the first syllable *woo-* and the rest of the word stem *-gicexi*.³ The resulting syllable would contain three moras and thus would receive primary stress on the third mora according to the rules in 1. However, since syllables made out of three moras are not permitted in Hocak, the long vowel of *woo-* is shortened. The primary stress does not rest on the pronominal infix, but

is moved one mora to the left to *wó-*, the so-called “initial stem component” (ISC) of the example.⁴ This movement is governed by the Vowel Strength Hierarchy proposed by Susman (1943) which controls primary stress placement in diphthongs in Hocak. According to this hierarchy (cf. (3)), open vowels are stronger than closed ones and back vowels are stronger than front vowels. The Vowel Strength Hierarchy determines whether the first vowel or the second vowel in a long syllable with a diphthong is stressed, no matter where the diphthong comes from. In all cases the stronger vowel of the diphthong is stressed.

(3) Vowel Strength Hierarchy (Susman 1943):

STRONG a o u e i WEAK
 ạ ụ ị

The noun *'áipa* ‘bracelet’ is one of the few words in the Hocak lexicon with a diphthong. The fact that diphthongs in Hocak emerge primarily by means of morphological processes (e. g. deletion of consonants in the combination of morphemes) makes it highly likely that *'áipa-* is a lexicalized composition of *'áa* ‘arm’ and *hipá* ‘near’. In this case, /h/ of *hipá* drops, and the long vowel of *'aa* is reduced. Primary stress is then set following the Vowel Strength Hierarchy.

3.3 Secondary stress

The syllable with primary stress may be followed by syllables which receive a secondary accent. Secondary stress is not as important as primary stress in Hocak. For example, secondary stress is never distinctive and only plays a minor role in the process of marking word boundaries. Table 4 presents some examples which illustrate the distribution of primary and secondary stress and (4) gives the rule for secondary stress assignment.

Table 4. Stress placement patterns: secondary stress

Patterns	Examples	English Gloss
(1) Four syllables, five moras	<i>taan̩ʒurà</i>	‘the sugar’
(2) Six syllables, seven moras	<i>haakítujìkgajà</i>	‘after I pulled’

(4) Secondary Stress Rule:

Words that are longer than three moras receive primary stress on the third mora and a significant weaker secondary stress on each even numbered mora thereafter.

Phonetically, secondary stress is much weaker than primary stress and often difficult to identify in recordings, which may be caused by the enclitic nature of many grammatical formatives following the verb stem.⁵

As was said in Section 1, there are many systematic deviations from the third-mora-from-the-left-edge rule, cf. (1) on page 120, which contradict the typological classification of Hocak by Goedemanns and van der Hulst (2005a). The main reason is that the basic rules interact with phonological and morphological rules resulting in deviations of various kinds.

4 Interaction of stress placement with phonological rules

4.1 Dorsey's Law and Dorsey's Law-sequences

It was in the late 19th century that Dorsey (1883) discovered two important sound laws separating Hocak from the closely related language Chiwere (and the other Siouan languages as well, but these will not be considered here). First, he found that there is a regular interruption of obstruent-sonorant consonant clusters in Hocak by means of a vowel-copying process, while the same clusters remained unaffected in Chiwere. The second important correspondence Dorsey found was the regular stress shift from the second syllable or mora in Chiwere to the third syllable or mora in Hocak.

The first example of Table 5 on the next page illustrates the vowel-copying rule, coined *Dorsey's Law* by Miner (1979: 25, 33). The Hocak word *hoikéwe* 'to go back inside' consists of a lexicalized initial stem part *hoi-* (presumably derived historically from a combination of two locative applicatives *ho-* 'in' + *hi-* 'towards' of which *hi-* is no longer productive) and stem *-kewe*. The cognate form of *-kewe* in Chiwere is *-gwe*, with the vowel /e/ missing between the obstruent /g/ and the sonorant /w/. Evidence from other Siouan languages shows that Chiwere did not lose this vowel, but instead that Hocak inserted the vowel into the obstruent-sonorant cluster. For further examples, compare the forms in Table 5. These examples also exhibit other sound correspondences such as /g/ ~ /k/, /r/ ~ /l/, /r/ ~ /n/, /s/ ~ /θ/, and /p/ ~ /b/, which will not be further discussed here.

Table 5. Some Hocąk-Chiwere cognates illustrating vowel-copying in Hocąk (cf. Miner 1979: 27)

Hocąk	Chiwere	English Gloss
<i>hoikéwe</i>	<i>ugwé</i>	‘go back inside’
<i>jiikére</i>	<i>jiglé</i>	‘start’
<i>wasúnú</i>	<i>waθlú</i>	‘cook over open fire’
<i>raparás</i>	<i>labláðe</i>	‘tear apart with mouth’

Table 6. Some Hocąk-Chiwere cognates illustrating the Hocąk stress shift (cf. Miner 1979: 31)

Hocąk	Chiwere	English Gloss
<i>wašośé</i>	<i>wasóse</i>	‘brave’
<i>wagujé</i>	<i>agúje</i>	‘footwear’
<i>wišćjéga</i>	<i>miščjje</i>	‘Hare’
<i>wanaǵí</i>	<i>wanáxe</i>	‘spirit’
<i>hinǵbáha</i>	<i>inúwenaha</i>	‘second’

A few cognate forms in Hocąk and Chiwere that illustrate the stress shift are presented in Table 6. Comparative Siouanists believe that both processes – vowel-copying in obstruent-sonorant clusters and stress shift to the right – are causally related. Wolff (1950: 172–173) claims that the stress shift preceded the vowel copy process, while Miner (1979: 25, 33) claims that the vowel copy process preceded the stress shift. I will leave this question open for further historical-comparative research. For the synchronic analysis of stress in Hocąk, this question as well as the potential answers to it are not relevant. In contrast, the vowel-copying process called Dorsey’s Law is still a synchronic phonological rule in Hocąk, as was shown convincingly in Miner (1979, 1990). The evidence will briefly be summarized in the various subsections to follow (cf. Sections 4.2 through 4.6).

To present the evidence that proves that Dorsey’s Law is a synchronic phonological rule in Hocąk, words containing so-called “Dorsey’s Law-sequences” (short: DL-sequences) have to be defined. They can be defined almost entirely in phonological terms as presented in (5).

- (5) A sequence of the form $[C_1 V_1 C_2 V_2]$ is called a DL-sequence iff:
- C_1 is voiceless obstruent /p, k, c, s, š, x/
 - C_2 is one of the following sonorant consonants /n, r, w /

- $V_1 = V_2$
- The sequence is spoken faster than average CVCV sequences

Some examples of Dorsey's Law-sequences illustrating the rules in (5) are given in (6). The Dorsey's Law-sequences are rendered in boldface. Table 7 gives an exhaustive list of Dorsey's Law-sequences in Hocak. There is only one Dorsey's Law-sequence allowed by the rules in (5) which does not occur, namely the cV_1nV_1 sequence.

- (6) ***kaṇqák*** 'put something not having length'
keré 'depart returning'
hiperés 'know'
hakewé 'six'

Table 7. List of Dorsey's Law-sequences in Hocak

-pV ₁ nV ₁ -	-šV ₁ nV ₁ -	-kV ₁ nV ₁ -	-xV ₁ nV ₁ -	-sV ₁ nV ₁ -	-cV ₁ wV ₁ -
-pV ₁ rV ₁ -	-šV ₁ rV ₁ -	-kV ₁ rV ₁ -	-xV ₁ rV ₁ -	-sV ₁ rV ₁ -	-cV ₁ rV ₁ -
-pV ₁ nV ₁ -	-šV ₁ wV ₁ -	-kV ₁ wV ₁ -	-xV ₁ wV ₁ -	-sV ₁ wV ₁ -	

The following sections explicate the evidence that the Dorsey's Law-sequences are a synchronic rule in Hocak. The main point is that Dorsey's Law-sequences are affected as a whole by various synchronic phonological processes. For instance, reduplication affects in most cases only open syllables. However, in Dorsey's Law-sequences the whole sequence is reduplicated (cf. Section 4.2). The same holds for nasalization (cf. Section 4.3), stem final metaphony (cf. Section 4.4), inflection (if Dorsey's Law-sequences are produced, cf. Section 4.5), and for stress placement (cf. Section 4.6).

4.2 Reduplication of Dorsey's Law-sequences

That Dorsey's Law is a synchronic phonological rule in Hocak can be shown in reduplication patterns that can be observed in Hocak. Like most of the other Siouan languages, Hocak has extensive reduplication as illustrated in examples in (7) through (9). Normally, only CV or CCV sequences are reduplicated.

- (7) a. *gihú* 'swing'
b. *gihuhú* 'wag tail'

- (8) a. *hit'é* 'speak, talk'
 b. *hit'et'é* 'speak, talk'
- (9) a. *racgá* 'drink'
 b. *racgacgá* 'drink repeatedly'

In contrast, Dorsey's Law-sequences of the form CVCV reduplicate as a whole. They are the only sequences longer than C(C)V which may reduplicate, as illustrated in the examples in (10) through (12). This special behavior follows naturally when Dorsey's Law-sequences are interpreted as the result of a regular synchronic derivation from underlying CCV sequences.

- (10) a. *-keréš* 'make designs'
 b. *kerekéreš* 'spotted'
- (11) a. *cíwí* 'sound causing vibration'
 b. *cíwícíwí* 'sound causing vibration'
- (12) a. *šará* 'bald'
 b. *šarašára* 'bald in spots'

4.3 Nasalization in Dorsey's Law-sequences

The peripheral vowels /a/, /i/, and /u/ in Hocak have an oral /nasal distinction. However, there is a synchronic rule which causes these vowels to be automatically nasalized after a nasal consonant /n/ or /m/. Thus, the language has syllables such as *kə*, *tí*, *mə*, *ní*, but none like **ma* or **ni*. This means, the opposition between oral and nasal vowel is neutralized after nasal consonants. There are no oral counterparts to the words in (15 a–c), in contrast to the pairs in (13) and (14).

- (13) a. *haák* 'rear part'
 b. *haák* 'woodchuck'
- (14) a. *gisú* '(de)husk'
 b. *gisú* 'upset'
- (15) a. *máq* 'earth'
 b. *ní* 'water'
 c. *wamqnúke* 'thief'

In Dorsey's Law-sequences, both vowels are nasalized, i. e. the original and the copied one. The original vowel is automatically nasalized after a nasal consonant before it is copied into the obstruent-nasal consonant cluster. Since the automatic nasalization of the vowel after nasal consonants is a synchronic rule of Hocak, copying this nasalized vowel in a Dorsey's Law-sequence must be a synchronic rule, too. The result of these processes is illustrated in (16 a–c).

- (16) a. *kənák* 'marry'
 b. *siní* 'cold'
 c. *boopúnũš* 'hit at random'

4.4 Stem final /-e/ to /-a/ metaphony

Another synchronic rule of Hocak is that stem final /-e/ is changed to /-a/ before certain suffixes. This metaphony is widespread among Siouan languages, particularly among the Mississippi Valley group. It is illustrated in example (17 a–b). The SBJ.3PL suffix *-ire* triggers a change of /-e/ to /-a/ in the final vowel of *máqcé* in (17b).

- (17) a. *máqcé* 'he cuts off a piece (of soft substance)'
 b. *máqcá-ire* 'they cut off a piece (of soft substance)'
 cut.off-SBJ.3PL

However, when the root is a Dorsey's Law-sequence, both vowels – the original one and the copied one – are changed; cf. the examples (18) through (20).

- (18) a. *keré* 'to leave returning'
 b. *kara-íre* 'they leave returning'
 leave.returning-A.3PL
- (19) a. *máqpére* 'to slice thin'
 b. *máqpára-na* 'could slice thin'
 slice.thin-POT
- (20) a. *gisewé* 'to calm down'
 b. *gisawa-nák* 'calming down (sitting)'
 calm.down-be(sitting)

So, Dorsey's Law applies after the /-e/ to /-a/ metaphony. And because the e/a-sound change is a productive rule in Hocak, Dorsey's Law must be productive too. The order of the application of these rules reflects the historical development. It is assumed that the stem final e/a-sound change goes back at least to Proto Mississippi Valley Siouan, while Dorsey's Law represents a late development in Hocak.

4.5 Dorsey's Law-sequences across morpheme boundaries

When a root or suffix ending in an obstruent is followed immediately by a root or suffix beginning with a resonant, Dorsey's Law does not apply. In (21b), the nominal stem *wanɨk* 'bird' ends in an obstruent followed by the nasal consonant of the diminutive suffix *-nɨk*. This is a Dorsey's Law-sequence, but Dorsey's Law does not apply across this morpheme boundary.

- (21) a. *wanɨk* 'bird'
 b. *wanɨg-nɨk* 'little bird'

In contrast, there are situations in which Dorsey's Law applies even across morpheme boundaries. Specifically, two prefixes, each consisting of a single obstruent, will trigger Dorsey's Law when attached to a stem beginning with a resonant. This can be demonstrated with regard to the A.2SG prefix of the second conjugation /š-/ (cf. the examples in (22) and (23)), and with regard to the possessive reflexive marker /k-/ (cf. the examples in (24) and (25)). When verb stems do not begin with a resonant, Dorsey's Law does not apply.

(22) Stems not beginning with a resonant

- | | | |
|----|--------------------|-----------------------|
| a. | <i>guú</i> | 'leave returning' |
| | š- <i>guú</i> | 'you leave returning' |
| | A.2SG-leave.return | |
| b. | <i>t'eé</i> | 'die' |
| | š- <i>jeé</i> | 'you die' |
| | A.2SG-die | |
| c. | <i>'ǰǰ</i> | 'live' |
| | š- <i>'ǰǰ</i> | 'you live' |
| | A.2SG-live | |

(23) Stems beginning with a resonant

- | | | |
|----|-----------------|-------------|
| a. | <i>waší</i> | ‘dance’ |
| | <i>ša-waší</i> | ‘you dance’ |
| | A.2SG-dance | |
| b. | <i>rugás</i> | ‘tear’ |
| | <i>šu-rugás</i> | ‘you tear’ |
| | A.2SG-tear | |
| c. | <i>reé</i> | ‘go’ |
| | <i>še-reé</i> | ‘you go’ |
| | A.2SG-go | |

The possessive reflexive prefix /k-/ indicates that the actor of the event designated by the verb possesses the undergoer of the same event (‘A does it to his own U’). The possessive reflexive marker also triggers Dorsey’s Law across morpheme boundaries, when the stem begins with a resonant; cf. the examples in (24) and (25).

(24) Stems not beginning with a resonant

- | | | |
|----|--------------------|------------------------|
| a. | <i>’uú</i> | ‘make, wear’ |
| | <i>k-’uú</i> | ‘make, wear one’s own’ |
| | POSS.REFL-wear | |
| b. | <i>hi’é</i> | ‘find’ |
| | <i>hi-k-’é</i> | ‘find one’s own’ |
| | ISC-POSS.REFL-find | |

(25) Stems beginning with a resonant

- | | | |
|----|---------------------|-----------------------|
| a. | <i>rušíp</i> | ‘pull down’ |
| | <i>ku-rušíp</i> | ‘pull down one’s own’ |
| | POSS.REFL-pull.down | |
| b. | <i>racgá</i> | ‘drink’ |
| | <i>ka-racgá</i> | ‘drink one’s own’ |
| | POSS.REFL-drink | |

4.6 Stress placement in words with Dorsey’s Law-sequences

In principle, words containing Dorsey’s Law-sequences follow the basic stress placement rules described in Section 3.1. However, there are a few

Table 8. Words with Dorsey's Law sequences that do not follow the third mora rule

Pattern	Example	English Gloss
a. CV[CVCV]'CV	<i>hikurukják</i> <i>hikurunġ</i> <i>hokarawé</i>	'to miscalculate' 'tangled' 'go and meet'
b. CV[CVCV]{'CVCV]	<i>gikaŋaŋaŋap</i> <i>wakiripáras</i> <i>wakirikírik</i>	'shiny' 'flat insect' 'sub-layer of basswood bark (used for healing)'
c. CV[CVCV]{'CVCV][CVCV]	<i>wakiripóropòro</i>	'spherical insect'

patterns that cause special stress placement, as listed in Table 8. All examples there have primary stress on the fourth mora from the left edge of the word which happens to be also the fourth syllable. This does not fit into the basic rules of primary stress placement as presented in Section 3.1. Pattern (a), which is quite frequent in the lexicon of Hocak, suggests that the stress placement rules apply before Dorsey's Law. The inserted copied vowel of the Dorsey's Law-sequence then moves the primary stress one mora to the right (cf. Miner 1990: 156), resulting in a primary stress on the fourth mora. However, this explanation does not work for the two other patterns (b) and (c). If stress were placed on the third mora before the copying rules for the vowels applies, the main stress would have to fall on the last syllable. Fortunately, examples for the patterns (b) and (c) are rare: for (c) the only example I am aware of is the one shown here from Miner (1990: 155).

5 Interaction of stress placement with morphophonological rules

5.1 Prefixes

Hocak has a complex verbal morphology, in particular concerning possible combinations of prefixes. All in all, there are ten prefix slots, which cannot be filled all at a time, or in all combinations that are possible in theory. Some of the prefixes such as the applicatives and the instrumental prefixes have a tendency to lexicalize with the verb stem to the effect that there is systematic infixation in Hocak. For a detailed treatment of this morphological property of Hocak see Helmbrecht and Lehmann (2008).

What is important with regard to stress placement is the fact that pronominal prefixes merge with various other prefixes and parts of the verbal stem in Hocak to the effect that long syllables emerge. In almost all of these cases, the third mora rule applies, but the primary stress then falls on the second syllable or even on the first syllable because of the morphophonological reduction. Speakers of Hocak have even a Hocak term for these morphophonological processes: *hirašórop* ‘to glue it together’. In (26), an example illustrates what is meant with this. The verb *hogirák* ‘to tell someone something’ has an initial stem component *ho-* which changes to *woo-* if it is combined with an U.3PL *wa-* prefix. The /h/ drops and the vowel assimilates to the vowel of the prefix. The result is that the primary stress is placed on the second syllable. In Table 9 on the next page, the most important morphological processes with the same result are summarized.

- (26) *woogírakšəŋə*
wa-hogirak-šəŋə
 U.3PL-he.tells-DECL
 ‘He / she told them.’

Similar morphological processes can lead to the first syllable as the bearer of primary stress. This is exemplified in (27). The personally inflected verb *wáak’uŋə* ‘I gave them something’ contains an U.3PL prefix *wa-*, an initial stem component *ho-*, and an A.1SG *ha-/a-* infix. All three of them merge to a single syllable *wáa-* bearing primary stress. Other morphological processes to the same effect are summarized in Table 10 on page 133.

- (27) *wáak’uŋə*
wa-ho-ha-k’u-nə (< *hok’ú* ‘to give’)
 U.3PL-ISC-A.1E-give-DECL
 ‘I gave them something.’

There are other instances of primary stress placement in Hocak that are not covered by these rules. The combination *wa-* (U.3PL / ‘something’) and *ha-/a-* (A.1SG) at the beginning of a verb tends to attract the primary stress to the effect that the first syllable is stressed. This is a clear violation of the third mora rule, illustrated in (28).

- (28) *wáagikàrahè*
wa-ha-gikarahe
 U.3PL-A.1E-invite
 ‘I invite them.’ (Fieldnotes; cf. Helmbrecht and Lehmann (eds.) (2006: 19))

Table 9. Prefix combinations resulting in CVV'CV... pattern with primary stress on the second syllable

Combinations of pronominal prefixes with applicatives / ISC	Result	Examples
ha- (APPL.SUPESS) + ha-/a- (A.1 SG)	haa-	<i>hapé</i> 'to wait for so.' <i>haapéwi</i> 'we(E) wait for so.'
hi- (A.1 DI) + ha-/a- (APPL.SUPESS)	hij-	<i>hapé</i> 'to wait for so.' <i>hiipéwi</i> 'we(I) wait for so.'
ha- (APPL.SUPESS) + hi-/i- (U.1 SG)	hij-	<i>hapé</i> 'to wait for so.' <i>hiipé</i> 'so. waits for me' / 'you and I wait for sb.'
wa- (U.3 PL / 'sth.') + ha-/a- (A.1 SG)	waa-	<i>wa'ú</i> 'to do, make', <i>waa'ú</i> 'I do, make'
ho- (APPL.INESS) + ha-/a- (A.1 SG)	waa-	<i>hogirák</i> 'to tell so.' <i>waagítak</i> 'I tell him / so.'
hi- (A.1 DI) + ho-/o- (APPL.INESS)	hoo-	<i>hogirák</i> 'to tell so.' <i>hoogítagwi</i> 'we(I) tell him'
ho- (APPL.INESS) + hi-/i- (U.1 SG)	huy-	<i>hogirák</i> 'to tell so.', <i>huygírak</i> 'he / so. tells me'
wa- (U.3 PL / 'sth.') + ho-/o- (APPL.INESS)	woo-	<i>hogirák</i> 'to tell so.' <i>woogírak</i> 'he / so. tells them'
hi- (APPL.INST) + ha-/a- (A.1 SG)	yaa-	<i>higé</i> 'to say to so., call so. a name' <i>yaagé</i> 'I say to, call him a name'
hi- (A.1 DI) + hi-/i- (APPL.INST)	hij-	<i>hipérés</i> 'to know', <i>hiipéreswi</i> 'we(I) know'
hi- (APPL.INST) + hi-/i- (U.1 SG)	hij-	<i>higé</i> 'to say to so., call so. a name' <i>hiigé</i> 'he says to me, calls me a name'
wa- (U.3 PL / 'sth.') + hi-/i- (APPL.INESS)	wii-	<i>hipérés</i> 'to know' <i>wiipéres</i> 'he knows them'
ha- (A.1 SG) + gi- (INSTR.striking)	hai-	<i>gixánq</i> 'to move away', <i>haixánq</i> 'I move away'
ra- (A.2 SG) + gi- (INSTR.striking)	rai-	<i>raixánq</i> 'you move away'
wa- (U.3 PL / 'sth.') + wa- (INSTR.pressure)	waa-	<i>wat'é</i> 'to make it sore' <i>waat'é</i> 'he makes them sore'; but also <i>wawat'é</i> 'he makes them sore'
ho- (APPL.INESS) + gi- (INSTR.striking)	hoi-	<i>git'ék</i> 'to be bruised' <i>hoit'ék</i> 'bruise'
hi- (ISC) + hi-/i- (U.1 SG)	hij-	<i>higé</i> 'to be tired', <i>hiigé</i> 'I am tired'
kara- (A possesses U) + gi- (INSTR.striking)	karaí-	<i>karaíšiš</i> 'to break one's own' (< <i>kara-</i> <i>gi-</i> <i>šiš</i>)

Table 10. Prefix combinations resulting in a 'CVVCV... pattern with primary stress on the first syllable

Combinations of pronominal prefixes with applicatives / ISC	Result	Examples
wa- (U.3PL / 'sth.') + ho- / o- (APPL.INESS) + ha- / a- (A.1SG)	wáa-	<i>hok'yú</i> 'to give them', <i>wáak'yú</i> 'I give them'
wa- (U.3PL / 'sth.') + ho- / o- (APPL.INESS) + hi- / i- (U.1SG)	wói-	<i>woogícexi</i> 'to be difficult for so.' <i>woígicèxi</i> 'It is difficult for me, I am afraid to do it'
wa- (U.3PL / 'sth.') + hi- / i- (APPL.INST) + ha- / a- (A.1SG)	wiá- yáa-	<i>hiperes</i> 'to know', <i>wiáperes</i> 'I know them' <i>yaágawiniššunųnq</i> wa-hi-há-ge-wi-ni-ššunų-nq U.3PL-ISC-A.1E-say-PL-NEG-HAB-DECL 'we usually did not say to them' (Susman 1943: 53)
boo- (INSTR.shooting) + ha- / a- (A.1SG)	boá-	<i>boojánik</i> 'to knock s.o. goofy', <i>bođjanik</i> 'I knock s.o. goofy'
boo- (INSTR.shooting) + hi- / i- (U.1SG)	bói-	<i>booksáp</i> 'to sober up', <i>bóiksap</i> 'I sober up'
mqq- (INSTR.cutting) + ha- / a- (A.1SG)	máq-	<i>máqxóro</i> 'to peel sth.', <i>máqxoro</i> 'I peel sth.'
mqq- (INSTR.cutting) + hi- / i- (U.1SG)	máq-	<i>máqšjǵ</i> 'to be strong', <i>máqšja</i> 'I am strong'
nqq- (INSTR.foot) + ha- / a- (A.1SG)	náq-	<i>náqǵǵú</i> 'to hear, understand', <i>náqǵǵu</i> 'I hear, understand'
nqq- (INSTR.foot) + hi- / i- (U.1SG)	náq-	<i>náqǵǵu</i> 'he hears me'
taa- (INSTR.heat) + ha- / a- (A.1SG)	táa-	<i>taagás</i> 'to tear sth. (paper) by heat', <i>táa-gas</i> 'I tear sth. (paper) by heat'
taa- (INSTR.heat) + hi- / i- (U.1SG)	tái-	<i>taakác</i> 'to be hot', <i>táikac</i> 'I am hot'
roo- (ISC) + ha- / a- (A.1SG)	ráa-	<i>roogú</i> 'to want', <i>ráagu</i> 'I want'
kii- (RFL) + gi- (INSTR.striking)	kü-	<i>giš'ák</i> 'to respect someone' <i>küš'agire</i> 'they respect each other'

5.2 The special case of *wqaga*

The undergoer paradigm of pronominal prefixes contains a morpheme that is possibly better analyzed as a proclitic than as a prefix. It is the U.1DI proclitic *wqaga=*, which is composed of a long and a short syllable [CVVCV] amounting to three moras by itself. If it were an independent word, it should bear primary stress on the third mora followed either by secondary stress on the second mora thereafter, or by a new stress domain. We find indeed cases in our text corpus in which *wqagá=* behaves like an independent word, as shown in (29).

(29) *wqagá guṽswíra*

wqagá guṽs-wi-ra

U.1DI create-PL-DEF

‘He (the great spirit) created us (PL.1).’ (WIL017)

The expression in (29) consists of two stress domains (i. e. there are two independent words) with two primary stresses, one on the third mora of *wqagá*, and the second on the third mora of the verb, i. e. on the plural suffix *wí*. Independent evidence that the expression in (29) consists of two words rather than one inflected verb comes from the fact that the monosyllabic verbal stem *guṽs* ‘to create’ is not shortened. This would be expected if *wqagá* were a prefix, because all monosyllabic verbs have a long stem vowel which is regularly shortened when prefixes are attached to it.

However, in many other cases *wqaga=* is not stressed at all, but has the effect to attract the primary stress to the first following syllable. In (30), *wqagá-* is not stressed itself but causes the primary stress to be on the first syllable of the verbal stem. This effect contradicts the rules given in (1) on page 120. Primary stress here comes to be on the fourth syllable of the word in this case. I have no explanation for this rather exceptional behavior.

(30) *wqagagíkarahè*

wqagá=gíkarahè

U.1DI=he.invites

‘He invites us (first dual inclusive).’ (Fieldnotes; cf. Helmbrecht and Lehmann (eds.) (2006: 19))

In the text corpus of Hocak, *wqaga=* often fuses with the first syllable of the verb stem resulting in primary stress placement on the second syllable as

Table 11. Prefix combinations resulting in a CVV'CVV... pattern with primary stress on the second syllable

Combinations of <i>wqaga=</i> with applicatives / ISC	Result
<i>wqaga-</i> (U.1DI) + <i>ha-</i> (APPL.SUPES)	<i>wqagáa-</i>
<i>wqaga-</i> (U.1DI) + <i>ho-</i> (APPL.INES)	<i>wqagóo-</i>
<i>wqaga-</i> (U.1DI) + <i>hi-</i> (APPL.INST)	<i>wqagú-</i>

illustrated in (31) (cf. Table 11 for a summary of the most important contractions). In these cases, the placement of primary stress obeys the rules, since it is located in the third mora.

- (31) *wqagáaganìnqwi*
wqaga=haganj-ire-wi
 U.1DI=have.for-SBJ.3PL-PL ⁶
 'They have for us (first plural inclusive).' (ECO054)

Rarely, *wqaga=* is infixed in the stem and then fully integrated into the phonological structure of the word as is the case in (32). Here, the first long syllable of *wqaga=* is shortened (like long monosyllabic words, which are shortened when they receive a prefix), and the second vowel /a/ is changed to an unstressed schwa. With regard to stress placement, this occurrence of *wqagá-* is fully in accordance with the general rules for primary stress placement.

- (32) *jaawqágesge hajá(w)inq* (< *jaasgé* 'how')
jaa<wqaga>sgé ha-jée-wi-nq
 <U.1DI>-how A.1E-be(standing.position)-PL-DECL
 'How we are today.' (WIL026)

5.3 Suffixes / enclitics

There is a huge variety of grammatical formatives following the verb stem which were traditionally analyzed as suffixes (cf. Lipkind 1945; Susman 1943). These grammatical formatives express predominantly tense, mood, modality, and aspectual categories. However, a distributional analysis shows that most of them are better analyzed as enclitics or even independent grammatical words (cf. Helmbrecht 2006). The main problem with regard to

predicting stress placement is to decide whether a construction [verb plus enclitic] constitutes one or rather two stress domains. The second problem is that many of these grammatical enclitics do not behave homogeneously with regard to this distinction, i.e. in one context they are integrated prosodically, but in the next context they are not, constituting a new stress domain. In what follows, I will consider only examples of the first case, i.e. cases in which they are integrated into the prosodic structure of the verb.

The operation of the basic rules for the placement of primary and secondary stress can be illustrated with the complex stem+enclitics construction in (33). The verbal stem *hihokú* ‘to preach’ in (33) receives one derivational infix (RFL) and three suffixes/enclitics, namely a SBJ.3PL suffix, an iterative (ITER), and a quotative (QUOT) enclitic. Primary stress is placed on the RFL infix in accordance with the three mora rule. Secondary stress is distributed on every second mora thereafter. The first secondary stress is placed on the last vowel of the stem (this is an effect of the vowel strength hierarchy), and the second secondary stress on the iterative enclitic.

(33) *hihokíkùjines’àže*

hiho<kí>kù-ire-s’à-že

<RFL>preach-SBJ.3PL-ITER-QUOT

‘They were preaching this way, it is said.’ (WIL013)

In principle, all suffixes/enclitics may bear primary or secondary stress when they are integrated into the prosodic structure of the word. Some exceptions exist, however. All depends on the mora pattern of the word form (stem alone, or prefix-stem combination). An enclitic receives primary stress when the stem has less than three moras, or exactly three moras including a Dorsey’s Law-sequence. A typology of stable and unstable syllabic patterns is given in Table 12 on the next page.

When enclitics are integrated prosodically into the word form, then they may receive primary stress after stems or prefix-stem forms with an unstable pattern; cf. (34). In this example the verbal stem consists of a long vowel, i.e. two moras, thus primary stress falls on the first syllable of the future enclitic. Unstable patterns are stem types that are too short to attract the main stress by themselves so that subsequent suffixes/enclitics bear it. Stems with stable patterns are long enough to attract primary stress by themselves. If there are suffixes or enclitics, they may bear only secondary stresses.

- (34) *cuykjene*
cuy-kjene
 have.plenty-FUT
 ‘There will be plenty.’

When enclitics are integrated prosodically, then they will receive secondary stress after stems or prefix-stem forms with a stable pattern; cf. (35). In this example the noun stem has three moras with primary stress on the second syllable. Secondary stress will then fall on the second mora / syllable of the indefinite article.

- (35) *nijkajikižə*
nijkajik=hižə
 child.little=one
 ‘a / one little child’

Hocak has numerous mostly verbal enclitics and there is some variation in their behavior with regard to stress placement. Tables 13 on the following page and 14 on page 139 present a summary of the different enclitics and the rules they follow in primary and secondary stress placement (cf. Helmbrecht in prep.).

- (36) *kanak=irèe=nə*
 marry=SBJ.3PL=DECL
 ‘they marry’

As can be seen from (36), *=irèenə* restarts mora counting, otherwise the primary stress would be on /i/. The underlying form is *=ire-e-nə* with the secondary stress on the /e/, the lengthened vowel before *=nə* (DECL).

Table 12. Classification of stems / prefix-stems according to Susman (1943: 46–55).

Stem / Prefix+Stem patterns in terms of:	Syllables	Moras
Stable patterns	CVV'CV	3 moras
	CVCV'CV	3 moras
	CV[CVCV]'CV	4 moras – incl. a DL sequence
Unstable patterns	'CVV	2 moras
	CV'CV	2 moras
	CV[CV'CV]	3 moras – incl. a DL sequence

Table 13. Hocak enclitics and their different behavior in primary and secondary stress placement

Type	Forms	Stress rules
Mono-syllabic suffixes/enclitics	=ak/ =qk (=ak after V, =qk after C), =xji, -kje, -ke, -ge, -gi, =nq, =nqk, -ni, =ra, -re, -re (IMP), -s'a, že/še	(a) primary stress after stem with unstable pattern. (b) no stress after stem with stable pattern. (c) secondary stress after unstressed syllable (<i>waborášip-kè</i> 'you shot them down often') (d) primary stress after contraction with last syllable of a stem with unstable pattern (<i>hajáak</i> 'he is seeing')
Disyllabic enclitics	=eja, =ege, =(i)kje (epenthetic /i/ after C), =gaja, =kjane, =giži, =guni, =šunq, =šqnaq, =nagre, =žeži	(a) The first syllable is accented after a stem with unstable pattern — this may be a primary or secondary stress. (b) The second syllable is accented after a stem with a stable pattern — this is always a secondary stress.
Disyllabic enclitics	=ire, =jire, =gini, =isge, =išge	These disyllabic enclitics with two moras have (a) no effect on the stem and (b) are themselves lightly accented
Disyllabic enclitics	=žaaare, =jaané, =nqk'ú	These disyllabic enclitics with stable patterns, i. e. three moras have (a) no effect on the preceding stem/host, and (b) constitute their own stress domain
Trisyllabic enclitics	=nunige, =kjanahe	These trisyllabic words/enclitics have the primary stress on the first syllable and a secondary stress on the third syllable, if the preceding stem/host has an unstable pattern.

Table 14. Peculiarities of some specific enclitics

Enclitics	Function	Specific rules
<i>=nq</i>	DECL	<p><i>-nq</i> (DECL) appears after stem-final V which is lengthened and stressed, cf. <i>pjĩna</i>, <i>nĩgĩgĩĩna</i>. <i>-nq</i> (DECL) may also receive secondary stress, if the preceding syllable is unstressed. – combinations of <i>kjanqhe+nq</i>, <i>jee+nq</i>, <i>jire+nq</i> are like stem+<i>nq</i> compositions.</p> <p>– <i>-ire+nq</i> and <i>-nĩ+nq</i> throw light stress on <i>-nq</i>, if preceded by an unstressed syllable, cf. <i>'aĩrenq̃</i> ‘they say’ (cf. Susman 1943: 54).</p>
<i>=nq̃qk</i>	‘they are (in a sitting position)’	<p>– Bears its own stress and hence blocks the regular count of moras. The stress resembles the regular secondary stress. This holds also for the derived demonstrative pronouns <i>=nq̃qgre</i> ‘these’.</p> <p>– Grammatical forms following <i>=nq̃qk</i> are unstressed on the first syllable, and lightly stressed on the second syllable; cf. <i>hirá=nq̃qkĩĩ</i> ‘if they think’; <i>wa'ĩ=nq̃qkra</i> ‘what they do’ cf. Miner (1990); Susman (1943: 49, 54)</p>
<i>=nq̃qgre</i>	‘these’	<i>wa'ĩ=nq̃qgre</i> ‘the ones who do this’
<i>=jee / =jq̃q</i>	‘to be (in a standing position)’	– Bears its own stress and hence blocks the regular count of moras. This special stress resembles the regular secondary stress, cf. <i>nq̃qwá=jéenaq</i> ‘he is singing (standing)’.
<i>=ire</i> <i>=jire</i> <i>=anqga</i> <i>=žaaire</i>	SBJ.3PL ‘to start’ ‘and’ EMPH	– These enclitics block and restart the mora count for stress placement, cf. ex. (36) on page 137
<i>ğáak=anqgá</i> cry=and		<i>=anqgá</i> is a proper stress domain.

5.4 Compounding

Often, the placement of primary stress is the only means in Hocak to distinguish between juxtaposition and compounding. If two words are juxtaposed, they retain their own stress domains, i. e. each of them has its own primary stress. If two words, two nouns for instance, are compounded, then they constitute a single stress domain in which the third-mora-from-left rule is operative, as illustrated in the examples in (37) through (39). In all three cases, the first word / stem would have received primary stress if it would have been the first element in a juxtaposition construction.

- (37) *həqbókahi*

həp-hokahi

day-every

‘every day’

- (38) *wəgwácek jáane*

wək-wácek jáane

man-young this

‘this young man’ (WIL063)

- (39) *hinəgwácek waanínərə*

hinək-wácek wa-hanj-ire-ra

woman-young OBJ.3PL-own-SBJ.3PL-DEF

‘their young women’ (WIL050)

6 Conclusions

Hocak stress placement is a typological rarity, but perhaps in a more complex way as the typology of stress systems by Goedemanns and van der Hulst (2005a, 2005b) suggest. Their classification of languages as having either a fixed stress location stress system, or a weight sensitive stress system does not do justice to Hocak, where both principles play a role. Their classification of Hocak is not fully correct either. Primary stress falls on the third syllable in Hocak only under certain conditions. Hocak is mora counting – this holds for primary and secondary stress – and syllable accenting; i. e. once a syllable received stress, the second mora of this syllable does not count anymore.

We saw the type of problems that arise when grammatical or phonological systems of an individual language are described in terms of a rough typology.

Nuances, particular properties of the individual language are neglected or ignored, because the individual system does not entirely fit into the categories and category values of a typological description. From this perspective, each language is a *rarissimum*.

Abbreviations

- = morpheme boundary; _ = locus of personal inflection; < > = infix boundaries; 1, 2, 3 = first, second, and third person; A = actor; APPL.BEN = benefactive applicative; APPL.INESS = inessive applicative; APPL.INST = instrumental applicative; APPL.SUPRESS = superessive applicative; C = consonant; D = dual; DECL = declarative; DEF = definite article; E = exclusive; EMPH = emphatic; FUT = future; HAB = habitual; I = inclusive; IMP = imperative; INST.shooting = instrumental prefix “by shooting”; ISC = initial stem component; ITER = iterative; NEG = negation; OBJ = object; PL = plural; POSS.REFL = possessive reflexive; QUOT = quotative; RFL = reflexive; SBJ = subject; SG = singular; U = undergoer; V = vowel / verb

Notes

1. More information can be found on the project website http://www2.uni-erfurt.de/sprachwissenschaft/Vgl_SW/Hocank/home.html and on the DoBeS website <http://www.mpi.nl/DOBES/projects/hocank>
2. Alternative attempts to determine the primary stress placement on the basis of the metrical structure of words, as proposed in Hale and White Eagle (1980), are problematic in that they presuppose that the metrics in actual word forms are easy to identify. This is, however, not the case. There are other more theoretical problems with this approach, see Miner (1979, 1990) for details.
3. The two forms *i-* and *hi-* are allomorphs, *hi-* occurs word-initially only.
4. The initial stem component (ISC) is in most cases a lexicalized prefix. See Helmbrecht and Lehmann (2008) for further details.
5. For the formal and functional properties of enclitics in Hocąk, see Helmbrecht (2006).
6. The first plural inclusive undergoer pronominal affix is a discontinuous morpheme *wąqgá... wi* (PL).

References

- Dorsey, James Owen
 1883 On the Comparative Phonology of Four Siouan Languages. In *Smithsonian Institution Annual Report for 1883* (1885): 919–929.
- Goedemans, Rob and Harry van der Hulst
 2005a Fixed Stress Locations. In *The World Atlas of Language Structures*. Martin

- Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 62–65. Oxford: Oxford University Press.
- Goedemans, Rob and Harry van der Hulst
2005b Weight-Sensitive Stress. In *The World Atlas of Language Structures*. Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.), 66–69. Oxford: Oxford University Press.
- Goddard, Ives
1996 Introduction. In *Languages (Handbook of North American Indians Vol. 17.)* Ives Goddard (ed.), 1–17. Washington, D. C.: Smithsonian Institution.
- Gordon, Raymond G. Jr. (ed.)
2005 *Ethnologue: Languages of the World*. Fifteenth edition. Dallas: SIL International. <http://www.ethnologue.com/>
- Hale, Kenneth and White Eagle, Josephine
1980 A preliminary metrical account for Winnebago accent. *International Journal of American Linguistics* 46: 117–132.
- Hale, Kenneth
1985 A note on Winnebago metrical structure. *International Journal of American Linguistics* 51: 427–429.
- Helmbrecht, Johannes and Christian Lehmann
2008 Hočank's challenge to morphological theory. In *A World of Many Voices: Lessons from Documented Endangered Languages*. David Rood, David Harrison and Arienne Dwyer (eds.), 271–315. Amsterdam: Benjamins.
- Helmbrecht, Johannes and Christian Lehmann (eds.)
2006 *Learners' Dictionary Hocank-English / English-Hocank*. (= Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt; 21) Erfurt: University of Erfurt. <http://www2.uni-erfurt.de/sprachwissenschaft/ASSidUE/ASSidUE21.pdf>
- Helmbrecht, Johannes
2006 The problem of tense, mood and aspect enclitics in Hocank – how annotated corpora and concordancing tools can support descriptive linguistics. In *Contrastive Studies and Valency. Studies in Honor of Hans Ulrich Boas*. Petra Steiner, Hans Christian Boas and Stefan Schierholz (eds.), 171–199. Bern: Lang.
- Helmbrecht, Johannes
in prep. *A Hocak (Winnebago) Grammar*. University of Regensburg.
- Lipkind, William
1945 *Winnebago grammar*; New York; King's Crown Press.
- Matthews, Hubert G.
1958 *Handbook of Siouan Languages*. Doctoral dissertation, University of Pennsylvania.
- Miner, Kenneth L.
1979 Dorsey's law in Winnebago-Chiwere and Winnebago accent. *International Journal of American Linguistics* 45: 25–33.
- Miner, Kenneth L.
1981 Metrics, or Winnebago made harder. *International Journal of American Linguistics* 45: 25–33.

- Miner, Kenneth L.
1990 Winnebago accent: The rest of the data. *Anthropological Linguistics* 31: 148–172.
- Mithun, Marianne
1999 *The Languages of Native North America*. Cambridge: Cambridge University Press.
- Susman, Amelia
1943 *The accentual system of Winnebago*. Doctoral dissertation. Columbia University, New York.
- Wolff, Hans
1950 Comparative Siouan: Parts I–IV. *International Journal of American Linguistics* 16: 61–66, 113–121, 168–178, 197–204.

Affixation by place of articulation: the case of Tiene

Larry M. Hyman

1 Introduction

As is well known, certain phonological, morphological and syntactic properties occur with great frequency in the world's languages, while others are quite rare. To account for this difference, linguists have generally assumed that properties will be frequent if they have a "natural" motivation in terms of production, processing, communication etc., whereas properties which lack such a functional basis are "unnatural" and hence expected to be rare. Qualitative differences between frequent vs. rare properties may be attributed to diachrony: Widely attested synchronic properties tend to resemble each other in terms of substance, as they often arise transparently from universal processes of phonologization and grammaticalization. More restricted properties, on the other hand, often have an unmotivated or arbitrary character requiring a history involving multiple such processes and/or restructuring (cf. Bach and Harms' 1968 notion of "crazy rules"). What would be striking is a property which appears to be motivated, but still is rare.

In this paper I consider one such case from the verb morphology of Tiene, a Bantu language of the Teke subgroup spoken in the Democratic Republic of the Congo, which is not only rare, but as we shall also see, mysterious. In order, first, to appreciate how unique the Tiene facts to be presented are within Bantu, consider in (1) the more "canonical" derived verb structures from Ikalanga, a Southern Bantu language spoken in Botswana:

- (1) Derivational suffixes ("verb extensions") in Ikalanga
(Mathangwane 2001)
- a. Productive extensions (cf. *dabíl-a* 'answer')
- | | | | |
|-----------------------|-------------------|---------------------|------------|
| i. Causative: | <i>dabíl-ís-a</i> | 'cause to answer' | < PB *-IC- |
| ii. Applicative: | <i>dabíl-íl-a</i> | 'answer for / at' | < PB *-ID- |
| iii. Reciprocal: | <i>dabíl-án-a</i> | 'answer each other' | < PB *-an- |
| iv. Stative / Neuter: | <i>dabíl-ík-a</i> | 'be answerable' | < PB *-ik- |
| v. Passive: | <i>dabíl-w-a</i> | 'be answered' | < PB *-U- |
- (→ [*dabig-w-*])

- b. Non-productive extensions (**amb-*, **fum-* are bound roots)
- i. Extensive: *amb-al-a* ‘dress, put on (clothes)’ < PB **-ad-*
 - ii. Reversive tr.: *amb-ul-a* ‘take off (clothes)’ < PB **-ud-*
 - iii. Reversive intr.: *amb-uk-a* ‘come off (clothes)’ < PB **-uk-*
 - iv. Impositive: *fum-ik-a* ‘cover up (sth.)’ < PB **-uk-*
- c. Multiple extensions:
dabil-is-an-il-a ‘cause to answer each other for/at’

The productive derivational suffixes or “verb extensions” are illustrated in (1a), where the verb base is /*dabil-*/ ‘answer’ and *-a* is a final vowel (FV) inflectional morpheme. With the exception of the passive, these extensions all have the shape *-VC-*, corresponding with the Proto-Bantu (PB) reconstructions given to the right (Meeussen 1967; Schadeberg 2003). The data in (1b) illustrate some typically non-productive suffixes. Although there are constraints, the form in (1c) shows that extensions typically can be combined with each other to form quite long derived verb stems, here consisting of six syllables.

If we now compare the Tiene data in (2), we see a quite striking difference (the corresponding Ikalanga suffixes are shown in parentheses to the right):

- (2) Verb infixes in Tiene (Ellington 1977)
- a. i. Causative:
lók-a ‘vomit’ → *lósek-ε* ‘cause to vomit’ (cf. *-is*)
 - ii. Applicative:
yók-a ‘hear’ → *yólek-ε* ‘listen to’ (cf. *-il-*)
 - b. i. Extensive:
kab-a ‘divide’ → *kalab-a* ‘be divided’ (cf. *-al-*)
 - ii. Reversive tr.:
sook-ε ‘put in’ → *solek-ε* ‘take out’ (cf. *-ul-*)

As in Ikalanga, the causative and applicative suffixes are productive in Tiene, while the extensive and reversive are not. While Ikalanga (and Bantu in general) has *-VC-* verb suffixes, the (underlined) consonant of the corresponding suffix appears to be infixed in Tiene. (Other differences, e. g. in the vowels, will be addressed below.) The question is why? What has caused Tiene to change the pan-Bantu suffixes into infixes?

Such questions are addressed in the following sections. In order to answer the above questions I first take a fuller look at Tiene verb morphology in Section 2. As we shall see, the exact realization of the above and other verb extensions depend on the phonological properties of the verb base. In Section 3 I show that these properties follow from restrictions on the “prosodic stem” in Tiene, which is then compared with other Bantu languages in Section 4. Section 5 seeks diachronic and external evidence for the unusual Tiene properties, which are shown to have parallels in other Teke languages (Section 6) and a Nigerian Plateau language, Izere (Section 7). A speculative account of the rare and mysterious distributions and affixation by place of articulation is presented in Section 8 followed by a brief summary in Section 9.

2 Tiene verb extensions

In this section we shall examine the realization of the verb extensions in Tiene. Except where indicated, all of the data are taken from Ellington (1977) to whom I owe an additional debt for personal communications.

In (3) we begin by considering the two ways in which the language forms a causative:

(3) Causative formation (PB *-*IC-i*- > -*es*-)

a. C₂ = coronal (alveolar or palatal)

<i>mat-a</i>	‘go away’	<i>maas-a</i>	‘cause to go away’
<i>bót-a</i>	‘give birth’	<i>bóos-ε</i>	‘deliver (child)’
<i>tiit-a</i>	‘grow’	<i>tiis-ε</i>	‘cause to grow’
<i>kɔlɔ</i>	‘become tired’	<i>kɔɔsɔ</i>	‘tire (tr.)’
<i>pal-a</i>	‘arrive’	<i>paas-a</i>	‘cause to arrive’
<i>kal-a</i>	‘be’	<i>kaas-a</i>	‘cause to be / become’
<i>lil-a</i>	‘cry’	<i>liis-ε</i>	‘cause to cry’
			(< Guthrie 1960)
<i>taan-a</i>	‘get thin’	<i>taas-a</i>	‘cause to get thin’
			(< Guthrie 1960)
<i>píín-a</i>	‘be black’	<i>píís-ε</i>	‘blacken’
<i>bany-a</i>	‘be judged’	<i>baas-a</i>	‘caused to be judged’

b. C_2 = non-coronal (labial or velar)

i.	<i>lab-a</i>	'walk'	<i>lasab-a</i>	'cause to walk'
	<i>lók-a</i>	'vomit'	<i>lósek-ε</i>	'cause to vomit'
	<i>bik-a</i>	'become cured'	<i>bisek-ε</i>	'cure' ($<$ Guthrie 1953)
	<i>kuk-a</i>	'be sufficient'	<i>kusik-ε</i>	'make sufficient' ($<$ Guthrie 1960)
ii.	<i>dím-a</i>	'become extinguished'	<i>díseb-ε</i>	'extinguish (tr.)'
	<i>yóm-a</i>	'become dry'	<i>yóseb-ε</i>	'make dry'
	<i>tóm-a</i>	'send'	<i>tóseb-ε</i>	'cause to send'
	<i>suom-ɔ</i>	'borrow'	<i>sɔsɔb-ɔ</i>	'lend'

c. $C_2 = \emptyset$

	<i>le</i>	'eat'	<i>lees-ε</i>	'feed'
	<i>vu</i>	'fall'	<i>vuus-ε</i>	'cause to fall'

As seen in (3a,b), most roots in Tiene, as in Bantu in general, consist of a $C_1V(V)C_2$ - structure, where VV represents a long vowel. As can now also be seen, the realization of the causative depends on the place of articulation of the root C_2 . In (3a), where C_2 = coronal, the /s/ of the causative suffix replaces the C_2 , and the vowel lengthens, if it is not already long. (In these and other forms, the final vowel of a derived verb stem is /-ε/, which assimilates to a preceding /ɔ/ or /a/.) In (3b-i), where C_2 = non-coronal, the /s/ of the causative is infixes, with the root C_2 now being realized as C_3 . The same is observed in (3b-ii), with the addition that root C_2 /m/ is denasalized to [b]. The two verbs in (3c) show that when there is no root C_2 , the causative suffix takes the expected -Vs- shape.

Turning to the applicative suffix, the data in (4) show the same skewing on the basis of the place of articulation of the root C_2 :

(4) Applicative formation (PB *-Id- > -el-)

a. C_2 = coronal (alveolar or palatal)

	<i>bót-a</i>	'give birth'	<i>bóot-ε</i>	'give birth for'
	<i>bel-a</i>	'speak'	<i>beel-ε</i>	'speak to'
	<i>sal-a</i>	'work'	<i>saal-a</i>	'work for' ($<$ Guthrie 1953)
	<i>yal-a</i>	'spread'	<i>yaal-a</i>	'spread for'
	<i>kas-a</i>	'fight for'	<i>kaas-a</i>	'fight on behalf of'

<i>kón-a</i>	‘plant’	<i>kóon-a</i>	‘plant for’
<i>són-ɔ</i>	‘write’	<i>sóon-ɔ</i>	‘write for’
<i>kony-a</i>	‘nibble’	<i>koony-ε</i>	‘nibble for’
b. C ₂ = non-coronal (labial or velar)			
i. <i>yɔb-ɔ</i>	‘bathe’	<i>yɔɔb-ɔ</i>	‘bathe for’
<i>bák-a</i>	‘reach’	<i>bálak-a</i>	‘reach for’
<i>yók-a</i>	‘hear’	<i>yólek-ε</i>	‘listen to’
ii. <i>dum-a</i>	‘run fast’	<i>dunem-ε</i>	‘run fast for’
			(NB: <i>l</i> → [n])
<i>súom-ɔ</i>	‘buy’	<i>sónem-ε</i>	‘buy for’
<i>tim-a</i>	‘dig’	<i>tinem-ε</i>	‘dig for’
			(< Guthrie 1953)
<i>lɔŋ-ɔ</i>	‘load’	<i>lɔnɔŋ-ɔ</i>	‘load for’
c. C ₂ = ∅			
<i>tá</i>	‘throw, strike’	<i>téel-ε</i>	‘throw to / for’
<i>díá</i>	‘wrap’	<i>díil-ε</i>	‘wrap for’
<i>síε</i>	‘whittle’	<i>síil-ε</i>	‘whittle for’

In (4a), where the root C₂ is coronal, the /l/ of the applicative suffix /-el-/ is lost, the only reflex of the applicative being the observed vowel lengthening.¹ In (4b-i), the /l/ is infixated exactly like the /s/ of the causative in (3b-i). The same infixation is observed in (3b-ii), but since the root C₂ is nasal, the /l/ is nasalized to [ɲ]. The three verbs in (4c) which lack a C₂ form their applicative by suffixing -Vl-.

Tiene stative formation is shown in (5).

(5) Stative formation (PB *-lk- > -ek-)

a. C ₂ = coronal (alveolar or palatal)			
i. <i>yaat-a</i>	‘split’	<i>yatak-a</i>	‘be split’
<i>ból-a</i>	‘break’	<i>bólek-ε</i>	‘be broken’
<i>faas-a</i>	‘drive through’	<i>fasak-a</i>	‘be driven through’
ii. <i>són-ɔ</i>	‘write’	<i>sónɔŋ-ɔ</i>	‘be written’
<i>vwuny-a</i>	‘mix’	<i>vwunyey-ε</i>	‘be mixed’
b. C ₂ = non-coronal (labial or velar) (? PB *-ad- > -al-)			
i. <i>kab-a</i>	‘divide’	<i>kalab-a</i>	‘be divided’
<i>nyak-a</i>	‘tear’	<i>nyalak-a</i>	‘be torn’
ii. <i>kam-a</i>	‘twist’	<i>kanam-a</i>	‘be turned over’

c. $C_2 = \emptyset$

<i>kaa</i>	‘fasten’	<i>kaal-a</i>	‘be fastened’
------------	----------	---------------	---------------

The forms in (5a-i) show that when the root C_2 is coronal the stative is derived by the historical *-Vk-* suffix, which becomes *-Vŋ-* in (5a-ii), where the root C_2 is nasal. However, when the root C_2 is non-coronal, as in (5b-i), an *-l-* is infix, which again becomes *-n-* when the root C_2 is nasal, as in (5b-ii). In this case it appears necessary to recognize two allomorphs: a suffix *-Vk-* which occurs after coronals vs. an infix with *-l-* which occurs before non-coronals (and is, therefore, identical to the applicative). The one example Ellington cites without a root C_2 takes the *-l-* allomorph, as seen in (5c).

There are fewer examples of reversives in the materials, all of which are reproduced in (6).

(6) Reversive formation in Tiene (PB ‘reversive’ **-ʊk-*, **-ʊd-* > *-ok-*, *-ol-*)a. C_2 = coronal (alveolar or palatal)(PB ‘reversive transitive’ **-ʊd-* > *-ol-*)

<i>kót-a</i>	‘tie’	<i>kóót-ε</i>	‘untie’
<i>yal-a</i>	‘spread’	<i>yaal-a</i>	‘roll up’

b. C_2 = non-coronal (labial or velar)(PB ‘reversive transitive’ **-ʊd-* > *-ol-*)

<i>sook-ε</i>	‘put in’	<i>solek-ε</i>	‘take out’
<i>sum-a</i>	‘stick in ground’	<i>sunem-ε</i>	‘pull out of ground’

c. C_2 = coronal (alveolar or palatal)(PB ‘reversive intransitive’ **-ʊk-* > *-ok-*)

<i>kót-a</i>	‘tie’	<i>kótek-ε</i>	‘be untied’
--------------	-------	----------------	-------------

d. Coronal / non-coronal alternation

<i>vuol-a</i>	‘open’	<i>vuok-a</i>	‘close’
---------------	--------	---------------	---------

(< impositive **-ɪk-*?)

(6a,b) show the same realizations as the applicative: vowel lengthening if the root C_2 is coronal vs. *-l-* infixation if it is non-coronal. The one form in (6c) shows a *-Vk-* suffix being added to a coronal root C_2 (as in the stative), while (6d) shows an alternation between /l/ and /k/, where the latter is perhaps related to the PB impositive suffix **-ɪk-* illustrated in Ikalanga in (1b-iv).

While we shall be primarily interested in why affixal *-s-* or *-l-* appears before non-coronal root C_2 , let us first note the types of phonological analyses

one might consider to account for the above alternations. Working within the framework of Chomsky and Halle (1968), Ellington (1977) proposes a solution involving :

- (7) Metathesis: CVPVT, CVKVT → CVTVP, CVTVK

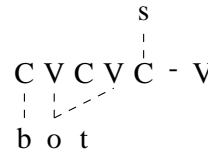
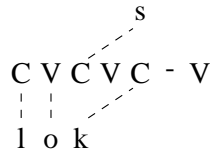
a.	b.	c.	d.	
/lók-es-/	/yók-el-/	/kab-el-/	/sook-el-/	Underlying representations
			sok-el-	Vowel shortening
lós-ek-	yól-ek-	kal-eb-	sol-ek-	Metathesis

Whenever there is an input where root C_2 = labial or velar and suffixal C_3 = coronal, the two consonants are metathesized. In addition, (7d) shows that a long root vowel must be shortened.

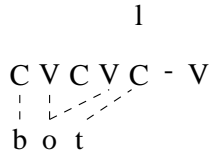
Had Ellington been working a few years later, Tiene metathesis or infixation could have been neatly represented in the non-linear framework of McCarthy (1981). With the CV template in (8a), the effect of metathesis could be achieved by having the suffixal /s/ of the causative associate to the second C just in case the root ends in a non-coronal:

- (8) Multiple tier analysis with CV templates à la McCarthy (1981)

- a. /lók-/ + /-es-/ → lósek- b. /bót + es/ → bóos-



- c. /bót + el/ → bóot-



When the root C_2 is coronal, as in (8b), it is overridden by suffixal /s/, which links to the second C, while the coronal root C_2 overrides suffixal /l/ in (8c).

Finally, as summarized in (9), a constraint-based analysis within the framework of McCarthy and Prince (1999) is also possible (cf. Hyman and Inkelas 1997):

- (9) Metathesis / fusion can also be driven by OT-style output constraints
- a. $T \supset P, K$
 - b. $\text{MAX}(s) \gg \text{MAX}(\text{Root}) \gg \text{MAX}(l)$

(9a) directly encodes that coronals should precede non-coronals, while the ranked constraints in (9b) say that input /s/ should be preserved over other consonants of the root, which in turn should be preserved over affixal /l/.

The three approaches in (7)–(9a) of course assume that the realization of affixal /s/ and /l/ as the C_2 of CVCVC-V verb stems should be accounted for in phonological terms. The alternative would be to set up allomorphs (as is required for the stative, in any case): The causative and applicative would have the infixal allomorphs /-se-/ and /-le-/ when the root C_2 is non-coronal, otherwise the allomorphs would be suffixal /-es-/ and /-el-/, which fuse with a coronal C_2 but are transparently suffixed to a CV- root (see Section 3 concerning the realization of the underlying vowel /e/).

Considering the above (and perhaps other) solutions, we can conclude that the Tiene data are easy to describe, but raise a number of questions: Why doesn't Tiene have only derivational suffixes as elsewhere in Bantu? Why does the derivational morphology show such sensitivity to place of articulation ("fusion" vs. "infixation"; -Vk- vs. -lV- allomorphy)? Why do affixal -l- and -k- nasalize to [ɲ] and [ŋ], respectively, and the root C_2 /m/ denasalize to [b] (e. g. *tóm-a* 'send' → *tóseb-ε* 'cause to send')? Why does the vowel of a CVVC- root shorten when appearing in a trisyllabic verb stem (e. g. *yaata* 'split' → *yataka* 'be split')? As I will argue in Section 3, all of these properties can be explained as the result of phonological restrictions on the "prosodic stem" in Tiene.

3 The prosodic stem in Tiene

As summarized in (10), there are severe restrictions on the size and distribution of consonants and vowels within the stem in Tiene:

- (10) The "prosodic stem" in Tiene
- a. Five shapes: CV, CVV, CVCV, CVVCV, CVCVCV

- b. In case of $C_1VC_2V_2C_3V_3$:
- C_2 must be coronal
 - C_3 must be non-coronal
 - C_2 and C_3 must agree in nasality
 - V_2 is predictable (with few exceptions)

As seen in (10a), in Tiene stems can be monosyllabic, bisyllabic or trisyllabic, with vowel length being limited to monosyllables and the first syllable of bisyllabic stems. In addition, trisyllabic stems are subject to the conditions indicated in (10b). We have mostly focused on the first two, which have to do with place of articulation. In addition, C_2 and C_3 must agree in nasality: This causes infixal *-l-* and suffixal *-k-* to nasalize to [ɲ] and [ŋ] when the root C_2 is nasal.² When the infix is *-s-*, root C_2 /m/ denasalizes (since it is difficult, perhaps impossible to nasalize an [s]). With few exceptions, the V_2 is limited to /e/, which is realized [e] after /i, u, e, o/, but is identical to a preceding /ɛ, ɔ, a/.³ These phonological conditions define what may be referred to as the “prosodic stem” in Tiene.

Further evidence that the prosodic stem is subject to a maximum of three syllables is seen from the definitive aspect forms in (11).

- (11) Definitive aspect formation (cf. Proto-Bantu **-idɪd-* > *-elɛl-* ‘completive’)

a.	<i>kaa</i>	‘fasten’	<i>kalal-a</i>	‘fasten permanently’
	<i>nɔɔ</i>	‘look at’	<i>nɔlɔl-ɔ</i>	‘fix gaze on’
	<i>bɛɛ</i>	‘become ripe’	<i>bɛlɛl-ɛ</i>	‘ripen once and for all’
	<i>sía</i>	‘hate’	<i>sílel-ɛ</i>	‘hate definitively’
	<i>twa</i>	‘crush’	<i>túlel-ɛ</i>	‘crush definitively’
	<i>fue</i>	‘become violent’	<i>fuɛlɛl-ɛ</i>	‘become permanently violent’
	<i>suɔ</i>	‘show’	<i>suɔlɔl-ɔ</i>	‘show once and for all’
b.	<i>yɔb-ɔ</i>	‘bathe’	<i>yɔbɔb-ɔ</i>	‘bathe thoroughly’
	<i>mat-a</i>	‘go away’	<i>matat-a</i>	‘go away once and for all’
	<i>yak-a</i>	‘believe’	<i>yakak-a</i>	‘believe once and for all’
	<i>kén-a</i>	‘dance’	<i>kénen-a</i>	‘dance once and for all’
	<i>lɔŋ-ɔ</i>	‘load’	<i>lɔŋɔŋ-ɔ</i>	‘load once and for all’
c.	<i>kóóm-ɔ</i>	‘sweep’	<i>kómmɔm-ɔ</i>	‘sweep once and for all’
	<i>maas-a</i>	‘cause to go away’	<i>masas-a</i>	‘cause to go away for good’

The forms on the right show that the definitive aspect is characterized by a $C_1VC_2VC_3-V$ template, where $C_2 = C_3$. If a verb root lacks a C_2 , as in (11a), the C_2 and C_3 of the template are filled with a $/-lel-/$ sequence, whose $/e/$ assimilates to the root vowel as expected. As seen in (11b), if the root has a C_2 , the definitive is formed by adding $/e/$ + a copy of the C_2 . (11c) shows that if the root vowel is long, it must shorten, as expected. Since the C_2 and C_3 are identical, these forms are obviously exceptions to the distribution of coronal + non-coronal indicated in (10b). The requirement of the definitive that $C_2 = C_3$ thus overrides the otherwise general place restrictions on the two consonant positions. Now consider what happens if the verb base already has a C_3 :

“... verbs having the canonical shape -CVCVC- (including extended radicals) ... do not accept the Definitive Aspect Morpheme. For such verbs, this aspect must be rendered by adding the expression *nkó móte* to the conjugated verb in the Neutral Aspect” (Ellington 1977: 93).

The morphological definitive is blocked just in case the condition $C_2 = C_3$ cannot be met without either truncating part of the base or exceeding the maximum trisyllabic size constraint on stems.

Further evidence for a trisyllabic maximum is seen from the vestiges of the reciprocal extension in (12).

- (12) Reciprocal “vestiges” (PB $*-a(n)g-an-$ > $-ɲena$ > $-neɲa$)
- | | | | | | |
|----|------------|-----------------|----|----------------|-----------------------|
| a. | <i>le</i> | ‘eat’ | b. | <i>lé-neɲa</i> | ‘eat with each other’ |
| | <i>nwa</i> | ‘drink’ | | <i>nú-neɲa</i> | ‘drink each other’ |
| | <i>pa</i> | ‘give’ | | <i>pé-neɲa</i> | ‘give each other’ |
| | <i>ta</i> | ‘throw, strike’ | | <i>té-neɲa</i> | ‘injure each other’ |

As seen, the above four C(V)- roots occur with traces of the reciprocal extension $-neɲ-$ inherited from the Proto-Bantu plural + reciprocal sequence $*-a(n)g-an-$ found in a number of daughter languages (cf. Haya $-angan-$, Luganda $-agan-$). In the Tiene reflex, the velar + coronal sequence is metathesized to coronal + velar, in conformity with the place restrictions on prosodic stems. Significantly, there are no vestiges of the reciprocal with CVC- or CVCVC- verb bases, precisely because $-neɲ-$ would require a fourth syllable. It is again clear that derived stems are maximally trisyllabic in Tiene.

The same is true of the non-derived verb stems I have been able to extract from Ellington (1977). As seen in (13a), verbs consisting of a synchronically non-derived CVCVC- base + FV observe all of the constraints in (10):⁴

(13) “Non-derived” CVCVCV stems in Ellington (1977)

a.	<i>kótob-a</i>	‘chase’	C-t-b	
	<i>kótok-a</i>	‘gnaw’	C-t-k-	GCB * <i>kókot-</i>
	<i>vútek-ε</i>	‘come back’	C-t-k-	GCB * <i>bútok-</i>
	<i>pálab-a</i>	‘sprout’	C-l-b-	
	<i>pεlεb-ε</i>	‘fly’	C-l-b-	
	<i>tóleb-ε</i>	‘pierce’	C-l-b-	GCB * <i>tóbod-</i>
	<i>sélek-ε</i>	‘tease’	C-l-k-	GCB * <i>cék-ed-</i>
	<i>sɔlɔk-ɔ</i>	‘go out’	C-l-k-	GCB * <i>cɔkod-</i>
				‘pull out’
	<i>binem-a</i>	‘sleep’	C-n-m-	
	<i>dínem-a</i>	‘get lost’	C-n-m-	GCB * <i>dímed-</i>
	<i>kanam-a</i>	‘turn over (sth.)’	C-n-m-	GCB * <i>kámod-</i>
				‘wring’
	<i>kɔnɔm-ɔ</i>	‘crawl’	C-n-m-	
	<i>panam-a</i>	‘frighten’	C-n-m-	
b.	<i>m-pítiba</i>	‘darkness’, 9?	C-t-b-	“regular”
	<i>le-bóboki</i>	‘bird (sp.)’, 11 / 10	(PL. <i>m-bómboki</i>)	reduplication
	<i>ke-lélébe</i>	‘lip’, 7 / 8	(PL. <i>be-lélébe</i>)	reduplication
	<i>le-sásálá</i>	‘eyelash’, 11 / 10	(PL. <i>n-sánsálá</i>)	reduplication
	<i>m-fúmfálá</i>	‘armpit’, 9		reduplication
	<i>síkule</i>	‘school’		borrowing
	<i>n-gwánkete</i>	‘enemy’, 9	(~ <i>n-gbánkete</i>)	compound?

The schemas given to the right of these forms show that C_2 = coronal and C_3 = non-coronal, as expected. Several of these are reflexes of reconstructed Proto-Bantu forms. Ellington cites from Guthrie’s (1967–1971) Common Bantu, which I abbreviate as *GCB*. As seen, all but **bútok-* ‘come back’ require the metathesis of the GCB C_2 and C_3 consonants. Most interesting is **kókot-* ‘gnaw’, which, although possibly an archaic reduplication **kó-kot-*, becomes *kótok-a* in Tiene. Since such forms cannot be interpreted as infixation, it is clear that metathesis occurred as historical process in the language.

Except for the first form, the trisyllabic noun stems in (13b) all violate one or another of the conditions in (10). It looks, however, like four are reduplications, one a borrowing, and one a possible compound. (The number(s) following each gloss indicate(s) the Bantu noun SG./PL. noun class.) It is safe to say that there are very few trisyllabic noun stems.

Up to this point nothing has been said about possible constraints on the C₂ of bisyllabic verb stems. As seen in Table 1, both coronals and non-coronals appear in this position:

Table 1. Underlying consonants found in C₁ position and in C₂ position of C₁V(V)C₂V stems

	Labials	Alveolars	Proto-Palatals	Velars	
C ₁ (12):	p b m	t l n	s z ɲ	k g ŋ	(PB *c, *j > s, z)
C ₂ (9):	b m	t l n	s ɲ	k ŋ	

The main restriction is the voicing neutralization of /p, b/, /s, z/ and /k, g/ in C₂ position.⁵ In addition, there are exactly two constraints on C₁ + C₂ combinations: First, after C₁ /p/, we find [p] instead of [b], e. g. *kab-a* ‘divide’, *lób-ɔ* ‘fish with a line’ vs. *píp-a* ‘suck’, *m-pɛɛɛ* ‘wind’. Second, when C₁ is a nasal consonant, C₂ cannot be [b] or [l]. There is one interesting trisyllabic exception to this generalization: *nyalak-a* ‘be torn’. Deriving from earlier **nyakal-a*, the [l] escapes nasalization because it originates in C₃ position separated from the [+nasal] C₁ by a full syllable; cf. *nyak-a* ‘tear’. This form not only provides further evidence for metathesis, but also reveals that if we start with the pre-metathesis order, nasal agreement is both left-to-right and strictly local. Differing from nasal agreement between C₂ and C₃, a C₂ [k] may follow a C₁ nasal, e. g. *nók-ɔ* ‘to rain’. On the other hand, just as we saw that a (metathesized) C₂ [s] may not co-occur with a C₃ nasal, Ellington (1977) cites no forms where [s] follows a C₁ nasal.

4 Comparison with the prosodic stem in other Bantu languages

The Tiene facts presented in the preceding sections raise a number of questions: Why does Tiene have such a restrictive prosodic stem template? How did the prosodic stem template arise? Is Tiene unique, or can we relate these facts to what happens in other languages? In this section we compare Tiene with other Northwest (NW) Bantu languages spoken in the same vicinity.

The properties of the Tiene prosodic stem template can be subdivided as in (14a,b).

- (14) The properties of the Tiene prosodic stem template can be subdivided into those which are
- a. not so unusual
 - i. prosodic maximality
 - ii. decreasing # of oppositions from C_1 to C_2 to C_3
 - iii. stronger realization of consonants in C_1 vs. C_2 and C_3 positions
 - b. very unusual
 - i. coronal-noncoronal limitation on C_2 - C_3 sequences
 - ii. regular, nonlocal metathesis based on place of articulation

The properties in (14a) are also attested in neighboring languages. As seen in (15), whereas “canonical” Central, Eastern, and Southern Bantu languages do not place an upper limit on how many syllables can appear within the stem (recall Ikalanga from (1c) on page 146), stem-maximality conditions are very common in Northwest Bantu, as indicated in (15).

- (15) Maximal size limitations on the stem in certain Northwest Bantu languages and further West
- | | |
|-----------------------------------|--|
| a. four (~five) syllable maximum | Yaka (Hyman 1998),
Bobangi (Whitehead 1899)
Punu (Fontaney 1980, Blanchon 1995) |
| b. three (~four) syllable maximum | Koyo (Hyman 2008) |
| c. three-syllable maximum | Tiene (Ellington 1977),
Basaa (Lemb and Degastines 1973,
Hyman 2003),
Kukuya (Paulian 1975) |
| d. two (~three) syllable maximum | Mankon [Grassfields Bantu]
(Leroy 1982) |
| e. one (~two) syllable maximum | Ewe [Kwa] (Westermann 1930) |

Some of the languages are transitional, with the additional syllable (indicated in parentheses) typically being restricted to a single morpheme or construction.

As an illustration, consider in (16) the possible stem shapes in Koyo, a Northwest Bantu language spoken in nearby Congo (Hyman 2008):

(16) Possible stem shapes in Koyo (Congo-Brazzaville)

CV	<i>dz-a</i>	‘be, exist’	<i>my-a</i>	‘swallow’
CVV	<i>dzá-a</i>	‘eat’	<i>sá-a</i>	‘cultivate’
CVCV	<i>kór-a</i>	‘attach’	<i>bom-a</i>	‘kill’
CVCVCV	<i>sélum-a</i>	‘slip’	<i>ñɔbir-a</i>	‘tickle’
CVCVCVgV	<i>sélum-ag-a</i>	‘slip + DUR’	<i>ñɔbir-ag-a</i>	‘tickle + DUR’

In this language stems are maximally trisyllabic, although a fourth syllable is made possible only by the durative *-Vg-* suffix. That such a maximality condition is in effect is seen in (17).

(17) Maximum trisyllabic stem: verb extensions can be added only if there is room!

- a. *kór-a* ‘to tie’ *bar-a* ‘to bite’
kór-is-a ‘to cause to tie’ *bar-is-a* ‘to cause to bite’
kór-in-a ‘to tie each other’ *bar-in-a* ‘to bite each other’
- b. **kór-is-in-a* ‘to cause each other to tie’
**kór-in-is-a* ‘to cause to tie each other’
**bar-is-in-a* ‘to cause each other to bite’
**bar-in-is-a* ‘to cause to bite each other’
- c. *dzá-a* ‘to eat’ < /*dzé-a*/
dzé-s-a ‘to cause to eat, feed’
dzé-n-a ‘to eat each other’
dzé-s-in-a ‘to feed each other’
tá-a ‘to see’
tá-s-a ‘to cause to see, show’
tá-n-a ‘to see each other’
tá-s-an-a ‘to show each other’
- d. *yigin-a* ‘to get accustomed to’
yig-is-a ‘to cause to be accustomed’
súndzin-a ‘to decrease, shorten’
súndz-is-a ‘to cause to decrease’

Causative *-is-* and reciprocal *-in-* are illustrated in (17a). As seen in (17b), they cannot both be present after a CVC- root, because that would result in four syllables. They do co-occur in the order causative + reciprocal after a CV- root in (17c), where they can be fit into the trisyllabic maximal template. The examples in (17d) show exceptional cases where the causative replaces the [in] of a monomorphemic CVCin- verb.

Another property of Northwest Bantu accompanying stem maximality is the decrease in the number of consonant oppositions as one goes from left-to-right. In Table 2 we see that the drop is precipitous in Koyo:

Table 2. Decrease in the number of consonant oppositions in each of the four stem syllables

	Labials	Alveolars	Proto-Palatals	Velars
C ₁ (18)	p b w m mb	t l s n nd	ts dz y ɲ ndz	k h ŋg
C ₂ (12)	b m mb	r l s n nd	y ɲ ndz	g
C ₃ (6)	m	r l s n		g
C ₄ (1)				g

This contrasts with Proto-Bantu and present-day canonical Bantu languages, where there is a near-free distribution of consonants in the different stem positions (cf. Hyman 2008; Teil-D’Autrey 2002, 2008).⁶

In addition to the decrease in contrasts by position, Table 2 also shows that a stem-internal stop must be voiced. As a result, /p/ contrasts with /b/ only in C₁ position (as in Tiene). The examples in (18) show that /t/ and /k/ are realized as [t, k] in C₁ position but as [r, g] in C₂, C₃ and C₄ positions:

(18) Realization of /t/ and /k/ in Koyo

C ₁	/tón-a/	[tóna]	‘refuse’	/kúl-a/	[kúla]	‘abandon’
C ₂	/bát-a/	[bára]	‘keep’	/mék-a/	[méga]	‘dare’
C ₃	/tsikit-a/	[tsigira]	‘tremble’	/pítak-a/	[píraga]	‘smear’
C ₄				/pítak-ak-a/	[píragaga]	(+DUR)

As seen, however, there are no significant place restrictions in Koyo, aside from the restriction of /ts/ and /dz/ to C₁ position. In fact, place restrictions seem to be found only in the Teke subgroup to which Tiene belongs (see Section 6).

The question is whether the place restrictions in Tiene are synchronically grounded, i. e. due to a linguistic tendency of some sort, or are synchronically accidental, i. e. attributable to specific historical factors which gave rise to it? Addressing this latter possibility first, it is easy to show that T ⊃ P, K is not inherited from Proto-Bantu, nor is it typical of Bantu languages in general, where non-coronals statistically *precede* coronals. Consider in this context the Proto-Bantu extension system, which is impressively uniform throughout most of Bantu:⁷

(19) Proto-Bantu verb extensions (Meeussen 1967; Schadeberg 2003)

- a. frozen, mostly unidentifiable -VC- “expansions”
(9 / 11 = non-coronal)
 - i. **-im-*, **-un-*, **-ing-*
 - ii. **-ang-*, **-ab-*, **-ag-*, **-ak-*
 - iii. **-im-*, **-om-*, **-ong-* (only after CV-)
 - iv. **-ut-*
- b. unproductive extensions often restricted to post-radical position
(4 / 7 = non-coronal)
 - i. **-ik-* ‘impositive’
 - ii. **-am-* ‘positional’ (contactive)
 - iii. **-a(n)g-* ‘repetitive’ (tr. / itr.)
 - iv. **-ad-* ‘extensive’
 - v. **-at-* ‘tentive’
 - vi. **-ud-* / **-uk-* ‘reversive / separative’
- c. productive extensions (3 / 4 clear cases = coronal)
 - i. **-i-* ‘causative’
 - ii. **-iC-i-* ‘causative’
 - iii. **-iD-* ‘applicative’
 - iv. **-Ik-* ‘neuter / stative’
 - v. **-an-* ‘reciprocal / associative’
 - vi. **(-iC)-u-* ‘passive’

In (19) the reconstructions are presented in the order in which they are expected to occur: frozen “expansions”, unproductive extensions, productive extensions. As seen, as one moves out from the verb root the suffixal consonants become more coronal. The preponderance of (productive) coronal suffixes appears to be characteristic of Niger-Congo in general, and is particularly striking in the Atlantic subgroup (cf. the table in Becher 2000: 31). Although less impressive in this respect, for completeness, the Proto-Bantu inflectional endings are given in (20).

- (20) Proto-Bantu final inflectional endings (Meeussen 1967; see also Grégoire 1979; Bastin 1983)
- | | | | |
|-----------|---------------|-----------|----------------|
| a. *-I | (past) | d. *-ag-a | (imperfective) |
| b. *-ε | (subjunctive) | e. *-a | (“default”) |
| c. *-id-ε | (perfective) | | |

Since Proto-Bantu did not provide the source, the Tiene place restrictions must be an innovation. We might therefore expect nearby Northwest Bantu languages to show at least a statistical tendency for coronals to come earlier among post-root consonants. In Table 3 we see that both Koyo and Bobangi instead show a preponderance of coronals in later verb stem positions (more than expected frequencies are in bold).⁸ In both Koyo and Bobangi coronals range from 43.8% to 55.3% in the two root positions, C₁ and C₂. The percentage rises dramatically to 70.5–79.2% in C₃ and C₄ positions. The last two columns show that the C₂ in Koyo and the C₃ in Bobangi are even more likely to be coronal if they are the last consonant of the verb stem. Since coronals cannot occur in C₃ position, Tiene appears to have exactly the opposite distribution from its neighbors outside the Teke group.

Table 3. Distribution of consonants by place in two nearby Northwest Bantu languages.

a. Koyo (Hyman 1996): Max = CVCVCV (fourth CV must be durative -gV).								
	Total	P	K	T	Y	{T, Y} %		
C ₁	1,536	419 (+9%)	316 (-21%)	613 (-8.1%)	118 (+47.2%)	47.6%		
C ₂	1,308	324 (-0.3%)	409 (+20.4%)	549 (-2.8%)	24 (-156.7%)	43.8%	CVC ₂ V 47.4%	CVC ₂ VCV 38.5%
C ₃	192	11 (-333.5%)	31 (-54.2%)	148 (+44%)	1 (-804.3%)	77.6%		
b. Bobangi (Whitehead 1899): Max = CVCVCVCV (3/3.324 verbs have fifth CV)								
	Total	P	K	T	Y	{T, Y} %		
C ₁	7,619	2,055 (+20.0%)	1,502 (-18.9%)	3,048 (-19.2%)	1,012 (+45.3%)	55.3%		
C ₂	7,246	1,508 (-3.7%)	2,289 (+25.8%)	3,257 (-6.1%)	192 (-174.3%)	47.6%		
C ₃	2,930	395 (-60.1%)	470 (-46.2%)	1,955 (+28.5%)	110 (-93.6%)	70.5%	CVCVC ₃ V 81.5%	CVCVC ₃ VCV 43.6%
C ₄	852	65 (-182.8%)	112 (-78.4%)	634 (+35.9%)	41 (-51.0%)	79.2%		

5 In search of an explanation

Having eliminated the possibility that Tiene inherited a skewed distribution of consonants by place, we must now seek an account of how and why Tiene innovated as it did. There are two aspects of the Tiene situation which require an explanation. First, why cannot $C_2 = C_3$ in place of articulation? Second, why cannot C_2 - C_3 = non-coronal + coronal?

The first question seems appropriately answered by relating Tiene to the avoidance of sequences of homorganic consonants in lexical items, well-known from the study of Arabic roots (cf. Frisch et al. 2004). Pozdniakov and Segerer (2007) have recently shown that there is, universally, an avoidance of successive homorganic consonants within words, even when such consonants are separated by a vowel. They calculate the expected vs. attested number of consonant combinations by place of articulation. Some of their results are reproduced in (21).

(21) Similar Place Avoidance (SPA) (Pozdniakov and Segerer 2007)

Fula (n = 672)

	P	K	T	Y
P	--	--	+	
K	--	--		++
T	++	++	-	-
Y	+	++		--

Malagasy (n = 1,944)

	P	K	T	Y
P	--		+	+
K		--		+
T	++		-	-
Y		++	-	--

Basque, Euskara (n = 3,140)

	P	K	T	Y
P	--			
K	--	--	+	
T	+	++	-	+
Y	++			--

Pidgin English,

Port-Moresby (n = 2,215)

	P	K	T	Y
P	--		+	
K		--		
T	++	+	-	
Y		+		

Quechua (n = 5,254)

	P	K	T	Y
P	--	-	++	
K		--		
T	+	++		-
Y	+	++	-	

Classical Mongolian

(n = 66,407)

	P	K	T	Y
P	--	-	+	++
K		--	++	
T		++	--	
Y	+			-

Bantu Reconstructions (n = 12,426)					Swahili (n = 1,481)				
	P	K	T	Y		P	K	T	Y
P	–		+		P	–	–	+	
K		–	–	+	K		–	–	++
T	+	+		–	T	+		–	–
Y	+	+	–	–	Y	+		–	–

In the above tables P = labial, K = velar, T = dental-alveolar, and Y = palatal. A single + or – indicates that the attested number of lexical entries is 15–30% off from the number that would be expected if there were no co-occurrence restrictions. A double ++ or –– means that the discrepancy is over 30%. As seen, the bulk of minuses occurs along the homorganic cells which descend diagonally from left to right, i. e. where the consonant combinations involve an identical place of articulation. Besides the underrepresentation of consonants of identical place, the upper left and lower right quadrants of the above tables show a statistical avoidance of successive consonants from the same superclass: P+K (= peripheral / grave) and T+Y (= medial / acute).

Pozdiakov and Segerer (2007) show that their statistical universal is (with the exception of Y) also in force within Bantu. While we expect the productive suffixes to combine freely with bases ending in all places of articulation, I have found that unproductive suffixes avoid homorganic bases, at least in Chichewa, cf. Table 4.⁹ I will therefore assume that the statistical universal of Similar Place Avoidance (SPA) is responsible for the categorical prohibition of the coronal / non-coronal superclasses in Tiene.

Having provided a possible motivation for why C₂-C₃ consonants cannot be [αcoronal], the remaining question we must address is why they cannot

Table 4. Distribution of unproductive suffixes *-am-, *-uk-, *-Vt- and *-ud- in Chichewa

		P -Vm-	K -uk- / -ok-	T -Vt-	-ul- / -ol-
C2	P	–77 %	+23 %	+52 %	+12 %
	K	–25 %	–78 %	+150 %	+17 %
	T	+43 %	+17 %	–77 %	–11 %
	Y	–7 %	+15 %	–11 %	–3 %

occur in the order non-coronal + coronal (*P-T, *K-T). As indicated in (22) on the current page, there are three logical possibilities for ruling out such sequences (henceforth, T stands for all coronals).

- (22) Three logical possibilities for ruling out non-coronal C₂ + coronal C₃
- a. a restriction on non-coronals: *P, *K in C₂ position
 - b. a restriction on coronals: *T in C₃ position
 - c. a restriction on sequences: *P-T, *K-T in C₂–C₃ positions

Having limited stems to a maximum of three syllables, the innovative prohibition may have been against medial non-coronals, final coronals, or non-coronal + coronal sequences (cf. Hyman and Inkelas 1997). The Tiene data thus do not point to a clear interpretation.

One possibility is that the coronal + non-coronal order is related to the special “unmarked” status of coronals in inventories, processes, language acquisition, and performance. It is often pointed out that coronals have greater frequency, assimilability, transparency, and distribution (Paradis and Prunet 1991). But why should unmarked T precede marked P and K? Perhaps there is a phonetic motivation: If it could be shown that coronals tend to be shorter in duration than coronals, perhaps Tiene phonologized “fast before slow”? As reported by Maddieson (1997: 630), however, available data are not consistent on this point:

- (23) Two studies of English stop consonant closure duration (in ms) by place of articulation
- | | |
|---------------------------------|----------------------|
| P > T > K | P > K > T |
| (Stathopoulos and Weismer 1983) | (Byrd 1993) |
| p : 96 t : 82 k : 72 | p : 69 t : 53 k : 60 |
| b : 92 d : 76 g : 68 | b : 64 d : 52 g : 54 |

While these studies measured stop closure, we would also want to know how the durations compare between nasals, but of oral sonorants, affricates and fricatives. It seems unlikely that Tiene intervocalic [s] and [l] are shorter in duration than [b] and [k] or were so at the time of phonologization.

In Section 3 it was argued that the historical process in developing the place constraints was one of metathesis: {P, K} V T > T V {P, K}. However, this metathesis goes against claims which have been made concerning the direction of historical metathesis or speech errors. As an example of long distance metathesis, Grammont (1933: 348) cites the example *beau-*

coup > *copou* as producing a sequence reflecting the “expiratory order”: $K \supset T \supset P$. Concerning contiguous CC metathesis, Blevins and Garrett’s (2004) generalizations in (24), taken together, produce the favored linear ordering: $K \supset P \supset T$.

(24) Generalizations of Blevins and Garrett (2004) concerning CC metathesis as sound changes

- | | |
|---------------------------|-----------------------------|
| a. $PK > KP$ | not: $*KP > PK$ |
| b. $*T\{P,K\} > \{P,K\}T$ | not: $*\{P,K\}T > T\{P,K\}$ |

Data from child language phonology also seems to be of little help in understanding Tiene (Macken 1996; Pater and Werle 2001; Fikkert and Levelt 2006; etc.) First, there is the well-known fact that “consonant harmony predominates”, e. g. *duck* → [gak], [dʌt] (Rose 2001). Second, as the examples in (25) show, metatheses often result in coronals being realized later:

(25) Non-coronal + coronal metatheses in child language (examples from Smith 1971)

- | | | | |
|-----------|-----------|---------------|-------------|
| a. [rikt] | ‘risk’ | b. [mizbərəl] | ‘miserable’ |
| [ɛplin] | ‘helping’ | [gæpəlin] | ‘galloping’ |

The examples in (25a) are cases of local CC metathesis, whereas those in (25b) involve consonant metathesis across a vowel. Concerning these latter, the generalization seems to go against Tiene:

“[cor] Cs metathesize to C₂ position in [‘CV.CV] and [CVC] words and to C₃ position in [‘CVCVCV], and to the right of string-adjacent [-cor]” (Macken 1996: 166).

What Tiene does have in common with child language phonology is that the latter is often templatic (Macken 1992, 1996; Vihman 2002). Macken distinguishes two kinds of learners (harmonic vs. templatic) who are both sensitive to coronality:

“... in the harmony systems ... coronals undergo harmony, particularly when they precede non-coronals; in melody templates, coronals are sequenced to the right of non-coronals” (Macken 1992: 350).

Macken considers non-coronals dominant and coronals non-dominant and adds:

“Nondominant features occur in prosodically nondominant positions, such as in codas or foot-internally, or in positions of neutralization; dominant features occur in prosodically dominant positions, such as foot initial and onsets generally” (Macken 1996: 166).

There is reason to think that this property of child language phonology extends to adult phonology as well. As the first of two examples, consider the realization of pluractional *-ta* in Kashaya:

(26) Pluractional *-ta* in Kashaya (Buckley 2000: 16)

- a. suffixation after T
 - i. *dahqoʔol-* → *dahqoʔol-ta-* ‘fail (to do)’
 - ii. *diʔan-* → *diʔan-ta-* ‘bruise by dropping’
- b. infixation before P, K
 - i. *bilaq^ham-* → *bilaq^ha-ta-m-* ‘feed’ **bi.la.q^ham.ta*
 - ii. *sima:q* → *sima-ta-q-* ‘go to sleep’ **si.ma:q.ta*

Whereas *-ta* is suffixed after a base which ends in a coronal in (26a), it is infix in (26b), where the base ends in a non-coronal consonant. Buckley explains that *-ta* is infix in order to avoid non-coronal codas (as suffixation would produce in the starred forms to the right in (26b)). Following Macken, Kashaya *-ta* affixation avoids producing dominant {P, K} in a non-dominant (i. e. coda) position.

The second example concerns the distribution of root-internal consonants in Mathimathi. As seen from the lexical counts made by Gahl (1996) in Table 5a, both consonants of CVC roots can be coronal or non-coronal. In Table 5b, however, we see that there are no non-coronal consonants in the C₂ position of CVCVC roots. This is indeed very reminiscent of Tiene. Using the term “weak” in place of Macken’s “non-dominant”, the facts seen thus far can be united as in (27).

(27) Tendency to align coronals with weak prosodic positions

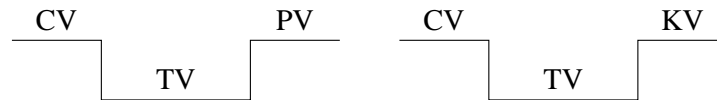
- a. Kashaya: coda vs. onset
- b. Bantu in general: late vs. early (roughly, root vs. post-root)
- c. Tiene and Mathimathi: medial vs. peripheral

Table 5. Root-internal C must be coronal in Mathimathi (Gahl 1996; data from Hercus 1969, 1986)

a. CVC roots			b. CVCVC roots			
	C ₁	C ₂		C ₁	C ₂	C ₃
T	66	79	T	25	104	30
P, K	110	97	P, K	79	0	74

As was seen in Section 4, Koyo, Bobangi, and most Bantu show more coronality in later consonant positions within the stem, especially those following the initial CV(V)C- root. The prosodic structure of such languages is therefore strong (root) vs. weak (post-root). In Tiene (and apparently Mathimathi), there is instead a weak medial CV or “nadir” (Hyman and Inkela 1997). As schematized in (28), the *prosodic trough* (τ), introduced for Yaka (Hyman 1998), consists of the middle C_2V_2 sequence of a CVCVCV stem in Tiene, where $C_2 = T$ and V_2 is predictable from V_1 :

(28) The prosodic trough in Tiene



If we are correct in extending Macken’s observations on child language phonology in this way, we can summarize the account reached for Tiene consonant place restrictions as follows: (i) [α coronal] C_2 - C_3 sequences are prohibited by a categorical version of the statistical universal of Similar Place Avoidance (Pozdniakov and Segerer 2007). (ii) Non-coronals are restricted from the prosodic trough. Taken together, this account perhaps represents a bit of an overkill in the sense that a non-coronal C_2 - C_3 sequence would be ruled out by either constraint. It also may not be as fine-tuned as needed, since it lumps together all coronals vs. all non-coronals. In order to get more perspective on both of these issues, the next sections briefly describe three other languages that have Tiene-like properties.

6 Other Teke

In the preceding sections reference has been made to the Teke subgroup of Bantu to which Tiene has sometimes been claimed to belong. It should not be surprising that other languages of this subgroup have at least some of the same properties found in Tiene. In this section I will briefly describe two Teke languages for which we have data and then turn in Section 7 to the Nigerian Plateau language Izere, which is quite distant, both geographically and genetically.

The prosodic stem has been described in great detail for Kukuya by Paulian (1975). The relevant properties of Kukuya prosodic stem are summarized in (29).

- (29) The prosodic stem in Kukuya (Paulian 1975; cf. Hyman 1987)
- a. Five syllable shapes: CV, CV.V, CV.CV, CVV.CV, CV.CV.CV
 - b. Six C₂ or C₃ consonants: /p, t, k, l, m, n/ (vs. large inventory of C₁ consonants)
 - c. Six C₂-C₃ combinations: C-n-m, C-t-k, C-l-k, C-l-p, C-t-p, C-k-p
 - d. Weak C₂, C₃ realizations: /p, t, k/ are realized [b ~ β], [r] and [k ~ g ~ ɣ], respectively.
 - e. V₂ of CVCVCV is almost totally predictable and is subject to reduction or deletion
 - f. Five tonal “melodies”: L, H, LH, HL, LHL

As seen, Kukuya stems are limited to the same five shapes as in Tiene, with a maximum of three syllables. In terms of consonant distribution it has gone beyond Tiene in two ways. First, only six consonants appear as C₂ or C₃, even in CV(V)CV stems. (We saw in Table 1 on page 156 that Tiene allows 9 out of its 12 consonants in this position.) Second, there are only six combinations of C₂-C₃ in trisyllabic stems, exemplified in (30).

- (30) Among CVCVCV stems, “ne sont attestées que” [are not attested but] (C₂-C₃): (Paulian 1975)
- a. C - n - m : /(kì-) .púnumà/ ‘accidentally knock over’ 34.57 %
 C - l - k : /(kì-) .bólókò/ ‘break’ 19.15 %
 C - l - p : /(kì-) .lèlèpè/ ‘slow down’ 9.84 %
 C - t - p : /(kì-) .nàtàpà/ ‘be fixed, stuck’ 3.19 %
 - b. C - k - p : /(kì-) .pákapà/ ‘crack, be torn (intr.)’ 6.92 %

The six combinations agree in nasality and respect the prohibition against [α coronal]. Those in (30a) mirror the coronal + non-coronal sequencing found in Tiene. Unlike Tiene, however, the velar + labial combination in (30b) is allowed. On the other hand, Kukuya has eliminated labials from C₂, possibly by metathesis: Proto-Bantu **papuk-a* ‘become torn’ > *pákapà* ‘crack, be torn’, Proto-Bantu **tombuk-a* ‘burst open’ > *ókopò* ‘burst open (abscess)’. Verb extensions are not productive in the language (Christiane Paulian, personal communication).

The second language is a variant of Teke spoken in Gabon whose properties in (31) are drawn from a lexicon of 1,466 items (Hombert 1993):

(31) The prosodic stem in Teke-Gabon

- a. Five syllable shapes: CV, CV.V, CV.CV, CVV.CV,
CV.CV.CV
- b. Seven C₂ or C₃ consonants: /b, r, g, l, m, n, ŋ/ (vs. large inventory of C₁ consonants)
- c. Five C₂-C₃ combinations: C-b-g C-m-ŋ, C-l-g C-r-g, C-n-ŋ

Again we see the same five stem shapes and a maximum of three syllables in (31a). While Teke-Gabon allows a wide array of C₂ consonants in CV(V)CV stems (including prenasalized /mb, nd, nj, ŋg/), the seven consonants in (31b) form the five C₂-C₃ combinations in (31c). As seen, velars are excluded from C₂, while only velars can occur in C₃ (cf. Koyo's C₄ in Table 2 on page 159). Again, C₂ and C₃ must agree in nasality. Given the seven consonants in (31b), which may be derived from /p, t, k, l, m, n, ŋ/, nasal agreement, and the bidirectional constraint "if C₃, then velar; if velar, then C₃", exactly the five combinations in (31c) will be well-formed.

The facts from *Tiene*, *Kukuya* and *Teke-Gabon* suffice to show that the languages of this area exhibit variations on a theme, as summarized in Table 6 in which I have added *Izere*, to be discussed next. As seen, all four languages prohibit non-coronal + non-coronal C₂-C₃. They also all allow CVTVK. Beyond this they differ: *Kukuya* allows CVKVP (but not CVPVK), while *Teke-Gabon* and *Izere* allow CVPVK (but not CVKVP). Finally, *Teke-Gabon* alone disallows CVTVP, since C₃ must be velar.

Table 6. Heterorganic C₂-C₃ combinations ranked in terms of most to least prohibitions

	C ₃ = T		C ₂ & C ₃ = P, K		C ₂ = T, C ₃ = P, K	
	CVPVT	CVKVT	CVKVP	CVPVK	CVTVP	CVTVK
<i>Tiene</i>	*	*	*	*		
<i>Kukuya</i>	*	*		*		
<i>Teke-Gabon</i>	*	*	*		*	
<i>Izere</i>	*	*	*			

What this means is that slightly different constraints have been imposed in the languages for which we have information. It is tempting to say that the initial trigger of the above variations was getting coronals out of C₃. However, this does not quite work for *Izere*, as we shall now see.

Table 7. Maximum stem = CVCVC

C ₂ /C ₃	P	K	T
P	*	CVbVk CVmVŋ	*
K	*	*	*
T	CVrVp CVrVm CVsVp CVsVm	CVrVk CVrVŋ CVnVŋ CVsVk CVsVŋ	CVrVs

7 Izere

Izere (a.k.a. Afuzare and Zarek) is a Northern Nigerian language belonging to the Plateau subbranch of Niger-Congo whose relevant phonological and morphological properties have been studied by Wolff and Meyer-Bahlburg (1979), Gerhardt (1984), and Blench (2001). Although distantly related to Bantu, Izere is geographically quite far from Tiene and its close relatives. Still, as seen in Table 7, based on 2,178 entries from Blench and Kaze (2000), the maximum CVCVC stem shows remarkable similarities with the Teke languages.

Izere is like Tiene except for the bolded sequences: Like Teke-Gabon, it allows C₂-C₃ to be CVPVK. Unlike any of the Teke languages, however, it allows one CVTVT shape, CVrVs, occurring only in pluractional forms (see rightmost column).

In addition, nasal agreement is unidirectional in Izere: If C₂ = nasal, then C₃ = nasal, but if C₃ = nasal, C₂ may or may not be. All 12 acceptable sequences are illustrated in the verbs in Table 8 on the facing page, where it should also be noted that Izere does not require an inflectional final vowel. Again, coronals are missing from C₃ position, except for /s/, which must be preceded by [r]. While Izere contrasts approximately 25 consonants in C₁ position, Table 9 on the next page shows severely restricted inventories in C₂ and C₃ positions.

Besides the striking distributional similarities, Izere has in common with Tiene that verb pluractional (~ habitual) formation may involve either suffixation or infixation:

Table 8. CVCVC verb stems

CVTVP		CVTVK		CVPVK		CVTVT	
<i>tíríp</i>	‘rub’	<i>burúk</i>	‘stir’	<i>túbùk</i>	‘stab’	<i>sàràs</i>	‘tear + PL’
<i>kurúm</i>	‘coax’	<i>káráŋ</i>	‘pay’	<i>rómóŋ</i>	‘bite’		
<i>gesèp</i>	‘stammer’	<i>kánáŋ</i>	‘fry’				
<i>kósóm</i>	‘cough’						

Table 9. C₂ & C₃ consonants (CVVp and CVVk are both marginal)

CVC			CVCVC			CVCVC			CVVC		
p	r	k	b	r		p		k	(p)	r	(k)
m	n	ŋ	m	n		m		ŋ	m	n	
	s			s			s			s	

(32) Suffix vs. infix -s-(~ -r-) in pluractional formation

a. CV → CVs (12 cases)

<i>bó</i>	→	<i>bós</i>	‘fetch’	<i>dí</i>	→	<i>dís</i>	‘see’
<i>kpà</i>	→	<i>kpàs</i>	‘fall’	<i>sɛ</i>	→	<i>sɛs</i>	‘locate, find’

b. CVP, CVK → CVVs (14 cases)

<i>rép</i>	→	<i>réés</i>	‘sell’	<i>nɔk</i>	→	<i>nóós</i>	‘build’
<i>káp</i>	→	<i>káás</i>	‘farm’	<i>fók</i>	→	<i>fóòs</i>	‘hear’
<i>tóm</i>	→	<i>tóós</i>	‘send’	<i>gaŋ</i>	→	<i>gáás</i>	‘finish’
<i>nyim</i>	→	<i>nyíís</i>	‘meet’	<i>tseŋ</i>	→	<i>tséés</i>	‘walk, go’

c. CVn → CVrVŋ (9 cases), CVsVŋ (4 cases)

<i>men</i>	→	<i>mérèŋ</i>	‘lie down’	<i>bún</i>	→	<i>búsúŋ</i>	‘break (wood, bones)’
<i>kon</i>	→	<i>koróŋ</i>	‘rub’	<i>tén</i>	→	<i>téséŋ</i>	‘cut’
<i>tún</i>	→	<i>túrúŋ</i>	‘remove’	<i>shán</i>	→	<i>sháshàŋ</i>	‘buy’ (+ sibilant harmony)

d. CVr → CVsVk (10 cases)

<i>gor</i>	→	<i>gósók</i>	‘pass’	<i>tár</i>	→	<i>tásák</i>	‘shout, yell’
<i>nár</i>	→	<i>násàk</i>	‘surpass’	<i>tsér</i>	→	<i>tsésék</i>	‘look for, want’

e. CVs → irregular (4 cases)

<i>mas</i>	→	<i>manaŋ</i>	‘laugh’	<i>rús</i>	→	<i>tsór</i>	‘hit, beat, strike’
<i>shésh</i>	→	<i>shíshék</i>	‘save (s.o.)’	<i>rus</i>	→	<i>tsor</i>	‘groan in pain’

In the data base there are 539 verbs of which 181 (or 34%) have a derived pluractional form. (The number of examples of each pattern is given in parentheses.) In (32a) we see that CV verbs take an -s suffix. The same appears to be true in (32b) when a CVC verb ends in a non-coronal. As seen, the labial or velar consonant drops out. Thus compare Izere *nək* → *nóós* ‘build + PL’ with its analogue in nearby Berom: *lok* → *logos* ‘build + PL’ (Blench 2005). The forms in (32c) show infixation of either -r- or -s-. In addition, the /n/ of the root becomes [ŋ] in accordance with the requirement that C₃ be non-coronal. The same interpretation is possible in (32d), if we assume that C₂ [r], which reconstructs as *t (Wolff and Meyer-Bahlburg 1979; Gerhardt 1984) is underlyingly /t/ (cf. Izere *nók*, Berom *not* ‘give’). Starting with /got/ ‘pass’, we first derive *gosot*, which then becomes *gosok* ‘pass + PL’ since C₃ must be non-coronal.

That infixation is required is clearly seen in the following derivations:

(33) CVPVK → CVsVP unambiguously requires infixation (and loss of velar C₃)

a. CVbVk → CVsVp (7 cases), CVsVm (1 case)

<i>fúbúk</i>	→	<i>fúsùp</i>	‘sip’
<i>túbùk</i>	→	<i>túsùp</i>	‘stab’
<i>kábák</i>	→	<i>kasàp</i>	‘share out’
<i>nabak</i>	→	<i>násap</i>	‘liftup, stretch’
<i>kóbók</i>	→	<i>kósóp</i>	‘loan, borrow’
<i>fébék</i>	→	<i>fésèm</i>	‘blow’

b. CVmVŋ → CVsVp (3 cases), CVsVm (1 case)

<i>rímíŋ</i>	→	<i>rísìp</i>	‘kick’
<i>shímíŋ</i>	→	<i>shíshìp</i>	‘wake up, rise’
<i>bòmóŋ</i>	→	<i>bosòp</i>	‘learn, try, teach’
<i>tómóŋ</i>	→	<i>túsóm</i>	‘push’

When the verb has the shape CVPVK, the plural form is derived by infixing -s- into the C₂ position.¹⁰ In (33a) the labial C₂ of the singular form appears as C₃, where it is devoiced, and the C₃ velar of the singular is deleted. Also reminiscent of Tiene, three of the four pluractional forms in (33b) in-

volve denasalization of /m/ as C₃. The last example of each set shows that there is some irregularity concerning nasality.

As seen in Table 10, when the base verb has the shape CVTVK, the corresponding pluractional forms show considerable variation:

Table 10. Multiple possible forms of the plural of CVTVK verbs with number of found cases of each

	# SG.	# PL.	CVrVs	CVrVk	CVsVk	CVsVɿ	CVs	CVr	CVk	CV
CVrVk	40	26	13	4	2		1	1	1	4
CVrVɿ	20	1	1							
CVnVɿ	24	12	2	2	1	4	2			
CVsVk	12	6					6			
CVsVɿ	9	2				1	1			

First note that the three shapes whose C₂ and C₃ agree in nasality (CVrVk, CVnVɿ, CVsVk) have the highest percentage of pluractional forms (cf. only one out of 20 CVrVɿ verbs). Examples of some of the patterns are given in (34).

(34) Different patterns when singular is CVTVK

a. CVrVk → CVrVs (13 cases)

<i>kárák</i>	→	<i>káràs</i>	‘open’
<i>kórók</i>	→	<i>kóròs</i>	‘pour’
<i>wórók</i>	→	<i>woros</i>	‘throw, fling’
<i>yírík</i>	→	<i>yírìs</i>	‘destroy, demolish’

b. CVrVk → CVrVk (with tone change) (5 cases)

<i>burúk</i>	→	<i>burùk</i>	‘stir’
<i>berék</i>	→	<i>bèrek</i>	‘support’
<i>shirík</i>	→	<i>shirik</i>	‘frighten, scare’
<i>birík</i>	→	<i>bírik</i>	‘cancel, erase’

c. CVsVk → CVs (all 6 such singular verbs undergo this process)

<i>basák</i>	→	<i>bás</i>	‘seal, paste’
<i>fósók</i>	→	<i>fós</i>	‘peel’ (tree bark)
<i>tásák</i>	→	<i>tás</i>	‘pierce, winnow’
<i>bísík</i>	→	<i>bís</i>	‘untie, unfold’
<i>kpísík</i>	→	<i>kpís</i>	‘split chunk off larger part’
<i>mísík</i>	→	<i>mís</i>	‘sprinkle, pour away’

d. CVrVk, CVnVɿ → CVrVs ~ CVrVk, CVsVɿ	
<i>bárák</i> → <i>bárás</i> ~ <i>barak</i>	‘throw’
<i>dorók</i> → <i>dóròs</i> ~ <i>dorok</i>	‘leave’
<i>fúruk</i> → <i>fúrùs</i> ~ <i>furuk</i>	‘jump’
<i>tárák</i> → <i>táràs</i> ~ <i>tàràk</i>	‘spread out’
<i>fíníɿ</i> → <i>fírìs</i> ~ <i>fírìk</i>	‘sun-dry’
<i>tónòɿ</i> → <i>tóròs</i> ~ <i>túsòɿ</i>	‘show’

Most significantly, all 16 CVrVs verbs derive from CVTVK, either from CVrVk (13 cases), CVrVɿ (1 case) or CVnVɿ (2 cases), suggesting that in just this one case it was hard to avoid a CVTVT output. Still, as seen in (34d), six of the 16 CVrVs pluractionals have a variant of the shape CVTVK. It is hard to determine whether one variant is older than the other.

The following summarizes the aspects of Izere which resemble the Teke languages:

- (35) Special properties of Izere verb stems
- five stem shapes: CV (100), CVC (165), CVVC (72), CVCV (9), CVCVC (183)¹¹
 - limited inventory of C₂ and C₃ consonants in (40) (cf. especially Kukuya in (29b))
 - C₂-C₃ is limited to CVTVP (27), CVTVK (109), CVPVK (36) and CVrVs (16PL)
 - V₁=V₂ in 167 / 183 or 91 % of CVCVC stems
 - s suffix overrides root non-coronal C₂ (vs. coronal C₂ in Tiene);
 - s- and -r- infixes precede non-coronal C₃ (cf. -s- in Tiene)

8 Towards an account

Recall that we have had some difficulty finding support for restricting coronals to internal position. It also is rare for a suffix to be infixes because of its place of articulation. Given that such a distant language as Izere has so much in common with the Teke languages, we confidently conclude that they must share a common motivation. I suggest that the precondition that allowed for these languages to progress as they did was the limitation of the prosodic stem (root + suffixes) to a maximal triconsonantal structure: CVCVC-V in Tiene, CVCVC in Izere. With this established it meant that there are exactly

three positions: initial, medial and final. Since these same languages show stem-initial prominence (e. g. more oppositions in the first CV of the stem), only the C₂ and C₃ were available to be identified with specific places of articulation. Languages which impose a greater maximum or no maximum are not likely to reorganize the system by place of articulation.

With prosodic maximality as a backdrop, I speculate that the Teke and Izere facts were initially triggered by the codification of Similar Place Avoidance (Section 5), followed by different restructurings. Using T to stand for all coronals, *CVPVP-, *CVTVT- and *CVKVK- became prohibited in various stages. It is known that the fusion or “imbrication” of perfective *-*id-ε*, applicative *-*id-* and causative *-*ic-i-* suffixes which is widely attested in Bantu (Bastin 1983) first applied to bases which ended in coronals. The prohibition against *CVTVT- may therefore have predated the others in Teke. It certainly predated the development of the non-coronal constraints *CVPVK-, *CVKVP-, which are differentially observed within the Teke languages (and Izere).

Given the relative non-productivity of non-coronal suffixes in Bantu seen in (19), the effect of these prohibitions may have been to marginalize or lose P and K suffixes. Since coronal suffixes were more productive, some kind of accommodation would have had to be made for *CVT-VT-. As we saw in (3a) and (4a), causative -Vs- and applicative -Vl- avoid *CVTVT- by fusing with a CV(V)T- root. Izere relies on the dissimilation of C₃ T to K and metathesis. This produces the derivations in (36).

(36) Infixation and velarization in Izere

- a. **got* [*gor*] ‘pass’ **got-Vs* > *gotos* > *gosot* > *gosok* ‘pass + PL’
- b. **tén* [*tén*] ‘cut’ **tén+Vs* > *ténés* > *tésén* > *téséŋ* ‘cut + PL’
- c. **kon* [*kon*] ‘rub’ **kon+Vt* > *konot* > *koton* > *koron* ‘rub+PL’

In (36a,b) the pluractional suffix is -Vs-, while in (36c) it is -Vt-. In each case the suffixal consonant metathesizes with the coronal C₂ of the root. **t* is realized [r] as C₂, but [k] as C₃. The in (36a,b) appears to be motivated by the fact that velarization of C₃ *s would have produced [x], which does not otherwise exist in the language. This, however, cannot explain why metathesis is required in (36c). There we note that a non-metathesized output [**konok*] would violate the Izere constraint that if C₂ is nasal, then C₃ must also be. One might try to fix this up by producing [**konon*], which may be avoided for morphological reasons: Of the 84 CVCVC pluractional verbs in

Blench and Kaze (2000), only two irregular verbs have nasal C_2 and C_3 : *mas* → *manaj* ‘laugh + PL’, *tséém* → *tséméj* ‘sift + PL’. Given that pluractionality is marked by [s] and [r] (*t), [**kononj*] may just sound too singular.¹²

The suggestion made here is that once a language both limits its maximal prosodic stem to CVCVC- and introduces severe constraints on C_2 - C_3 homorganicity, it invites other restructurings. For example, Tiene speakers would be justified in interpreting CVT-Vs- > CVVs- as a coronal C_3 becoming C_2 . If this can serve as the basis of analogy, it might be extended to shift other coronals to C_2 , i. e. CVP-Vs-, CVK-Vs- > CVsVP-, CVsVK-. On the other hand, from the other end of the stem, the C_3 velar of the once productive durative/imperfective suffix **-a(n)g-* may have served as the model for the requirement that C_3 = velar in Teke-Gabon.

There are, however, at least two problems in applying this account to Izere: (i) If CVPVT, CVKVT > CVVT by intervocalic P-/K-deletion, how or why do labials survive in CVPVK? (ii) If CVTVT is prohibited, how does (pluractional) CVrVs survive? Concerning the first question, considerably less is known about pre-Izere than pre-Tiene, which derives from Proto-Bantu. It is possible that we should reconstruct *CVCs and perhaps other clusters in pre-Izere. In this case *CVPs and *CVKs would become CVVs by cluster simplification. *CVPK, on the other hand might have escaped cluster simplification because the second consonant is not coronal or because an epenthetic vowel was inserted to produce CVPVK. Similarly, it is possible that *CVrs escaped cluster simplification and only later became CVrVs. Perhaps there was a contrast in pre-Izere between CVCC and CVCVC. Since I have argued that C_3 *T > K, there also are possible complications deriving from the neutralization of C_3 *T and *K. Finally, it should be noted that some of the C_3 consonants may have been suffixes even on singulars, and that there may even have been a possibility of multiple suffixation (Wolff and Meyer-Bahlburg 1979, Gerhardt 1984).

While the above account is admittedly speculative, and leaves open a number of questions, one can feel at least confident about Similar Place Avoidance as a trigger in paring down the number of consonant combinations in a CVCVC- stem. Compare in Table 11 on the next page the possible CVCVC verb stems shapes in Izere vs. nearby Berom (based on Blench et al. 2006).

While 9 out of 12 Izere CVCVC forms have coronal + non-coronal C_2 - C_3 , only 12 out of 30 conform to this template in Berom. In fact, even though C_3 may not be /n/, this still allows 16 different CVTVT shapes. What we see from Berom, however, is that the constraints are severest on combina-

Table 11. Comparison of CVCVC in Izere vs. Berom verbs.

a. Izere (9 / 12 = CVTVp, CVTVk).

b. Berom (12 / 30 = CVTVp, CVTVk)

a.				b.			
C ₂ /C ₃	P	K	T	C ₂ /C ₃	P	K	T
P	*	CVbVk CVmVɲ	*	P	*	CVmVɲ	CVbVt
							CVbVs
							CVmVt
							CVmVl
K	*	*	*	K	CVgVm	*	CVmVs
							CVgVt
							CVgVs
							CVɲVt
T	CVrVp CVrVm CVsVp CVsVm	CVrVk CVrVɲ CVnVɲ CVsVk CVsVɲ	CVrVs	T	CVrVp CVrVm CVIVm CVsVp CVsVm	CVrVk CVrVɲ CVIVk CVIVɲ CVnVɲ CVsVk CVsVɲ	CVɲVl
							CVɲVs
							CVIVt
							CVIVs
T	CVrVp CVrVm CVsVp CVsVm	CVrVk CVrVɲ CVnVɲ CVsVk CVsVɲ	CVrVs	T	CVrVp CVrVm CVIVm CVsVp CVsVm	CVrVk CVrVɲ CVIVk CVIVɲ CVnVɲ CVsVk CVsVɲ	CVnVt
							CVnVs
							CVnVl
							CVsVl

9 Summary

In the preceding sections we have seen that two separate groups of Niger-Congo languages have independently introduced place of articulation constraints on consonant positions within the prosodic stem, which in turn determine whether derivations will involve suffixation or affixation. Although many languages and most of the external evidence from language acquisition and language change suggest that coronals should be realized after non-coronals, these languages have a preference for placing coronals before non-coronals, i. e. CVTVP and CVTVK. Historically, these systems result from the interaction of a number of changes, including metathesis of the C₂ and C₃ of inherited CVPVT and CVKVT sequences. The fact that this happens independently in the Teke Bantu region and on the Jos Plateau of Nigeria suggests that such sequencing is not isolated or accidental. Based on additional evidence from Kashaya and Mathimathi, it was suggested that the most likely synchronic motivation is that coronals gravitate to “weak” positions. While it appears more common that a language will define strong vs. weak in terms of early vs. late positions within a prosodic domain, in these languages, the weakest position is the medial prosodic trough. Tiene speakers appear to have given an [s-w-s] interpretation to the maximal CVCVC-V stem vs. the more usual [s-w-w] interpretation of Koyo speakers in Table 2 on page 159. Interestingly, C₂-C₃ [w-s] appears to be particularly sensitive to place of articulation while C₂-C₃ [w-w] shows a gradual weakening along the traditional strength hierarchies (voiceless → voiced; non-continuant → continuant) (cf. Williamson 1979). In other words, it is [s-w-s] structure that accounts for the relative rarity of Tiene and Izere affixation by place of articulation. That being the case, we would learn a lot more if the rare and (perhaps) mysterious variations in the Teke and Plateau areas were studied in greater comparative detail.

Abbreviations

C = consonant; FV = final vowel; GCB = Guthrie Common Bantu; H = high tone; itr. = intransitive; K = velar; L = low tone; MAX = a constraint in Optimality Theory; NB = *nota bene*; NW = Northwest; P = labial; PB = Proto-Bantu; pl. = plural; sg. = singular; SPA = Similar Place Avoidance; T = dental-alveolar; tr. = transitive; V = vowel; Y = palatal

Notes

1. This process of fusion is reminiscent of a widespread process of root-suffix fusion in Bantu known as “imbrication”, most frequently involving the PB **id-ε* perfective ending (Bastin 1983). Since imbrication sometimes extends to applicatives and causatives, it could be that the Tiene forms derive from this process.
2. Note that C₂ /ŋ/ is extremely rare, occurring mostly (only?) in borrowings, e. g. *lɔŋ-ɔ* ‘load’, *lɔŋŋ-ɔ* ‘load for’. While it does not occur in Ellington’s materials, the expected causative form would be *lɔsɔk-ɔ*, with denasalization of /ŋ/.
3. Compare the final vowel morpheme which is /-ε/ after “extended” verbs, otherwise /-a/. /-ε/ assimilates to a preceding [ε] or [ɔ], while /-a/ assimilates to a preceding [ε] or [ɔ].
4. Ellington (1977) does not report a corresponding root for any of the CVCVC- verbs in (13a). However, because their FV is -ε rather than -a, (cf. note 3) the forms *vútek-ε* ‘come back’ and *tóleb-ε* ‘pierce’ suggest that they may have been morphologically complex at one time.
5. /s/ and /z/ are indicated as “proto-palatals” since they derive from Proto-Bantu **c* and **j*, respectively. The glides /w/ and /y/ are not indicated in Table 1, as they do not appear in C₂ or C₃ position. Finally, /pu/ and /bu/ are realized [fu] and [vu], respectively.
6. Concerning vowels, /i, e, ε, u, o, ɔ, a/ contrast in V₁ position, while only /i, u, a/ contrast as V₂, V₃ and V₄ in Koyo.
7. The following discussion is restricted to verbs, since noun stems are typically limited to two syllables unless they are derived (e. g. from verbs), reduplicated, or borrowed.
8. My thanks to Michael Cysouw for doing the percentage calculations as well as other helpful comments. The fact that palatals are overrepresented in C₁ position is largely due to special status of the glide [y] in Bantu (cf. Swahili in (21) on page 162).
9. The calculations in Table 4 on page 163 are based on 1,412 trisyllabic or longer verbs extracted from Scott and Hetherwick (1957) as adapted by Al Mtenje for the Comparative Bantu On-Line Dictionary project.
10. There are four exceptions: *rómóŋ* → *rós* ‘bite-PL’, *tsíbík* → *tsip* ‘twist-PL’, *téméŋ* → *tém* ‘cut-PL, chop down-PL’, *zímíŋ* → *zim* ‘fling-PL, swing-PL’. Forms like the first three motivate Wolff and Meyer-Bahlburg (1979) and Gerhardt (1984) to propose a singular suffix -k (~ -ŋ) which is missing in the plural. Blench’s (2005) reconstruction of singulative **-tV* in Berom suggests that this -k may have originally been **-t* in Izere as well.
11. The numbers do not add up to 539 because of exceptional verbs. For example, two verbs have the shape CVCVVC with exceptional C₂-C₃ sequences: *kutáám* ‘lose direction’, *témòòr* ‘throw (piece of food) into mouth at once’.
12. A few verbs form their pluractional by denasalization, or by -s- affixation + denasalization, e. g. *fíníŋ* → *fírik* ~ *fírís* ‘sun dry-PL’.

References

- Bach, Emmon and Robert Harms
 1968 How do languages get crazy rules? In *Linguistic change and generative theory*, Robert P. Stockwell and Ronald K. S. Macaulay (eds.), 1–21. Bloomington: Indiana University Press.

- Bastin, Yvonne
 1983 La finale *-IDE* et l'imbrication en bantou. Tervuren: *Annales du Musée Royale de l'Afrique Centrale*. Série IN-8. 114.
- Becher, Jutta
 2000 Verbaextensionen in den Atlantischen Sprachen. *Hamburger Afrikanistische Arbeitspapiere* 1: 1–36.
- Blanchon, Jean
 1995 Punu lexical database in Filemaker Pro™.
- Blench, Roger
 2001 Plural verb morphology in Fobur Izere. Manuscript. Cambridge, UK.
- Blench, Roger
 2005 Plural verb morphology in Eastern Berom. Manuscript. Cambridge, UK.
- Blench, Roger and Bitrus Bulus Kaze
 2000 A dictionary of the Izere language of Fobur. Manuscript. Cambridge, UK.
- Blench, Roger et al.
 2006 Berom dictionary. Manuscript. Cambridge, UK.
- Blevins, Juliette and Andrew Garrett
 2004 The evolution of metathesis. In *Phonetically-based phonology*, Bruce Hayes, Robert Kirchner and Donca Steriade (eds.), 117–156. Cambridge etc.: Cambridge University Press.
- Bouquiaux, Luc
 1970 *La langue Birom*. Paris: Les Belles Lettres.
- Buckley, Eugene
 2000 Explaining Kashaya infixation. *Proceedings of the Berkeley Linguistics Society* 23: 14–25.
- Byrd, Dani
 1993 54,000 American stops. *UCLA Working Papers in Phonetics* 83: 97–116. Los Angeles: University of California, Los Angeles.
- Chomsky, Noam and Morris Halle
 1968 *The sound pattern of English*. New York: Harper and Row.
- Ellington, John
 1977 *Aspects of the Tiene language*. Doctoral dissertation, University of Wisconsin, Madison.
- Fikkert, Paula and Claartje Levelt
 2006 How does place fall into Place? Manuscript. Raboud University of Nijmegen and Leiden University.
- Fontaney, Louise
 1980 Le verbe. In *Eléments de description du punu*, François Nsuka-Nkutsi (ed.), 51–114. Centre de Recherches Linguistiques et Sémiotiques de Lyon, Université Lyon 2.
- Frisch, Stefan, Janet Pierrehumbert and Michael Bro
 2004 Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22: 179–228.
- Gahl, Susanne
 1996 Syllable onsets as a factor in stress rules: the case of Mathimathi revisited. *Phonology* 13: 329–344.

- Gerhardt, Ludwig
1984 More on the verbal system of Zarek (Northern Nigeria). *Afrika und Übersee* 67: 11–29.
- Grammont, Maurice
1933 *Traité de phonétique*. Paris: Librairie Delagrave.
- Grégoire, Claire
1979 Les voyelles finales alternantes dans la conjugaison affirmative des langues bantoues centrales. *Journal of African Languages and Linguistics Leiden* 1: 141–172.
- Guthrie, Malcolm
1953 *The Bantu Languages of Western Equatorial Africa*. London/New York: Published for the International African Institute by Oxford University Press.
- Guthrie, Malcolm
1960 Teke radical structure and common Bantu. *African Language Studies* 1: 1–15.
- Guthrie, Malcolm
1967–1971 *Comparative Bantu: an introduction to the comparative linguistics and pre-history of the Bantu languages*, vols. 1–4. London: Greggs.
- Hercus, Luise A.
1969 *The languages of Victoria: a late survey*. Canberra: Australian Institute of Aboriginal Studies.
- Hercus, Luise A.
1986 *Victorian languages, a late survey*. (Pacific Linguistics; B-77). Canberra: Department of Linguistics Research School of Pacific Studies, Australian National University. [Expanded version of Hercus (1969)]
- Hombert, Jean-Marie
1993 Lexique teke. Electronic data base. Lyon: Laboratoire Dynamique du Langage.
- Hyman, Larry M.
1987 Prosodic domains in Kukuya. *Natural Language and Linguistic Theory* 5: 311–333.
- Hyman, Larry M.
1996 Lexique koyo. Electronic database, 1598 entries. Berkeley, CA: Department of Linguistics, University of California, Berkeley.
- Hyman, Larry M.
1998 Positional prominence and the ‘prosodic trough’ in Yaka. *Phonology* 15: 41–75.
- Hyman, Larry M.
2003 Basaa A.43. In *The Bantu languages*. Derek Nurse and Gérard Philippson (eds.), 257–282. London: Routledge.
- Hyman, Larry M.
2008 Directional asymmetries in the morphology and phonology of words, with special reference to Bantu. *Linguistics* 46: 309–349.

- Hyman, Larry M. and Sharon Inkelas
 1997 Emergent templates: the unusual case of Tiene. In *Selected phonology papers from H-OT-97*. Viola Miglio and Bruce Morén (eds.), 92–116. (University of Maryland Working Papers in Linguistics; 5). Baltimore: Department of Linguistics, University of Maryland.
- Lemb, Pierre and François de Gastines
 1973 *Dictionnaire basaa-français*. Douala: Collège Libermann.
- Leroy, Jacqueline
 1982 Extensions en mankon. In *Le verbe bantou*. Gladys Guarisma, Gabriel M. Nissim, Jan Voorhoeve (eds.), 125–138. Paris/Leiden: Société d'Etudes Let Anthropologiques de France.
- Macken, Marlys A.
 1992 Lexical templates. *Chicago Linguistics Society* 28: 345–355.
- Macken, Marlys A.
 1996 Prosodic constraints on features. In *Proceedings of the UBC International Conference on Phonological Acquisition*. B. Bernhardt, J. Gilbert and D. Ingram (eds.), 159–172. Cambridge: Cascadilla Press.
- Maddieson, Ian
 1997 Phonetic universals. In *Handbook of Phonetic Sciences*. William J. Hardcastle and John Laver (eds.), 619–639. Oxford: Blackwell.
- Mathangwane, Joyce
 2001 Suffix ordering in the Ikalanga verb stem: A case against the Repeated Morph Constraint. *South African Journal of African Languages* 21: 396–409.
- McCarthy, John J.
 1981 A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* 12: 373–418.
- McCarthy, John J. and Alan Prince
 1999 Faithfulness and identity in prosodic morphology. In *The prosody-morphology interface*, René Kager, Harry van der Hulst and Wim Zonneveld (eds.), 218–309. Cambridge etc.: Cambridge University Press.
- Meeussen, A.E.
 1967 Bantu grammatical reconstructions. *Annales du Musée Royal de l'Afrique Centrale Série 8*, 61: 81–121. Tervuren: Musée Royal de l'Afrique Centrale.
- Paradis, Carole and Jean-François Prunet
 1991 Asymmetry and Visibility in Consonant Articulations. In *The Special status of coronals: internal and external evidence*, Carole Paradis and Jean-François Prunet (eds.), 1–29. San Diego: Academic Press.
- Pater, Joe and Adam Werle
 2001 Typology and variation in child consonant harmony. In *Proceedings of the Fifth Holland Institute for Generative Linguistics Phonology Conference (HILP5)*, Caroline Féry, Antony Dubach Green and Ruben van de Vijver (eds.), 119–139. Potsdam: University of Potsdam.
- Paulian, Christiane
 1975 *Le kukuya: langue teke du Congo*. Paris: Société d'Etudes Let Anthropologiques de France (SELAF).

- Pozdniakov, Konstantin and Guillaume Segerer
 2007 Similar place avoidance: A statistical universal. *Linguistic Typology* 11.3: 307–348.
- Rose, Yvan
 2001 Licensing and Feature Interaction Processes in Child Language. *West Coast Conference on Formal Linguistics* 20: 484–497.
- Schadeberg, Thilo C.
 2003 Derivation. In *The Bantu languages*, Derek Nurse and Gérard Philippson (eds.), 71–89. London: Routledge.
- Scott, David and Alexander Hetherwick
 1957 *Dictionary of the Nyanja language*. London: United Society for Christian Literature.
- Smith, Neil V.
 1971 *The acquisition of phonology*. Cambridge, etc.: Cambridge University Press.
- Stathopoulos, E. T. and G. Weismer
 1983 Closure duration of stop consonants. *Journal of phonetics* 11 (44): 395–400.
- Teil-D'Autrey, Gisèle
 2002 *Lexiques proto-bantous: étude des cooccurrences segmentales et supra-segmentales*. Doctoral dissertation, Université Lumière-Lyon 2.
- Teil-D'Autrey, Gisèle
 2008 Et si le Proto-Bantou était aussi une langue...avec ses contraintes et ses déséquilibres. *Diachronica* 25: 54–110.
- Vihman, Marilyn May
 2002 Getting started without a system: From phonetics to phonology in bilingual development. *The International Journal of Bilingualism* 6: 239–254.
- Westermann, Dietrich
 1930 *A study of the Ewe language*. London: Oxford University Press.
- Whitehead, John
 1899 *Grammar and dictionary of the Bobangi language*. London: Kegan Paul and Trench.
- Williamson, Kay
 1979 Consonant distribution in Ijo. In *Linguistic and literary studies presented to Archibald Hill Vol. 3*, E. C. Polome and W. Winter (eds.) 341–353. Lisse: Peter de Ridder Press.
- Wolff, Ekkehard and Hilke Meyer-Bahlburg
 1979 Morphologie und Semantik der erweiterten Verbalstämme in der Sprache der Afuzare (Zarek). *Afrika und Übersee* 62: 1–38.

Future tense to express habitual past or present, and past tense to express immediate future

Leena Kelkar-Stephan

1 Introduction

In this paper I discuss two particularities which arise solely due to a special language contact situation between Tamil and French in India. The two particularities – which at this point I will not call “*rara*” – consist of using the past tense forms of certain verbs to express immediate future and future tense to express present or past habitual in the French variety spoken in Pondichery.

The paper is divided into six sections. In Section 2, a brief outline of Tamil – a language of the Dravidian family – will be presented. Since the two languages in the language contact situation under study are typologically different, it is important to summarise some of the typological similarities and differences between the two languages. A typological study is deemed necessary to be able to comprehend the various observations in the case study to be analysed later. Since in this study Modern Tamil is in contact with French, the descriptions below concern only Modern Tamil. In Section 3, certain unexpected polysemy of tense-aspect forms in Tamil will be summarised. Since these polysemies are in no way unique, parallels found in other languages will be presented. Origin and developmental path of these features in Tamil and within the language family will then be summarised based on diachronic studies of Dravidian languages. Section 4 focuses on the brief diachronic study of tense usage in Tamil with the intention to point out the evolution of tenses in the language family. After having prepared the terrain, in Section 5 the case study of Pondicherien Creole French which has been influenced by the Tamil tense-aspect system will be dealt with. Selective examples from a Pondicherien Creole French corpus displaying anomalies will be discussed and an explanation will be presented. Empirical data and their analysis alone do not suffice to claim that the observed phenomena are unique. Hence, in Section 6, the question as to whether the said phenomena can be classified as *rara* will be discussed by comparing them with observations from other

French based language contact situations. In the last section (Section 7), the source of the influence and the direction of change from English to Tamil to French will be speculated.

2 Outline of Tamil and its typological comparability with French

Tamil, the most wide-spread Dravidian language, is today spoken in India, Sri Lanka, Malaysia, Mauritius, Réunion, Thailand and Singapore by approximately 74 million people. It is the administrative language of the Indian state of Tamil Nadu, and has been quite recently (in 2004) recognized as a classical language by the government of India. Tamil belongs to the Southern branch of the Dravidian languages, a family of around twenty-six languages native to the Indian subcontinent. Tamil literature is said to span 3,500 years and the only one among the old languages that has a history of 2,000 or more years of existence.

Typological research and language contact research are concerned with both similarities and differences between languages, usually a large sample of languages. Typology cannot be formulated of whole languages but only of sub-components of languages. Since the phenomena discussed in this paper arise solely due to the contact of two typologically different languages the typological features of French and Tamil are briefly dealt with here focusing on certain frequently discussed characteristics such as word order, negation, pro-drop, morphological and phonological typology.¹ The intention of this discussion is not only to draw attention to the typological differences between the languages involved in the linguistic contact situation in Pondichery but also to facilitate comprehension of the French-Tamil case study dealt with in Section 5.

2.1 Word order

Tamil basically has a SOV order, or more precisely a verb final order. Components preceding the verbal predicate are said to have a free word order since orders like:

$$[SO_1O_2V], [SO_2O_1V], [O_2SO_1V], [O_1SO_2V], [O_1O_2SV], [O_2O_1SV]$$

are possible (where O_1 and O_2 are first and second object). Topics related to word order include pre/post-positions, which probably also implies the

lack or presence of a category of articles and a strong tendency of SOV languages to be the prenominal type (Cf. Greenberg, 1963). The noun phrase in Tamil can have only prenominal modifiers, adjectives are placed before the noun, even relative clauses and sentential complements occur pre-nominally. Adverbial phrases (of time: *nālaikku pōvēṇ* ‘tomorrow I will go’, of place: *vīṭṭukku pōvēṇ* ‘to house I will go’ etc.) precede the verb in Tamil and lack a separate category of articles.

The general word order in modern French – as opposed to Tamil – is a Subject-Predicate-Object (direct object, except personal pronouns *me, le, les*, etc., follow the predicate), however variations in this order are possible for stylistic or rhetorical reasons.² From the point of view of the syntax, this SVO order of French has to be strictly followed due to the reduced morphology which is a general property of the French system (Ineichen 1999: 117). French is a prepositional language. The noun phrase in French consists of a head noun, and optionally, prenominal modifiers such as determiners, etc. and postnominal modifiers such as prepositional phrases, relative clauses, sentential complement etc. Adjectives can be pre- or post-nominal modifiers. The locative, direct object or indirect object markers are pre-positional in French grammar. Adverbial phrases (of time: *à 2 heures* ‘at two o’clock’, of place: *à la maison* ‘at the house / at home’ etc.) may either precede or follow the verb in French.

2.2 Negation

Negation strategies in Tamil and French differ significantly. Tamil possesses three ways of expressing negation, namely: (1) morphologically by a negative verbal suffix realised by two allomorphs *-ā, -āt*, e. g. *pas inṅē var-ā-tu* [bus here come-NEG] ‘The bus does not come here’, (2) lexically by a negative verb, *il* ‘be not’ which marks a locative, existential and copula function, and *al* ‘be not’ marks a copula function alone, as in *kumār vīṭṭ-il il.l-ai* [kumar house-LOC be not] ‘Kumar is not at home’ and *kumār vakkīl il.l-ai / al.l-a* [kumar lawyer be not] ‘Kumar is not a lawyer’ respectively, and (3) by a negative auxiliary verb *il* which occurs after a main verb in the infinitive, *ravi nērru var-a.v-il.l-ai* [ravi today come not] ‘Ravi will not come today’ (Lehmann 1993: 228-231). For a more extensive discussion on negation and negators in Tamil and in the Dravidian language family refer to Pilot-Raichoor (this volume) and Miestamo (2010).

Negation in French has been linked with word order changes since there is said to be a gradual shift from pre-verbal to post-verbal negation marking. Originally *non* and then *ne(n)* were pre-verbal negators, then *ne* placed pre-verbal and reinforced by *pas* (other possible post-verbal negation particles being *plus*, *rien*, *personne*, *que*, *guère*, *point*, etc.) post-verbally. However in modern non-standard varieties the *ne* of the bipartite negator is often deleted and post-verbal *pas* (and the other post-verbal negation particles) remains as the only main verb-phrase negator (Posner 1997: 369–370).

2.3 Pro-drop

Some questions that arise related to the word order of a language are: Is it an adposition/complement (*prepositional*) or a complement/adposition (*postpositional*) language? Is it a null-subject (*pro-drop*) language or not?

Tamil is a postpositional and a pro-drop language. All postpositions in Tamil are formally inflected or uninflected noun forms or non-finite verb forms. Postpositions can occur as bound or free forms. The pro-drop character of the language makes it possible to have subjectless sentences and verbless clauses, both in the written and spoken language.

French is a prepositional language. Standard European written French³ is a non-pro-drop language. In other words: due to unstressed verb endings subject pronouns are obligatory,⁴ hence a subjectless sentence cannot exist. In impersonals, where no logical subject is involved, a dummy or expletive pronoun *il* is used as in a ‘weather-verb’ *il pleut* ‘It’s raining’, and constructions of a main clause verb without a subject, as in **mangent* (*ils mangent* ‘they eat’) are grammatically unacceptable. Contrary to the norms of the written language, colloquial spoken European French allows for a certain number of subjectless sentences. Some of the most frequent cases of null-subject are the dropping of *il* in (*il*) *faut*, *paraît*, *y a*, *s’agit de*, *suffit*, *vaut mieux*.⁵ *Il* can sometimes be replaced by *ça*: *ça semble* ‘it seems’ (cf. Gadet 1989, 1992). A characteristic feature of the variety of Canadian French spoken by less-educated people is its null-subject and pro-drop character (Dahmen 1995: 232–233; Neumann-Holzschuh 2000: 263–264). Neumann-Holzschuh observes that the unstressed subject pronoun is/can be dropped for all persons; it is however more frequently deleted in the first person singular. For example *Moi ... moi melève le matin là* ‘I got up the other morning’ [my translation]. She adds that the subject pronoun can be deleted even in subordinate clauses with the precondition that the subject in the main clause is the

same, e. g. *J'aime un ... un hiver avec eh ... de la neige, de la belle neige qui tombe et pis ... quand est l'a, qu'a reste* 'I like a ... a winter ... with snow, with lovely snow that falls and then ... when it is there and stays' [my translation].

2.4 Morphological typology

Tamil morphology is agglutinating, i. e. morphs are 'stuck on' or agglutinated in a sequence. Agglutinating morphology in Tamil always involves suffixation, i. e. morphs are added in a sequence as suffixes after a stem. Tamil makes productive use of suffixes which include: case markers, postpositions (bound and free forms), tense suffixes-person endings in verbs, emphatic-interrogative markers, adjectival-adverbial suffixes, etc. Tamil does not have expressive suffixes like diminutive or augmentative. Section 3 concentrates solely on the specific case of the Tamil tense-aspect system.

French – in the passage from Latin to Middle French to Modern French – has been widely claimed to have gradually changed from a synthetic into an analytic (isolating) language.⁶ Modern French uses syntactic, rather than morphological, devices to show relationships between free-standing words, and it relies on obligatory subject pronouns to mark verb positions.⁷ That is, verb person are principally marked by proclitic pronouns rather than by inflectional endings since these endings are not stressed in the spoken language. Most verbs have zero markers for the first, second, third person singular and third person plural in the present indicative whereas the first and second person plural can be differentiated by [ɔ̃] and [e] endings or sometimes by difference of stem. Loss of unstressed final vowels and consonants has led to homonymy of all the stem-stressed forms and obligatory use of subject pronouns (cf. Posner 1996: 37, 45–49, Posner 1997: 294).

2.5 Phonological typology

In phonology, a central concern has been phoneme inventories. It appears that there are clear hierarchies in the distribution of particular vowels and consonants across languages. Stress, tone, and syllabicity are also important points in phonological typology.

Tamil is written with its own, alpha-syllabic Script which suggests similarities to both alphabetic and phonetic systems. One of the characteristic fea-

tures of Tamil consonants is retroflexion, i. e. voiced sublamino-palatals such as [ɻ] and [ʒ]. Tamil is a syllable-timed language. The canonical syllable type is (C)(C)V(C)(C), where V is either a short or a long vowel. V, VC and VCC occur in word initial positions only (Asher 1982: 228).

French: The complexity of the relation between the (ortho)graphic and phonic system in French is all too well known. The phonic code distinctly contrasts with the graphic code. The characteristic features of Modern French phonology are the schwa (*e muet*), the /h/ (*h aspiré* and *muet*) and liaison. French is a syllable-timed language. Each syllable contains one and only one vowel and a consonant alone cannot constitute a syllable. CV combination in syllables are the most frequent, other frequent open syllables are V, CCV, and CCCV; among the closed syllable structure CVC is the most frequent. Modern French shows a strong tendency towards monosyllabism as compared to Old French syllable structure $C_1V_1C_2(C_3)V_2C_4$, the loss of C_4 and V_2 has resulted in the generation of a new final C_2/C_3 (Campbell 2000: 574).

3 Tamil tense-aspect system

As discussed above, Tamil basically has a SOV order, is agglutinating, post-positional and suffixal. In this section only the discussion of the tense-aspect system will be dealt with.

Interpretation of tenses – mainly past and future – in Tamil is a rather complex matter since Modern Tamil grammar disposes only of the past, present and future tense forms. Present progressive, immediate future, past continuous, etc. need to be expressed using these available tenses. Hence, depending on the context the hearer has to infer the intended interpretation of tense to refer to various times. However, there is not always a correlation between the morphological tense form and the corresponding time reference (following examples and explanations are based on observations by Lehmann 1993: 65–68). In the next two sub-sections, polysemy of past and future tense forms in Modern Tamil are reviewed.

3.1 Past tense to express immediate future

The past tense form of the auxiliary verb *viṭu* ‘leave’ combined with the verbal participle form of the main verb *vā* ‘come’ is idiomatically used to express immediate future as in example (1):

- (1) a. *ṣānti iṅkē vā*
 shanti here come:IMP.S
 'Shanti come here'
 b. *nāṇ va-ntu- viṭ.ṭ-ēṇ*
 I come-VBP- leave-PST-1S
 'I'm coming.'

Further, in a situation of threatening or warning, a conditional clause followed by a verb in the past tense is used to refer to the future time, as in example (2):

- (2) *nī it-ai.t tot.ṭ-āl ce-tt-āy*
 you this-ACC touch-COND die-PST-2S
 'If you touch this, you will die.'

The third case involves the auxiliary verb *kīṇi* 'tear' expressing the speaker's disbelief. Thus, in response to a statement with future time reference, the addressee may use a past tense form of *kīṇi* which however has future reference, as in example (3):

- (3) a. *kumār nālai-kku.p paṇam koṭu-pp-āṇ*
 kumar tomorrow-DAT money give-FUT-3S.M
 'Tomorrow Kumar will give [me/us, etc.] some money.'
 b. *avaṇ kiṇ-tt-āṇ*
 he tear-PST-3S.M
 'He won't do it (I don't believe it).'

If this is not peculiar enough let us turn to the future in Tamil.

3.2 Future tense forms

In Tamil, the future tense can express both a temporal and an aspectual meaning so that a sentence like (4):

- (4) *nāṇ kālai.y-il kuḷi-pp-ēṇ*
 I morning-in bath-FUT-1S

may be interpreted as 'I will have a bath in the morning' or 'I (usually) have a bath in the morning' (Pillai 1986: 146) and a sentence like (5):

- (5) *kumār aṭikkati ciṇimā.v-ukku.p pō-v-āṇ*
 Kumar often cinema-DAT go-FUT-3SM

may be interpreted as ‘Kumar often goes to the cinema’ or ‘Kumar will frequently [in future] go to the cinema’ (Lehmann 1993: 66). A similar example is that of the Lezgian verbal suffix *-da* which can express both a temporal and an aspectual meaning, so that a sentence like *za ič-er qaču-da* [I:ERG apple-PL take-FUT] may be interpreted as ‘I will buy apples’ or ‘I (usually) buy apples’ (Haspelmath 1998: 32). Haspelmath further quotes an example from Yavaşa where a similar verb form called Aorist in Turkish may have both these meanings. Thus, the sentence *Ahmet üç ders okut-ur* [Ahmet three lesson teach-AOR] may mean ‘Ahmet will teach three lessons’ or ‘Ahmet (usually) teaches three lessons’.

The second function of the future tense in Tamil is a non-compositional case where the future tense expresses past habitual. Let us consider a Tamil sentence like (6):

- (6) *cinna vayat-il nān pamparam vīlaiyātu-v-ēn*
 small age-LOC I top play-FUT-1S
 ‘When I was young, I used to play with tops.’

Thus, the future tense in Tamil clearly seems to convey habitual meaning both present and past. This usage of the future tense forms to express past habitual meaning is, however, not uncommon as studies on the peculiarities of the Mohawk verbal system which has distinct morphology consisting of the habitual morpheme *-ha*’ or *-s*, followed by the past morpheme *-kwe*’ to express past habitual,⁸ and the related case of Udmurt past habitual which is formed by combining the future tense with the auxiliary *val* ‘was’ as discussed by Haspelmath (1998: 32) show. Thus, while *myn-o* means ‘I will go’, *myn-o val* means ‘I used to go’ which implies that although the future tense in Udmurt does not have a habitual reading by itself, it is used in a larger construction where it clearly seems to convey habitual meaning. Similar cases occur in Inuit, where the future marker *ssa* signals past habitual when it is used in a past narrative context, and in Margi where the progressive particle *əvər* indicates habitual action in the past (Bybee et al. 1994: 158).

A cross-linguistically rather unusual usage of the future tense concerns forms having non-future meaning in the special context of generics. Thus, the future in Tamil is also used with a generic meaning in a statement like in (7):

- (7) *mātu pil tin-um*
 cow grass eat-FUT-3SN
 ‘Cows eat grass.’

A similar case is that of Udmurt where Future is used in proverbs with a generic meaning and the Lezgian Future is also used in stage directions as in *perde agal že-da* [curtain fall become-FUT] ‘the curtain falls’ (cf. Haspelmath 1998: 32). Likewise, Bybee et al. (1994: 35) notice the use of future in generic expressions in English such as *Boys will be boys*. Thus, it seems that “most if not all, languages have another, more usual way of expressing the generic sentence” (Bybee et al. 1994: 35).

After having looked at some peculiarities of the tense-aspect system in Modern Tamil, it is of interest to examine the source and the development within Tamil and within the language family. I, thus, proceed with a synopsis of a diachronic study of the tense-aspect system in the Dravidian language family and in Tamil, in specific.

4 Diachronic study of tense usage in Tamil

Basically there are two tenses in Dravidian, past and non-past. Non-past includes the habitual (present / aorist / indefinite / generic) and future. Tamil and some other languages of the family have also developed a separate present tense (cf. Krishnamurti 2003; and for an extensive study of the development of the present tense in Tamil and the etymology of the present tense in Tamil refer to Steever 1993: 167–193). As the present tense paradigms evolved “both the inherited non-past paradigm and the non-past progressive paradigm were consistent with present time reference” (Steever 1993: 168). It is said that when they became restricted in South Dravidian (i. e. Tamil, Malayalam, Iruḷa, Kota, Toda, Badaga, Koḍagu, Kannada, and Tulu)

“[...] newly innovated non-past tense paradigms arose to assume most of the functions of these inherited non-past paradigms (e. g. present time reference, future time reference, habitual action, past time reference in historical narrative, and the “tenseless” present). This is what is now called the present tense paradigm” (Steever 1993: 168).

It, thus, can be asserted that since this newly innovated non-past took over the function of expressing (besides other functions) present time reference, future time reference, habitual action, past time reference in historical narrative, and the “tenseless” present, it is not surprising that generic statements can be expressed both with the future as well as a more usual way with present tense forms of verbs. Further, since the present-future (or otherwise non-past) semantically covers both the present and the future, i. e. it denotes ac-

tions which are or going to be realized (cf. Andronov 2003: 196), it also explains the use of future tense forms to express future time meaning as well as habitual actions in the present, as discussed in Section 3.2.

Diachronic studies of Tamil demonstrate that despite grammatical individuality of each stage of the language Tamil shows grammatical continuity from the old to the modern language (cf. Lehmann 1998: 75–99). A look at the historical morphology of older Tamil, where tense marking was not so common, reveals that especially for stative verbs emphasis was laid on marking habitual states (Schiffman 1982: 104–116). Andronov (2001: 100–106), additionally, cites several instances of discrepancies in the course of development of tense suffixes, and the meaning and function of emerging allomorphs in the Tamil literature. The approximate period at which the use of the future tense form started being used to signal past habitual in Tamil seems to be obscure. All that can be stated with certainty is that this usage of tense is not unusual both in written and colloquial Tamil.

A look at parallels from other languages of the family reveals that future tense forms in Malayalam can also refer to habitual actions in the past; future tense forms in Kolami refer to habitual actions as well as to events with future time reference. In Kolami, the future tense form is not used for past habitual since a separate durative form expresses continuous action with past time reference or in a habitual sense (Subrahmanyam 1998: 316). In Koṇḍa, the durative paradigm refers to an action in progress in the past or non-past time, and the non-past refers to a habitual or future action (Krishnamurti et al. 1998: 256–257). Thus, although not all languages in the linguistic area may share the feature, it is sufficiently widespread to be considered a characteristic of the Dravidian linguistic area.

As for the explanation of the use of past tense forms in certain cases for the future time meaning, an investigation of traditional Tamil grammar makes clear that additional uses of tense forms, as remarked by Tamilists like Steever (1993: 191), were recognised in Traditional Tamil grammar (cf. *Tolkāppiyam*,⁹ *Collatikāram*¹⁰ 241, 242). Thus, uses of the past tense form to signal present or future time reference in certain circumstances have been licensed. Andronov (2001: 102) has also observed several instances in Tamil literature where the past tense of verb *nil-* ‘to stop, stand’ was used to express the meaning of the present tense in periphrasis with verbal participles. These “sanctioned deviations in the use of tense/time” and “interchange of tense/time” have been extensively discussed in the Tamil grammatical literature.

The sort of movement future of certain auxiliary verbs, discussed in example (1), repeated here as (8) for convenience, expresses the notion of the agent being “on a path moving towards a goal” or that “the agent is already on the path and the movement is in progress” (cf. Bybee et al. 1994: 268).

- (8) a. *ṣānti iṅkē vā*
 shanti here come:IMP.S
 ‘Shanti come here’
 b. *nāṇ va-ntu- viṭ.ṭ-ēṇ*
 I come-VBP-leave-PST-1S
 ‘I’m coming.’

This Tamil statement may thus be interpreted as ‘I have left my original position and am now on the path to come to the goal in immediate future’. Since the action of leaving place A has to take place before the action of arriving at place B the verb ‘leave’ takes the past tense marker.

Furthermore, just as some languages use the future to express statements that amount to prediction, though not necessarily predictions about future time, Tamil seems to use the past tense in statements that amount to prediction, i. e. events whose “predictability” in future time is high. These predictions, that are based on past experiences and that apply to states of affairs that exist in the past, present and future, take the shape of generalized predictions (cf. Bybee et al. 1987: 120). These so called “general truths” based on past experiences seem in Tamil to logically demand past tense verbal forms.

After having so far created the foundation, I now turn to the case study of a variety of French spoken in Pondichery, on the south-east coast of India. Native French speakers in Pondichery reflect the influence of the Tamil tense-aspect system on their French tense-aspect system. The paper, thus, continues with an introduction of the French speaking Creole community and then turns to selected examples from a corpus of empirical data displaying these anomalies.

5 Case study: Pondicherien Creole French

Historical events in Pondichery on the south-east coast of India from about 1690 till 1954, when Pondichery gained independence from the French regime, explain the presence of French culture and account for the fact that Pondichery is the only French speaking conclave in the Indian peninsular till today.

5.1 The Pondichery Creole community

Descendants of mixed Indian and French parentage, called Creoles, are the remaining bastion of French native speakers in this multilingual society. Note, however, that the term ‘Creole’ strictly refers to descendants of French and Creoles who call themselves ‘Creoles’ till today and not to their language. ‘Creoles’, adding up to less than 500 (in the year 2001), are overseas French nationals bilingual in Tamil and French. French is however, mainly used with other French speakers and for administrative purposes. Tamil is used with the maids, the vendors in the market, or the rickshaw drivers. Informal interactions with friends or family take place in both Tamil and French and speakers usually code-switch. When they code-switch, the use of either language is not a random mixing.

The predominant use of Tamil is observed in narrating daily events, the scene in the market, talking about a certain television show, or the latest gossip. French, on the other hand, is significantly more frequently used to narrate events from the past — their childhood, their school, the nuns at school, etc. or official events — interaction at the consulate registration office, etc. This bilinguality brings about code switching of French in a Tamil conversation or Tamil in a French conversation. Both French and Tamil spoken in Pondichery seem to have influenced and still influence each other giving rise to varieties which I term Pondicherien Creole French and Pondicherien Creole Tamil.

Tamil and other regional languages have extensively influenced the French used by Creoles at the phonological, morpho-syntactic and semantic levels.¹¹ However, here, I concentrate only on the affect of Tamil on the use of past tense to refer to the future time as also the use of future tense in Pondicherien Creole French.

5.2 Past tense to refer to the future time

As discussed in Section 3.1, in Tamil the past tense forms of certain auxiliary verbs combined with verbal participle form of the main verb *vā* ‘come’ are idiomatically used to express immediate future. This unexpected polysemy of past tense forms which is not too common, however, does not seem to have influenced past tense usage in Pondicherien Creole French. Interestingly, as in Tamil in a situation of warning, a conditional clause followed by a verb in the past tense is used to refer to the future time. In the following sentence

(9),¹² some speakers of the Pondicherien Creole French community warn the author against wearing a “slightly expensive” watch when she is on buses or trains in India. An elliptical sentence, which, given the context, has to be a conditional clause, is followed by a verb in the past tense with future meaning.

- (9) PCF: *mets pas ça. tu **perdais** ça.*
 StdF: *Ne la [la montre] porte pas. [Si tu la portes], tu risques de la perdre.*
 ‘Don’t wear it [the watch]. [It’s quite likely] you loose it.’

5.3 Effect of Tamil future tense on habitualis in Pondicherien Creole French

As discussed in Section 3.2, in Tamil, the future tense can express both, a temporal and an aspectual meaning such that the future tense in Tamil not only expresses both a future and present habitual meaning, but it also expresses past habitual. These anomalies in Tamil which clearly seem to convey habitual meaning have conceivably influenced the tense-aspect system of Pondicherien Creole French due to the intensive language contact between the two languages.

Thus, in example (10) below, the future tense expresses present habitual meaning. In the absence of the temporal marker, the sentence may also be interpreted as expressing future meaning ‘She will go to church’.

- (10) PCF: *Tous les matins elle **ira** à l’église.*
 StdF: *Tous les matins elle **va** à l’église.*
 ‘She goes to church every morning.’

The future tense is also used for ongoing situations (mainly stative verbs) as in examples (11) and (12):

- (11) PCF: *elle **pourra** pas travailler avec ces jeunes enfants qu’elle a*
 StdF: *elle **ne peut pas** travailler à cause de ses [jeunes] enfants*
 ‘She cannot work due to her little children’
 (12) PCF: *il **sera** comme ça, de travers.*
 StdF: *Il **est** comme ça, de travers.*
 ‘He’s like that, askew.’

In the following examples (13), (14), and (15) the future is used to express habitual actions in the past with a narrative character.

- (13) PCF: *n'importe qui **aura** ciseau, je **prendrai** des longues tresses comme ça, **sera** lourdes, je **garderai**, je prendrai le ciseau et **cassera** en deux, mère me **donnera** jamais le ciseau*
 StdF: *À chaque fois que quelqu'un **avait** des ciseaux je les **prenais** [...] et je **coupais** les longues tresses de mes copines. Résultat: la Mère ne me **donnait** jamais de ciseaux.*
 'Whenever somebody had a pair of scissors I used to take them and cut the long plaits of my friends [with the consequence that] mother never trusted me with a pair of scissors.'
- (14) PCF: *les copines **diront** vous **êtes à tailler** toujours vous **êtes taillé** tous les plantes les tiges*
 StdF: *Mes copines me **disaient**: "tu es toujours **en train de couper** quelque chose. Tu **as coupé** toutes les plantes, les tiges ..."*
 'My friends used to tell me you are constantly cutting something. You have cut all the plants, the stalks ...'
- (15) PCF: *après en la nuit on **mettra** pour manger le riz, riz ...*
 StdF: *la nuit on nous **servait** le riz pour que nous mangions ...*
 'then at night we were [usually] served rice to eat ...'

Further, in example (16) the use of future is even extended to descriptions.

- (16) PCF: *il y **aura** une grotte devant la grotte il faut chanter pour la Vierge*
 StdF: *il y **avait** une grotte. On a dû chanter [quelques chansons à la gloire de la Vierge] devant.*
 'There was a grotto. We had to sing [in praise of] the Virgin Mary in front of it.'

6 Does this explanation suffice to call these phenomena *rara*?

Owing to the fact that the phenomenon discussed in this paper arises mainly due to a language contact situation – and more precisely a French based language contact – it would now be necessary to find out if the future is expressed in the same or in a comparably similar manner in other French based language contact situations, namely, in some of the French Creole languages. Most French Creoles typically dispose of only about one invariable verb form. They are nevertheless able to effectively express tense, mood and aspect due to various preverbal particles. Future is expressed by preverbal markers, namely:

1. *ka*, *ké*, and seldom \emptyset in Guyana Creole,
2. *ké* with variants (*kèy*, *kay*) in Martinique Creole,
3. *pu* along with *a*, *sava*, *sa*, *va* in Réunion Creole,
4. *a*, *ava* and *pu* in Mauritius Creole, and
5. *a*, *av* and *ap* in Haiti Creole.

These basic particles in combination with other preverbal particles are used to also express future in the past, present conditional and past conditional (cf. Hazaël-Massieux 1990, 1992; Conwell and Juilland 1963; Baker 1972; Chaudenson 1974, 1979). None of these creole languages use the future tense to express past habitual or present, and past tense to express immediate future as is the case in Pondicherien Creole French.

Furthermore, in another French based contact situation namely that of Québécois French, the tendency to frequently replace *futur simple* with the *futur périphrastique* has been researched by Neumann-Holzschuh (2000: 254–259). Verbs such as *être*, *avoir*, *pouvoir*, *falloir*, have also been noticed to be very frequently used in *futur périphrastique* constructions (i.e. *va être*, *va avoir*, *va falloir*) such as *je vas pouvoir manger* ['I will be able to eat'; my translation] which is very seldom used in spoken French in France. Interestingly, usage of the *futur périphrastique* to express habitual actions has also been noted in Québécois French, e.g. *je vas faire telle et telle chose dans ma semaine* ['I do one thing or another during my week'; my translation] (Neumann-Holzschuh 2000: 257, example 12). Chaudenson (1978: 554), Conwell and Juilland (1963: 156, 168), Renaud (1979: 429), and others have also noted the frequent use of the *futur périphrastique* to express *futur simple* in quite a number of Creoles and in overseas French varieties.

Contrary to tendencies noted in the Indian Ocean French Creole languages and in Canadian French discussed above, Pondicherien Creole French has *no* preverbal markers to mark future. Instead, Standard French future forms with appropriate verbal ending are used. However, this tense form is not used to express future alone but can be used to express events past, present and future. This highly marked situation, not noticed in any other Creole grammar, has two explanations. First, this usage of the future is very likely a result of the contact with and the influence of Tamil. As discussed in Section 4, Tamil grammar disposes only of the past, present and future tense forms. Present continuous, immediate future, past continuous, etc. need to be expressed using these available tenses. Depending on the context the tenses are interpreted

to refer to various times. Past tense not only refers to past events but also to the immediate future, for future time in situations of threatening or warning, and also for future time often to express the speaker's disbelief. Future tense expresses rather various conditions than referring simply to future time. A rather striking feature of the future tense is that it is used to express habitual or repeated actions or events in the past or present. The second assumption is that since Pondicherien Creole French speakers' wish to simplify tasks relating to storage and production of tenses, they use the relatively simple future forms for most narrations past or present habitual, and future. As stated by Plank in the abstract of his paper presented at the *Rara & Rarissima Conference 2006* on "*Third Person Plural in Distancing Pronominal Address for Individuals: Rare by Chance or Necessity?*" it seems that in this case frequency and rarity are directly connected with structural complexity; i. e. since the past tense system in French is more complex (five forms: *passé composé*, *imparfait*, *plus-que-parfait*, *passé simple*, *passé antérieur*) it is seemingly rarer than the less complex future tense system (two forms: *futur simple*, *futur antérieur*).

The phenomena discussed in this paper arise solely due to a specific French based language contact situation. Nonetheless, since these highly marked phenomena have not been noted in any of the other French based creoles, pidgins or language contact grammars they deserve to be classified as *rara*.

7 Direction of change English → Tamil → French?

In conclusion, I would like to discuss the direction of development thereby raising the question as to the role English might have played to give rise to these phenomena. My hypothesis is that Tamil is at the source of the anomalies noticed in Pondicherien Creole French, without English being the intermediary. In a multilingual situation as in India where several languages have been in intensive and long term contact with each other, it is natural that developments in languages include borrowings from other specific languages besides internal structural changes and other wide-spread tendencies of linguistic change. In this particular case, I rule out English influencing Tamil which in turn might be said to have stimulated the change in French although an analogous case of the English auxiliary *will* giving rise to polysemous expressions with future tense meaning as well as past habitual meaning exists, e. g. *would* in a sentence like *He would tell me the same thing over and over*

again (Bybee et al. 1994: 156). The reason why the change could not have happened from English to French directly is that there has not been a widespread bilingualism between French and English in the south of India. Most Tamil speakers in Tamil Nadu have English as the second language, and most bilingual speakers in Pondichery are either Tamil-English or Tamil-French bilinguals. English-French bilingualism in the absence of Tamil is practically non-existent. The direction of change evidently has to be from Dravidian to French since English-French bilinguals in French-English language contact situations elsewhere – in Canada for instance – do not display this tendency, nor do other French-English bilinguals (usually foreign nationals) who have resided for a long time in Pondichery and have not internalized Tamil to the extent that it starts affecting / influencing the tense-aspect system of the languages they speak. It has been observed that if a speech community is successful, in surviving and spreading, or if it is dominant it is able to pass on particularities of its own speech to other speech communities it is in contact with. If the dominance of Tamil or English or French, for that matter, in Pondichery is to be questioned, Tamil turns out to be the most dominant in several domains in most speech communities.

If I now turn to work out the time at which this development might have happened in Pondicherien Creole French I am confronted with the next difficulty. A diachronic study of these changes in Pondicherien Creole French is problematical since no records (written or oral) of this period testifying to the variety of French written or spoken during the French regime were made; and if they were, not much has survived. From the onset of the French trading activities in India till the end of the French regime in 1954, the history, political situation, missionary activities, and almost all aspects of life and customs of the locals was extensively documented. A detailed research at the *Centre des Archives d'Outre-Mer* in Aix-en-Provence which houses almost all of the salvaged material provides no distinct information. These documentations, most of them done in French, seem to have paid less attention to the linguistic development of French citizens and their descendants in Pondichery under the influence of indigenous languages. In conclusion, I have to say that since there is no oral transmission of French spoken during the French regime no more than a synchronic study of language use at the start of the twenty-first century Pondicherien Creole French is possible.

Abbreviations and symbols

-x- = letter(s) between hyphens are used for expository purposes to indicate morpheme boundaries; . = a dot used in Tamil data indicates that the following phoneme or phoneme cluster is the result of the operation of sandhi rules (morpho-phonological rules); 1 = first person; 2 = second person; 3 = third person; ACC = accusative case; AOR = aorist; COND = conditional mood; DAT = dative case; ERG = ergative; F = feminine; FUT = future; IMP = imperative; LOC = locative case; M = masculine; N = neuter; PCF = Pondicherien Creole French; PST = past tense; S = singular; StdF = Standard French (spoken and/or literary); VBP = verbal participle suffix

Notes

1. For reasons of space/time and my inability to do complete justice to this vast field of linguistics, only an infinitesimal part of the extensive discussions and literature can be dealt with here. The reader is referred to Comrie (2001); Corblin (1987); Cornish (1986, (1999); Fuchs (1997); Gadet (1989, 1992); Haase (2000); Ineichen (1999); Lehmann (1986); Hall (1984); Harris (1978); Posner (1996, 1997); Schwegler (1990); Schwegler et al. (1998); Wilmet (1997) for French; and Albert (1966); Annamalai et al. (1998); Asher (1982); Bright (1998); Emeneau (1970, 1994); Emeneau et al. (1962); Lehmann (1993, 1998); Mcalpin (1974); Paramasivam (1979); Schiffman (1999); Steever (1998); Zvelebil (1969, 1990) for Tamil and several others.
2. There has been extensive discussion and debate for some decades among linguists as to the word order in, say, French interrogatives, inversion, subject position in simple and in complex phrases, relatives, etc. There is however, no consensus as to *the* canonical word order. Since Subject-Predicate-Object is generally the most accepted word order I have stated it so. For details of the on-going discussion of French word order refer Fuchs 1997, the contributors and the bibliography therein.
3. This specific designation is necessary here to differentiate spoken European French (mainly French but also Belge and Swiss French) from written European French which in turn needs to be differentiated from written and spoken Canadian French.
4. For instance the verb *faire* 'do' in the 1, 2, 3 person singular is pronounced as [fɛ] due to the unstressed verb endings, likewise in most verbs of the first group '-er' the first, second, third person singular and the third person plural are all pronounced the same, e. g. [ariv] for *arriv(e)/(es)/(e)/(ent)*. To be able to distinguish the person and number (3.PL can – in verbs with initial vowel – be distinguished due to liaison, e. g. *ils arrivent* [ilzariv]) subject pronouns are necessary. Dropping of the subject pronoun might result in incomprehensible utterances. According to Schwegler (1990: 87) changes in the evolution of Modern French from Latin down to Old French to Middle French resulted in slow erasure of phonological oppositions which had distinguished the verbal endings in the previous stages, due to either phonological reduction or analogical development. These changes radically reduced the efficacy of personal suffixes and only the 1.PL and 2.PL continued to distinguish between tense and person/number. In the absence of any overt marking the preverbal markers became the only person/number indicators in many cases and hence became obligatory.

5. Lambrecht (1981: 27) states that “these non-referential subjects tend to be dropped systematically, though not obligatorily”.
6. This view is however not unanimously accepted among linguists some of who namely Bally and Ashby, and others cited by Schwegler (1990: 76) claim that Modern French is a synthetic language whereas the others (Guiraud and Jespersen and others also cited by Schwegler 1990: 75) consider it to have evolved from a synthetic into an analytic language. Schwegler analyses the reasons why linguists have widely accepted the view that Modern French has changed into an analytic language and states that

“analysis and synthesis are simply a consequence (rather than a cause) of language change [and that] every innovation ultimately affects the direction a speech unit takes on the analytic/synthetic axis, and that the frequently mentioned drift of IE [Indo-European] languages towards greater analyticity merely reflect a statistical trend whose unidirectional momentum had been exaggerated as a result of a widespread failure to recognise signs of synthesis” (Schwegler 1990: xvi).

This implies that French has been and still is a synthetic language (Schwegler 1990: 149).

7. See note 4 for phonological changes in the evolution of French which resulted in overt markings only in the 1.PL and 2.PL. As a result personal subject pronouns in Modern French are obligatory, bound forms severely limited in their syntactic mobility, unable to receive stress and have to be integrated with the verb. (Schwegler 1990: 87–92).
8. The following example shows the peculiarities of the Mohawk verbal system to express past habitual.

Kikv okára ne tsi ni-ye-ye-hah-kwe’ ak-sótha nónv
 this story is how PART-FSS-do-HAB-PAST my-grand.mother when
v-ye-nvhst-ohare-’ tánu’ v-ye-the’ser-uni-’ ohvtu tsi niyóre
 FUT-FSS-corn-clean-PUNC and FUT-FSS-flour-make-PUNC before
kana’taro-k-húwe v-ye-na’tar-isa-’.
 bread-Ø-REAL FUT-FSS-bread-FINISH-PUNC
 ‘This is the story of what my grandmother used to do when she would
 clean the corn and make the flour before she would make (traditional)
 corn bread.’ (Baker et al. 1997: 217, example 6)

9. The *Tolkāppiyam* is a work on the grammar of the Tamil language and the earliest known extant work of Tamil literature. It is the earliest long text in Old Tamil which is an early work on Tamil grammar and poetics, whose oldest layers could be as old as the first century BC.
10. *Tolkāppiyam* is divided into three parts, the third of which is the *Collatikāram*, which treats “words”, i. e. it treats parts of speech, morphology, reference and some aspects of syntax.
11. Cf. Kelkar-Stephan (2005) for an extensive analysis of this Pondicherien Creole French variety.
12. All examples, if not specifically mentioned, are from the larger corpus of conversation, part of which was analysed and commented in the research study on the language contact situation in Pondichery; cf. Kelkar-Stephan (2005). All the examples are verbatim as they were expressed and recorded.

References

- Albert, Paul
1966 Remarques sur la phonologie du Tamoul. *La Linguistique* 2: 93–110.
- Andronov, Michail S.
2001 *Dravidian historical linguistics*. München: LINCOM Europa.
- Andronov, Michail S.
2003 *A comparative grammar of the Dravidian languages*. Wiesbaden: Harrassowitz.
- Annamalai, E. and Sanford B. Steever
1998 Modern Tamil. In *The Dravidian Languages*, Sanford B. Steever (ed.), 100–128. New York: Routledge.
- Asher, Ronald E.
1982 *Tamil* (= Lingua Descriptive Studies; 7). Amsterdam: North Holland.
- Baker, Mark and Lisa Travis
1997 Mood as Verbal Definiteness in a ‘Tenseless’ Language’. *Natural Language Semantics* 5.3: 213–270.
- Baker, Philip
1972 *Kreol: a description of Mauritian Creole*. London: Hurst.
- Bright, William
1998 The Dravidian scripts. In *The Dravidian Languages*, Steever Sanford (ed.), 40–71. New York: Routledge.
- Bybee, Joan L. and William Pagliuca
1987 The evolution of future meaning. In *Papers from the VIIth International Conference on Historical Linguistics*, A. G. Ramat, O. Darruba, and G. Bernini (eds.), 109–122. Amsterdam: Benjamins.
- Bybee, Joan L., Revere Dale Perkins, and William Pagliuca
1994 *The evolution of grammar: tense, aspect, and modality in the languages of the world*. Chicago etc.: University of Chicago Press.
- Campbell, George L.
2000 *Compendium of the world’s languages*. Second edition. London: Routledge.
- Chaudenson, Robert
1974 *Le lexique du parler créole de La Réunion*, Paris: Champion.
- Chaudenson, Robert
1978 *Les Créoles français*. (= Langue Française; 37) Paris: Larousse.
- Chaudenson, Robert
1979 Le français dans les îles de l’Océan Indien (Mascareignes et Seychelles). In *Le Français hors de France*, Albert Valdman, Robert Chaudenson, and Gabriel Manessy (eds.), 543–617. Paris: Champion.
- Chaudenson, Robert
1992 *Des îles, des hommes, des langues: essai sur la créolisation linguistique et culturelle*. Paris: L’Harmattan.
- Chaudenson, Robert
2001 *Creolization of Language and Culture*. London / New York: Routledge.

- Comrie, Bernard
 2001 Different views of language typology. In *Language Typology and Language Universals / Sprachtypologie und sprachliche Universalien / La typologie des langues et les universaux linguistiques. An International Handbook / Ein internationales Handbuch / Manuel international*, Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, Wolfgang Raible (eds.), 25–39. (= Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science; 20.1). Berlin/New York: Mouton de Gruyter.
- Conwell, Marilyn and Alphonse Juilland
 1963 *Louisiana French grammar*. The Hague: Mouton.
- Corblin, Francis
 1987 *Indéfini, défini et démonstratif*, Geneva: Droz.
- Cornish, Francis
 1986 *Anaphoric Relations in English and French: A Discourse Perspective*. London: Croom Helm.
- Cornish, Francis
 1999 *Anaphora, discourse and Understanding*. New York etc.: Oxford University Press.
- Dahmen Wolfgang, Günter Holtus, Johannes Kramer, Michael Metzeltin, Wolfgang Schweickard, and Otto Winkelmann (eds.)
 1995 *Konvergenz und Divergenz in den romanischen Sprachen*. (= Romanistisches Kolloquium; VIII). Tübingen: Narr.
- Emeneau, Murray Barnson
 1970 *Dravidian comparative phonology: a sketch*. Annamalainagar: Annamalai University.
- Emeneau, Murray Barnson
 1994 *Dravidian studies, selected papers*. Delhi: Motilal Banarsidass.
- Emeneau, Murray Barnson and Thomas Burrow
 1962 *Dravidian borrowings from Indo-Aryan*. Berkeley: University of California Press.
- Emeneau, Murray Barnson and Charles A. Ferguson (eds.)
 1969 *Linguistics in South Asia*. The Hague: Mouton.
- Fuchs, Catherine (ed)
 1997 *La place du sujet en français contemporain*. Belgium: Duculot.
- Gadet, Françoise
 1989 *Le Français Ordinaire*, Paris: Armand Colin Éditeur.
- Gadet, Françoise
 1992 *Le Français Populaire* (= Que sais-je?; 1172). Paris: Presses Universitaires de France.
- Giacalone Ramat, Anna (ed.)
 1987 *Papers from the 7th International Conference on Historical Linguistics*. Amsterdam/Philadelphia: Benjamins.
- Greenberg, Joseph Harold (ed.)
 1963 *Universals of language*. Cambridge, MA: MIT Press.

- Haase, Martin
 2000 Das Französisch als exotische Sprache. In *Linguistica Romanica et Indiana*, Bruno Staib (ed.), 117–130. Tübingen: Narr.
- Hall, Robert A. Jr.
 1984 *Proto Romance Morphology. Comparative Romance Grammar, vol. III.* (= Amsterdam Studies in the Theory and History of Linguistic Science, Series IV: Current Issues in Linguistic Theory; 30). Amsterdam/Philadelphia: Benjamins.
- Harris, Martin
 1978 *The Evolution of French Syntax: a Comparative Approach*. London: Longman.
- Haspelmath, Martin
 1998 The semantic development of old presents: New futures and subjunctives without grammaticalization. *Diachronica* 15.1: 29–62.
- Hazaël-Massieux, Guy
 1990 Le guyanais et les créoles atlantiques à base française. *Études Créoles* XIII.2: 95–112.
- Hazaël-Massieux, Guy
 1992 Genèse des marques de TMA en créole de Guadeloupe. *Linguistique et Slavistique: Mélanges Offerts à Paul Garde, II.*, 643–669. Aix-en-Provence: Université de Provence.
- Hazaël-Massieux, Marie-Christine and Didier De Robillard (eds.)
 1997 *Contacts de langues, contacts de cultures, créolisation*. Paris: L'Harmattan.
- Ineichen, Gustav
 1999 *Typologie und Sprachvergleich im Romanischen*. Heidelberg: Winter.
- Kelkar-Stephan, Leena
 2005 *Bonjour maa: The French–Tamil Language Contact Situation in India*. Aachen: Shaker.
- Krishnamurti, Bhadriraju and Brett A. Benham
 1998 Koṇḍa. In *The Dravidian Languages*, Sanford B. Steever (ed.), 241–269. New York: Routledge.
- Krishnamurti, Bhadriraju
 2003 *The Dravidian languages*. Cambridge etc.: Cambridge University Press.
- Lambrecht, Knud
 1981 *Topic, antitopic and verb agreement in non-standard French*. (= Pragmatics and beyond 2.6). Amsterdam: Benjamins.
- Lehmann, Thomas
 1993 *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Lehmann, Thomas
 1998 Old Tamil. In *The Dravidian Languages*, Sanford B. Steever (ed.), 75–99. New York: Routledge.

- Lehmann, Winfried (ed.)
 1986 *Language Typology 1985. Papers from the Linguistic Typology Symposium Moscow.* (= Amsterdam Studies in the Theory and History of Linguistic Science; 4; Current issues in linguistic theory; 47). Amsterdam/Philadelphia: Benjamins.
- Mcalpin, David W.
 1974 Towards Proto-Elamo-Dravidian. *Language* 50: 89–101.
- Miestamo, Matti
 2010 Negatives without negators. In *Rethinking Universals: How rarities affect linguistic theory*. Jan Wohlgemuth and Michael Cysouw (eds.), 169–194. (= Empirical Approaches to Language Typology; 45). Berlin/New York: Mouton de Gruyter.
- Neumann-Holzschuh, Ingrid
 2000 Nous autres on parle peut-être pas bien français, ... mais In *Frankophone Sprachvarietäten – Variétés linguistiques francophones*, Peter Stein (ed.), 251–274. Tübingen: Stauffenberg.
- Paramasivam, K.
 1979 *Effectivity and causativity in Tamil*. Trivandrum: Dravidian Linguistics Association.
- Pillai, Nadaraja N.
 1986 *A Guide for Advanced Learners of Tamil* (= CIIL Central Institute of Indian Languages Occasional Monograph Series; 35). Mysore: Central Institute of Indian Languages.
- Pilot-Raichoor, Christine
 this volume The Dravidian zero negative: diachronic context of its morphogenesis and conceptualisation. In *Rara & Rarissima: Documenting the fringes of linguistic diversity*, Jan Wohlgemuth and Michael Cysouw (eds.), 267–303. (= Empirical Approaches to Language Typology; 46). Berlin/New York: Mouton de Gruyter.
- Posner, Rebecca
 1996 *The Romance Languages*. Cambridge etc.: Cambridge University Press.
- Posner, Rebecca
 1997 *Linguistic change in French*. New York etc.: Oxford University Press.
- Renaud, Patrick
 1979 Le Français au Cameroun. In *Le Français hors de France*, Albert Valdman, Robert Chaudenson, and Gabriel Manessy (eds.), 419–439. Paris: Champion.
- Schiffman, Harold
 1982 Negation in Tamil: Semantic and syntactic aspects. *South Asian Review* VI.3: 104–116.
- Schiffman, Harold F.
 1999 *Spoken Tamil*. Cambridge etc.: Cambridge University Press.
- Schiffman, Harold
 2005 The Grammaticalization of Aspect in Tamil, and its Semantic Sources. In *Poly-morphous Linguistics: Jim McCawley's Legacy*, Salikoko Mufwene, (ed.), 83–108. Cambridge, MA: MIT Press.

Schwegler, Armin

1990 *Analyticity and Syntheticity*. Berlin / New York: Mouton de Gruyter.

Schwegler, Armin, Bernard Tranel, and Myriam Uribe-Etxebarria (eds.)

1998 *Romance Linguistics: Theoretical perspectives*. Amsterdam / Philadelphia: Benjamins.

Steever, Sanford B.

1993 *Analysis to synthesis: the development of complex verb morphology in the Dravidian languages*. New York etc.: Oxford University Press.

Steever, Sanford B.

1998 Introduction to The Dravidian Languages. In *The Dravidian Languages*, Sanford B. Steever (ed.), 1–39. New York: Routledge.

Subrahmanyam, P. S.

1998 Kolami. In *The Dravidian Languages*, Sanford B. Steever (ed.), 301–327. New York: Routledge.

Traugott, Elizabeth Closs, and Bernd Heine (eds.)

1991 *Approaches to grammaticalization*. Amsterdam etc.: Benjamins.

Wilmet, Marc

1997 *Grammaire critique du Français*. Belgium: Duculot.

Zvelebil, Kamil V.

1969 Tamil. In *Linguistics in South Asia*, Thomas A. Sebeok (ed.) 343–371. (= Current Trends in Linguistics; 5). Paris / The Hague: Mouton.

Zvelebil, Kamil V.

1990 *Dravidian Linguistics: An Introduction*. Pondicherry: Pondicherry Institute of Linguistics and culture.

Abui tripartite verbs: exploring the limits of compositionality

Marian Klamer & František Kratochvíl

1 Introduction

Abui is a non-Austronesian (Papuan) language spoken in the Alor-Pantar Archipelago in Eastern Indonesia by about 16,000 people. A linguistic description of the northern Abui dialects is available (Stokhof 1984; Kratochvíl 2007). The Archipelago counts more than twenty languages, mostly not previously documented. These languages, with exception of Alorese and Malay, have been argued to belong to the Trans New Guinea family (cf. Pawley 2001; Ross 2005). The affiliation and internal subgrouping are currently investigated (Holton, Klamer and Kratochvíl 2009). The map in Figure 1 on the next page gives an overview of the linguistic situation of the Alor Island.

This paper describes the structure and semantics of the complex verbs in Abui that consist of three generic verbs. In Section 3, we first describe the semantic contribution of the head of the derivation, the final generic verb. We will see that the general function of generic verbs in this position is to act as a kind of ‘event classifier’, encoding various types of semantic information about the type of event denoted by the complex verb. In Section 4, the semantic contribution of generic verbs in medial position is discussed, and we argue that these medial generic verbs are in fact overt expressions of the inner aspect (‘aktionsart’) of the event. In Section 5, we discuss the semantic contribution of initial generic verbs, and show that they encode the ‘locus’ or spatial constellation of the event participants. In Section 6, we summarize the conclusions and point out some cross-linguistic parallels.

2 Abui generic verbs

Abui verbs belong to open or closed verb classes. One of the Abui closed verb classes are ‘generic verbs’ (GVs). Generic verbs are verbs with a non-specific, general semantics, whose interpretation is often contextually dependent. Abui has eighteen of such generic verbs. They are listed in Table 1 on page 211

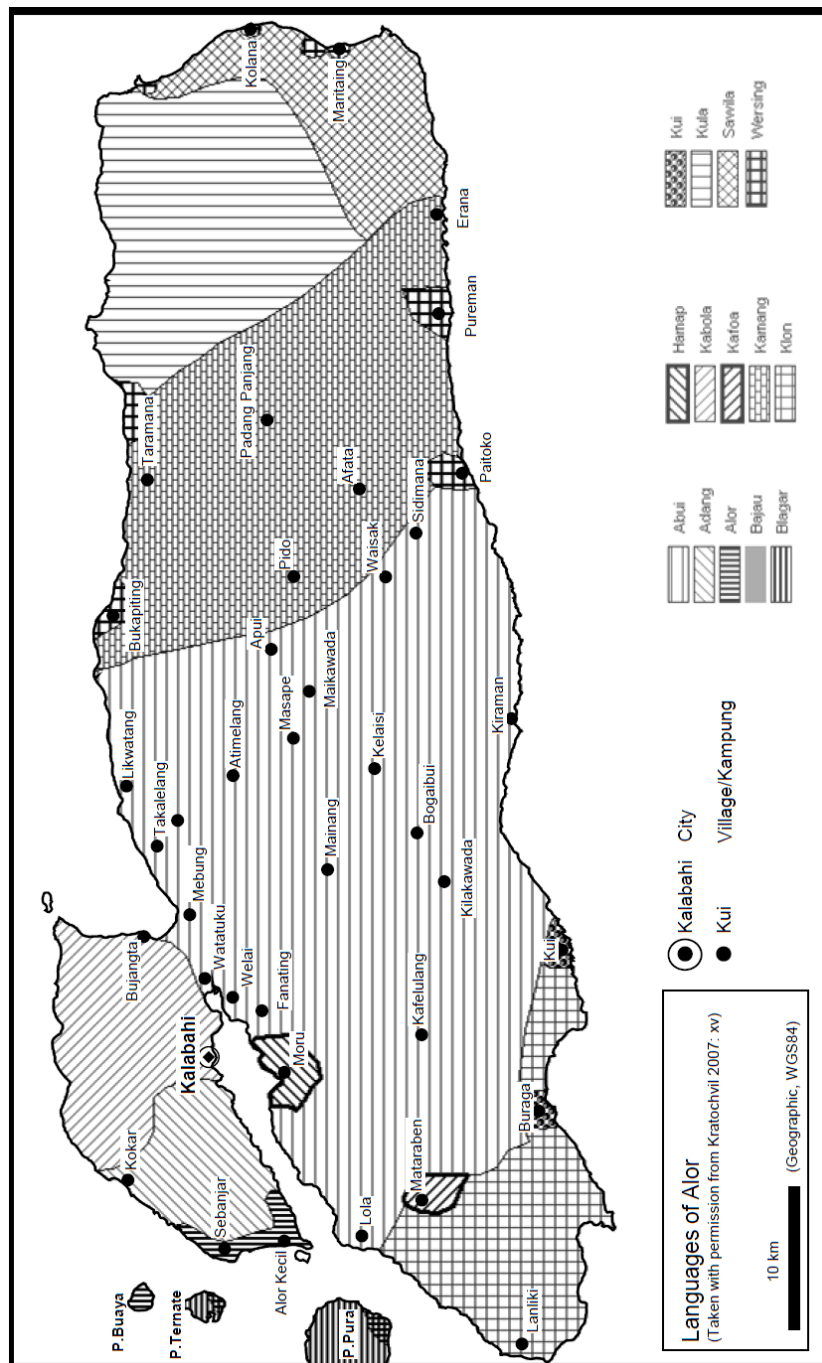


Figure 1. Linguistic situation in Alor Island

Table 1. Abui generic verbs

Root	Gloss	Meaning
<i>b</i>	JOIN	join, adjoin, absorb
<i>d</i>	HOLD	hold, get, control
<i>f</i>	SEVER	hide, loose, sever, lost, unknown
<i>h</i>	LACK	elsewhere, out, not here, not exist
<i>k</i>	THROW	throw, bring, receive, pass, feed on, move in
<i>l</i>	GIVE	give, make, affect
<i>m</i>	IN / TAKE	take, with, in
<i>ng</i>	LOOK	look, face, turn to
<i>n</i>	SEE	see, perceive, know
<i>p</i>	TOUCH	touch, be near, approach
<i>r</i>	REACH	reach, accomplish, given
<i>t</i>	LIE	lie, on, touch surface
<i>s</i>	LAY	set, move, displace
<i>i</i>	PUT	put, lay down, stop
<i>u</i>	LEAVE	leave, remote, demote, get away, gone
<i>a</i>	AT	be at, exist, last
<i>e</i>	MOVE	continue, add value
<i>o</i>	POINT	point, limit

with their gloss and some illustrations of how they are interpreted in context. Although the glosses are used throughout this paper, it should be kept in mind that they are only crude approximations of the semantics of the generic verbs. In the discussion below we will see that their semantics is generally more abstract than their English glosses suggest. The semantic functions of the generic verbs as circumscribed in Tables 5 on page 215, 10 on page 222, and 14 on page 226 may serve to illustrate this point.

As can be seen from Table 1, an Abui generic verb is very small — it consists of just one segment, a consonant or a vowel. In other words, generic verbs are mono-segmental roots. generic verbs *ng/n*, *t/s*, and *l/r* are analyzed as alternating roots. Their alternation reflects the internal temporal structure of the referred event (AKTIONSART). They refer to the same event that is viewed either as having an endpoint (COMPLETIVE), or an initial point (INCEPTIVE).

In this paper we propose that some morphologically complex verbs in Abui are built up by combining generic verbs with each other, that is, the generic verbs in Table 1 are used to derive verbal compounds. We consider

various paradigms of verbs, and investigate the regularities involved in their form and meaning. We argue that the form-meaning contrasts found in the paradigms discussed here are systematic.

While the morphological patterns described in this paper are systematic form-meaning relations, the system described here is neither productive, nor completely regular. In other words, Abui speakers do not coin new verbs by simply combining generic verbs, and not every verb discussed below has a meaning that is entirely predictable on the basis of its stipulated internal structure.

In our analysis, we take the conservative view that the segments of a verb can only be analyzed as morphemes when there is positive evidence to do so. A morphologically complex verb that is composed of generic verbs will only be identified as such under the strict condition that it occurs in a paradigm with other complex verbs, which shows a minimal contrast in form coupled to a systematic contrast in meaning. This is why the paradigms discussed in this paper are crucial. There are also many Abui verbs that do not show systematic form-meaning contrast with other verbs. It may be that such verbs are morphologically simple forms, but it may also be that some of them are derived forms whose semantics have shifted so much over time that we have no evidence of their former morphological composition. These verbs we therefore assume to be (synchronically) simplex forms.

Because an Abui phonological word minimally consists of one light syllable, only the vocalic generic verbs can be used as independent phonological words. An illustration is the verb *i* ‘put’, which functions as the main predicate in (1a). Consonantal generic verbs must always have a phonological host: they either occur with a pronominal prefix, e. g. *he-* in (1b), or attach to other lexical items, as illustrated in (2). In these examples, the generic verbs *l* ‘give’ and *ng* ‘look’ derive complex verbs from a nominal base. In line with the head-final character of Abui morpho-syntax, where objects precede verbs and negations and conjunctions occur at the end of a clause, in verbs that are derived by a generic verb this generic verb is placed at the end of the complex. We gloss the generic verbs in small case when they function as a single main predicate. This is illustrated by *i* ‘put’ in (1a) and *-l* ‘give’ in (1b). In all other functions the same gloss is in small capitals, as in (2).

- (1) a. *di sura nu meja taha-ng i-i*
 3A book SPC table on=LOOK put-PFV
 ‘He put the book on the table.’ [B07.034.06]

Table 2. Derivations from the stem *ar*

Root	Gloss	Meaning
<i>-ar</i>		‘burn’
<i>ar-a</i>	(burn- <u>AT</u>)	‘be on fire’
<i>ar-a-l</i>	(burn-AT- <u>GIVE</u>)	‘set on fire’
<i>ar-a-ng</i>	(burn-AT- <u>LOOK</u>)	‘glow’

b. *ama nu seng mi maama he-l*person SPC money take father 3II.LOC-give

‘Someone gives money to the father’, lit. ‘someone takes money gives father’

(2) a. *namu* ‘wound n.’ > *namu-l* (wound-GIVE) ‘wound v.’b. *bale* ‘bamboo sp.’ > *bale-ng* (bamboo-LOOK) ‘cook in bamboo tubes’

All morphologically derived verbs in Abui contain at least one generic verb. The final generic verb of a complex verb is considered the head of the derivation because it determines the meaning, category or valency of the derivation. The category-changing function was illustrated in (2a–2b), where *l* and *ng* derive verbs from nouns. In Table 2 it is illustrated how the final generic verb determines the meaning and valency of the derived form. The generic verb *a* ‘at’ derives stative intransitive verbs. The generic verbs *l* and *ng* derive verbs that may combine with one or two arguments.

Abui has two major types of complex verbs. One type combines a generic verb with a non-generic verb root form; in the other type, generic verbs are combined with each other. We saw examples of the former type in (2) above, where the non-generic verb root is a noun. Other complex verbs of this type are listed in Table 3 on the next page. The generic verb may combine with an open class verb, such as *kaang* ‘good’ or *ahel* ‘breathe’, a loan word, such as *buka* ‘open’, or an onomatopoeic root, such as *kirek* ‘sound of tearing (cloth)’.

Complex verbs that contain a non-generic verb root, like those in example (2) or Table 3 on the following page will be called ‘heterogeneous’ complex verbs. They are contrasted with ‘homogeneous’ complex verbs, which consist of combinations of generic verbs only (cf. Kratochvíl 2007: 299–345). An illustrative paradigm of the latter is given in Table 4 on the next page. It is this type of complex verb that we are concerned with in the present paper. Although most of the derivations in Table 4 contain more than three generic

Table 3. Some heterogenous derivations

Root	Gloss	Meaning
a.	<i>kaang</i>	(good) ‘good, proper’
	<i>kaan-r-</i>	(good.CPL- <u>REACH</u>) ‘make good, finish, ready’
	<i>kaang-d-</i>	(good- <u>HOLD</u>) ‘become good, ready’
b.	<i>ahel</i>	(breathe) ‘breathe, sniff’
	<i>ahel-r-</i>	(breathe- <u>REACH</u>) ‘make tired, fatigue’
	<i>ahel-i</i>	(breathe- <u>PUT</u>) ‘have a rest’
c.	<i>buka</i>	(open) ‘open’ (<i>Malay</i>)
	<i>buka-d-</i>	(open- <u>HOLD</u>) ‘open up’
d.	<i>kirek</i>	(tearing.sound) ‘sound of tearing’
	<i>kirek-d-</i>	(tearing.sound- <u>HOLD</u>) ‘get torn’

Table 4. Derivations sharing the stem *tu-*

Root	Gloss	Meaning
<i>t-u-ng</i>	LIE-LEAVE-LOOK	‘perforate, pierce’
<i>t-u-n-r-</i>	LIE-LEAVE-SEE-REACH	‘make a hole, inject’
<i>t-u-l</i>	LIE-LEAVE-GIVE	‘stick out, give out’
<i>t-u-k</i>	LIE-LEAVE-THROW	‘stick out, measure’
<i>t-u-k-u</i>	LIE-LEAVE-THROW-LEAVE	‘cut off’
<i>t-u-k-o-ng</i>	LIE-LEAVE-THROW-POINT-LOOK	‘cut at something’
<i>t-u-k-o-l-</i>	LIE-LEAVE-THROW-POINT-GIVE	‘cut in’
<i>t-u-k-o-l-r</i>	LIE-LEAVE-THROW-POINT-GIVE-REACH	‘make a hole’
<i>t-u-k-o-l-a-d-</i>	LIE-LEAVE-THROW-POINT-GIVE-AT-HOLD	‘leak, have a hole’

verbs, the present paper will focus on derivations that consist of three generic verbs. The main reason for this is that paradigms with contrasting tripartite complex verbs occur most frequently and have the most transparent semantics.

It must be noted here that some logical combinations of generic verbs do not occur in tripartite derived verbs because they would not constitute a well-formed syllable or minimal word in Abui. For example, derivations with two initial consonantal generic verbs are not possible (**pta*, **kpi*, or **lnu*), nor are tripartite verbs combining two non-high vowel generic verbs (**tao*, **koa*, **loe*). This is because Abui syllable structure is strictly (consonant)-vowel-(consonant), with only simple onsets and codas, and codas can consist of only consonants or high vowels.

Table 5. Description of the event type marked by the generic verb in final position

Root	Gloss	Event type
<i>b</i>	JOIN	impact
<i>d</i>	HOLD	internally caused change of state
<i>f</i>	SEVER	separation
<i>h</i>	LACK	perfect
<i>k</i>	THROW	motion away from the deictic centre
<i>l</i>	GIVE	externally caused change of state
<i>m</i>	IN / TAKE	position in the deictic centre
<i>ng</i>	LOOK	orientation towards deictic centre
<i>n</i>	SEE	completed orientation towards deictic centre
<i>p</i>	TOUCH	contact
<i>r</i>	REACH	completed externally caused change
<i>t</i>	LIE	completed horizontal contact, position
<i>s</i>	LAY	horizontal orientation
<i>i</i>	PUT	downwards motion, accomplishment
<i>u</i>	LEAVE	perfect
<i>a</i>	AT	state, position
<i>e</i>	MOVE	ongoing motion

3 The semantic contribution of final generic verbs

The final generic verb of a complex verb is the head of the derivation, as it determines the meaning, category and valence. Semantically, the final generic verb functions as a kind of ‘event classifier’, encoding various types of semantic information about the type of event expressed by the complex verb. These semantic dimensions include, for example, whether the event denoted by the verb is a state or a motion (and the orientation of the motion), whether the event involves causation (internal or external), or whether it is an achievement. Table 5 presents a survey of the various event types that are marked by final generic verbs. Note that the meaning of the generic verb root and its function as event classifier are semantically connected, though the link is quite abstract. In the remainder of this section, we illustrate some of these event types encoded by the third generic verbs using paradigms of complex verbs that only contrast in their final generic verb.

Not all of the generic verbs in Table 5 are equally productive as event classifiers. For example, in the tripartite derivations that are the topic of this paper, the verbs *h*, *k*, *l*, *r*, *ng*, *n*, *p*, *t*, and *i* frequently occur as final generic

verbs. The verbs *l*, *r*, *d*, *ng* and *i* are typically used as final generic verb in larger derivations or in heterogeneous verbs. The verbs *f* and *b* have an overall low frequency. The verb *u* is used most often in non-final position, or as the second generic verb in a bipartite derivation. The verbs *s* and *m* are typically used as initial generic verbs in verbs consisting of multiple generic verbs.

In Table 6, a paradigm is given where tripartite complex are derived from the base verb *t-a* ‘be on’. We will discuss each of these derivations in detail below.

Table 6. Derivations sharing the stem *ta-*

Root	Gloss	Meaning
<i>t-a</i>	LIE-AT	‘be on, lie on’
<i>t-a-k</i>	LIE-AT-THROW	‘shoot down, cover, bring on, plant’
<i>t-a-l</i>	LIE-AT-GIVE	‘drop on’
<i>t-a-i</i>	LIE-AT-PUT	‘put on’
<i>t-a-ng</i>	LIE-AT-LOOK	‘release, let drop’

In *t-a-k*, the verb *k* ‘throw’ derives a verb indicating a motion event, which causes the participant(s) to ‘throw or move’ an object to be lying down. Depending on the context in which the verb is used, it may get various interpretations, including ‘to shoot’, or ‘to cover with’. An illustration is (3), where *t-a-k* refers to shooting down birds.

- (3) *Arjun de-kartipel ong ba mara kuya ha-t-a-k*
 name 3I.AL-catapult make SIM go.up.CNT bird 3II.PAT-LIE-AT-THROW
 ‘Arjun made himself a catapult and went up shoot the birds.’ [B06.014.03]

In (4), *t-a-k* refers to covering one’s head with a piece of cloth, ‘to cover with’ thus being a possible interpretation of ‘move to lie on’.

- (4) *ama do di kabala do mi de-fui he-t-a-k-i*
 person PRX 3A cloth PRX take 3I.AL-vertex 3II.LOC-LIE-AT-THROW-PFV
 ‘and the people covered the top of their heads with a cloth’
 [B02.169.08:43]

The final generic verb *l* ‘give’ encodes an event that is externally caused and involves a transfer of participants. In the derivation *t-a-l* this renders the interpretation of ‘to drop on’. This is illustrated in (5). In (5a), the verb *t-a-l*

refers to the ‘sowing’ of tobacco seeds, in (5b), it refers to the ‘dropping’ of tears.

- (5) a. *maama di kafaak ha-t-a-l*
 father 3A tobacco 3II.PAT-LIE-AT-GIVE
 ‘Father sows the tobacco.’ [B08.051.02]
- b. *t-ieng nai t-a-l-i ayok-ayok-d-a*
 DISTR.INAL-eye tear LIE-AT-GIVE-PFV RED[two]-HOLD-DUR
 ‘Our eyes were dropping tears (each of us was crying).’ [B08.050.11]

In final position, the generic verb *i* ‘put’ derives a complex verb that refers to an achieved state or accomplished event. In *t-a-i*, the verb *i* ‘put’ denotes a state of being put down or being put in a certain location. This is illustrated in (6). In (6a), *t-a-i* is interpreted as ‘to put st. down’, in (6b), it refers to curing a person by ‘putting on’ medicine, while in (6c), it is used to express an order as something (a duty) that the old woman ‘imposes on’ the child.

- (6) a. *na yambuk mi t-a-i natia*
 1S glass take LIE-AT-PUT stand.CNT
 ‘I put down the glass.’ [B04.053.01]
- b. *ama he-l t-a-i tung ayok-d-a*
 person 3II.LOC-GIVE LIE-AT-PUT year two-HOLD-DUR
 ‘People were treating him for two years.’ [B07.022.04]
- c. *kalieta mayol do di moku do ho-t-a-i*
 old.person woman PRX 3A kid PRX 3II.REC-LIE-AT-PUT
 ‘the old woman, she ordered the child’ [B02.017.09:35]

In final position, the generic verb *ng* ‘look’ classifies events as motions towards a location. The Abui verbs ending in *ng* and applicative verbs in other languages sometimes function alike. In (7), the verb *t-a-ng* is thus interpreted as ‘to release (dogs)’, as Abui hunting is typically done by setting dogs off to track down the prey.

- (7) *neng loku ba he-n-u kaai ha-t-a-ng kaang*
 man PL TOP 3II.LOC-like.PRX-PRF dog 3II.PAT-LIE-AT-LOOK good
 ‘As for the men, these hunt.’ [B07.032.06]

Ng ‘look’ encodes motions towards a location as well as low involvement of the actor. For example, *t-a-ng* can contrast with *t-a-l* ‘to drop something’, discussed in (5), when it is interpreted as ‘let drop’, for example, ‘let water drop (on the floor)’.

Table 7. Complex verbs with a contrasting final generic verb *k*, *l*, *ng*, and *i*

	THROW	GIVE	LOOK	PUT
a.	<i>d-a-k</i> HOLD-AT-THROW 'cover'	<i>d-a-l</i> HOLD-AT-GIVE 'handle, grab'		<i>d-a-i</i> HOLD-AT-PUT 'cram(med)'
b.	<i>l-e-k</i> GIVE-MOVE-THROW 'point to'	<i>l-e-l</i> GIVE-MOVE-GIVE 'threaten, almost do'		<i>l-e-i</i> GIVE-MOVE-PUT 'miss (not hit)'
c.	<i>t-u-k</i> LIE-LEAVE-THROW 'stick out, measure'	<i>t-u-l</i> LIE-LEAVE-GIVE 'stick into'	<i>t-u-ng</i> LIE-LEAVE-LOOK 'perforate, pierce'	
d.	<i>l-u-k</i> GIVE-LEAVE-THROW 'rub, wipe, bend'		<i>l-u-ng</i> GIVE-LEAVE-LOOK 'be long-termed'	
	motion	caused event	oriented to location	accomplished

Some more complex verbs with final generic verbs *k* 'throw', *i* 'put', *l* 'give', and *ng* 'look' are given in Table 7. As was noted in Section 2, the derivational patterns shown by the paradigms are not always regular, and combinations that are logically possible are not always attested. Note also that *i* 'put' refers to some aspectual dimension, but how this semantic contribution relates to the verbal root meaning 'put' is much less clear than in the case of *k* 'throw', *l* 'give', and *ng* 'look'.

Now consider the paradigms in Table 8 on the next page. The verbs in the first column end in *k* 'throw' and thus express a motion event. In contrast, the verbs *t* 'lie' and *p* 'touch' both express a non-motion event: a horizontal state, or a state of contact. As such, they mark 'states' (achievements, results) in contrast to the motions encoded by *k*. Further, *t* and *p* contrast in the position of the event participants: *t* 'lie' refers to participant(s) in horizontal position, and *p* 'touch' refers to participant(s) in contact with another object. Again, the derivational pattern observed in the paradigms here is not totally regular.

The final two sets of examples in Table 8 on the facing page illustrate the contrast in participant positions most clearly. Example (8) illustrates the uses of the forms *t-e-k* and *t-e-t*. In (8a), the verb *t-e-k* refers to a motion event that leads to horizontal position. In a context where 'garden' is the object, the interpretation is that weeds and bushes are going to be cut down, or stepped on and broken, so that they end up on the ground. In (8b), the verb *t-e-t* refers

Table 8. Complex verbs with a contrasting final generic verb *k*, *l*, *ng*, and *i*

THROW	LIE	TOUCH
a. <i>t-u-k</i> LIE-LEAVE-THROW 'stick out, measure'	<i>t-u-t-</i> LIE-LEAVE-LIE 'emerge, surface'	<i>t-u-p-</i> LIE-LEAVE-TOUCH 'stuck out'
b. <i>t-a-k</i> LIE-AT-THROW 'shoot, empty'		<i>t-a-p-</i> LIE-AT-TOUCH 'shot (down)'
c. <i>k-a-k</i> THROW-AT-THROW 'stab'		<i>k-a-p-</i> THROW-AT-TOUCH 'stabbed (in)'
d. <i>l-u-k</i> GIVE-LEAVE-THROW 'rub, wipe, bend'	<i>l-u-t-</i> GIVE-LEAVE-LIE 'rubbed, bent'	
e. <i>m-o-k</i> IN-POINT-THROW 'put together'		<i>m-o-p-</i> IN-POINT-TOUCH '(already) put together'
f. <i>t-e-k</i> LIE-MOVE-THROW 'slide'	<i>t-e-t-</i> LIE-MOVE-LIE 'slid on'	
g. <i>k-e-k</i> THROW-MOVE-THROW 'prod'	<i>k-e-t-</i> THROW-MOVE-LIE 'prodded'	
motion	state, horizontal	state, contact

to a wall that collapsed during an earthquake. The final generic verb *t* 'lie' indicates the final horizontal position of the wall has been reached. *Tet* is the final verb of the sentence and has to be inflected for aspect, in this case with the perfective suffix *-i*.

- (8) a. *na yaa ne-'ut t-e-k*
 1S go 1S.AL-garden LIE-MOVE-THROW
 'I go clear my gardens',
 lit.: 'I go bring down my gardens.' [B01.035.32]
- b. *amakaang nuku kota ho-t-e-t-i*
 person one wall 3II.REC-LIE-MOVE-LIE-PFV
 'the wall collapsed down on one person' [B05.078.01]

Example (9) shows the contrast between *k-e-k* and *k-e-t*. The verb *k-e-k* expresses a vertical motion towards the object, the blossom of a banana tree; in this context it is translated as ‘to prod’, that is, use a stick to prod at a banana blossom hanging in a tree. In (9b), the event *k-e-t* involves a horizontal movement that has reached its final boundary.

- (9) a. *di baleei wataka do k-e-k he-we*
 3A banana blossom PRX THROW-MOVE-THROW 3II.LOC-leave
 ‘she went to prod at the banana blossom’ [B02.027.03:59]
- b. *na wi ha-k-o-k-u nemang tahai maar*
 1S stone 3II.PAT-THROW-POINT-THROW-PRF shells search cook.CPL
kaan-r-i, he-k-e-t ba nee
 GOOD-REACH-PFV 3II.LOC-THROW-MOVE-LIE SIM eat
 ‘I prodded the stones, searched for shells, cooked them, poked them out and ate them.’ [B05.070.02]

As the final example in this section, consider the paradigm derived from *l-o-* in Table 9. If the semantics of these derivations reflected their morphological make-up in a regular way, the complex verb ending in *k* ‘throw’ would refer to a motion event, a final *i* ‘put’ would mark an accomplishment, with *h* ‘lack’ the verb would have perfect aspect, and *ng* ‘look’ would mark orientation towards a location. As it is, it is very hard to find these semantics reflected in the translations as given.

While the different verb forms in Table 9 show some semantic overlap, it is impossible to analyse all of them as regularly derived compounds; the semantic contribution of their head (the final generic verb) is elusive. It is difficult to relate it to how it is used in other paradigms. In other words, the overall morphological function of particular generic verbs can only be spotted by comparing their function across various sets of paradigms (cf. the sets discussed above, and the ones to be discussed in the sections below). But even then, the patterns are not always regular and semantically transparent, and there are many paradigms with lexical gaps.

Table 9. Possible paradigm derived from *l-o-*

THROW	PUT	LACK	LOOK
<i>l-o-k</i>	<i>l-o-i</i>	<i>l-o-h-</i>	<i>l-o-ng</i>
GIVE-POINT-THROW	GIVE-POINT-PUT	GIVE-POINT-LACK	GIVE-POINT-LOOK
‘prick, push’	‘to chase, distance’	‘distant, long’	‘long’

The unpredictable, lexicalised nature of the derivations discussed here is due to the nature of the generic verbs involved. The complex verbs are built from generic verbs, and such verbs already start out with a rather abstract (or vague) lexical semantic content. Combining with each other to form new verbal compounds, they become reanalyzed as bound morphemes and lose some of their verbal semantics (which was already underspecified to begin with). Abstract verbs interacting with other abstract verbs in a derivation are also bound to undergo semantic drift or shift, with the result that the compound may become reanalyzed as a simplex word over time. Cross-linguistically, the lexicalization and subsequent semantic drift of compounds, as well as the existence of paradigmatic gaps is characteristic of all types of compounding (Spencer 1991: 312). This is because the meaning contribution of compound structure generally involves a high degree of abstractness or vagueness, since the semantic relation between the elements in a compound is not formally expressed (cf. Booij 2005: 210).

Despite these complications, however, we have argued that there is evidence to analyze the final generic verbs in complex verbs as ‘event classifiers’ that encode information about the type of event denoted by the verb. Interestingly, their function as final verb differs from the function they have when they occur as non-heads, i. e. in medial or initial position. This will be discussed in the next two sections.

4 The semantic contribution of medial generic verbs

For the phonological reasons set out in Section 1, a tripartite complex verb will always have a vocalic generic verb in medial position. The vocalic generic verbs are *a* ‘at’, *i* ‘put’, *e* ‘move’, *o* ‘point’, and *u* ‘leave’. In medial position, these verbs encode the inner aspect (aktionsart) of the event, providing information on its boundedness.

In Table 10 on the next page, we list the vocalic generic verbs with the aspectual function they have in medial position. The verb *a* ‘at’ encodes an unbounded event or state (‘progressive’), the verb *o* ‘point’ encodes an event that is bounded at both start and end (‘punctual’), the verb *i* ‘put’ encodes events that are bounded at the end (‘terminative’), and it contrasts with the verb *e* ‘move’ that encodes events which are bounded only at the start (‘ingressive’). Finally, the verb *u* ‘leave’ encodes events, which have reached beyond their final boundary (‘perfect’). In Table 11 on the following page some paradigms illustrating the contrast are given. The last column of verbs in Ta-

Table 10. The aspectual function of vocalic generic verbs in medial position

Generic verb	Gloss	Type	Boundary
<i>i</i>	PUT	terminative	bounded at end
<i>u</i>	LEAVE	perfect	beyond final boundary
<i>a</i>	AT	progressive	no boundary
<i>e</i>	MOVE	ingressive	bounded at start
<i>o</i>	POINT	punctual	bounded at start and end

Table 11. Derivations with contrasting generic verbs in medial position

<i>a</i> (AT)	<i>l-a-k</i>	<i>k-a-k</i>	<i>t-a-k</i>
progressive	GIVE-AT-THROW 'mark, count'	THROW-AT-THROW 'penetrate, stab'	LIE-AT-THROW 'put down, shoot'
<i>o</i> (POINT)	<i>l-o-k</i>	<i>k-o-k</i>	<i>t-o-k</i>
punctual	GIVE-POINT-THROW 'prick'	THROW-POINT-THROW 'prod (once)'	LIE-POINT-THROW 'drop, pour'
<i>i</i> (PUT)	<i>l-i-k</i>	<i>k-i-k</i>	<i>t-i-k</i>
terminative	GIVE-PUT-THROW 'bend'	THROW-PUT-THROW 'sweep'	LIE-PUT-THROW 'loosen, untie'
<i>e</i> (MOVE)	<i>l-e-k</i>	<i>k-e-k</i>	<i>t-e-k</i>
ingressive	GIVE-MOVE-THROW 'point to/at'	THROW-MOVE-THROW 'prod'	LIE-MOVE-THROW 'slide'
<i>u</i> (LEAVE)	<i>l-u-k</i>	<i>k-u-k</i>	<i>t-u-k</i>
perfect	GIVE-LEAVE-THROW 'rub, bend'	THROW-LEAVE-THROW 'push out'	LIE-LEAVE-THROW 'stick'

ble 11 is exemplified in (10)–(16). In (10), the verb *t-a-k* refers to shooting down of a pig. The root *a* 'at' in the complex verb *t-a-k* refers to the aktionsart of 'lying' and indicates that the event of 'lying' has no internal boundary. The final verb *k* classifies this verb as a motion event that brings about the result that a pig is 'lying down'; the actor does not directly affect the undergoer, and the resulting event of lying has no boundary. The undergoer participant 'pig' is expressed with the NP *fe nu* 'a pig' and co-indexed as patient with the prefix *ha-*.

- (10) *no-buoka ba pulang mi fe nu ha-t-a-k*
 1S.REC-far SIM arrow take pig SPC 3IL.PAT-LIE-AT-THROW
 'From far, I shoot a pig with an arrow.' [B05.067.01]

In contrast, the verb *t-o-k* in (11) refers to ‘pouring’ or ‘dropping’ water on the fire. Here the generic verb *o* ‘point’ encodes that this event is punctual: it has a limited time span and occurs once.

- (11) *na ya mi ba ara he-t-o-k*
 1S water take SIM fire 3II.LOC-LIE-POINT-THROW
 ‘I pour water on the fire.’ [B10.021.13]

In (12), the verb *t-o-k* is interpreted as ‘demolish’ or ‘break down’, an event with a limited time span. The verb *k* marks it as a motion away from the speaker. Note also that the undergoer must be expressed with the patient prefix *ha-* indicating a significant affectedness and change of state.

- (12) *na fala ha-t-o-k*
 1S house 3II.PAT-LIE-POINT-THROW
 ‘I demolish the house.’ [B03.002.03]

In (13), the complex verb *t-e-k* also ends in *k* and refers to a motion event, which now originates from where the wall once stood. The verb *e* encodes an event of ‘sliding down, collapsing’ that has just started. In other words, an event with an initial boundary.

- (13) *kota no-t-e-k-e*
 wall 1S.REC-LIE-MOVE-THROW-IPFV
 ‘the wall slides down on me’ [B05.078.01]

The inner aspect of the event *t-e-k* contrasts with that of *t-i-k*, as the latter has a final boundary. An example is given in (14), where a child that was locked up (the door is tied up with a rope), managed to untie the door and get out to play. The actual untying is encoded as a motion (encoded by final *k*) that brings about a final boundary (encoded by *i*) to ‘lying’. The single participant is expressed with two arguments. The free pronoun *di* marking the actor records the argument of the verb *t-i-k*. The coreferential prefix *de-* records the same participant as the benefactive of the auxiliary verb *l* ‘give’.

- (14) *di de-l t-i-k-i ba mui-l-a*
 3A 3I.LOC-GIVE LIE-PUT-THROW-PFV SIM game-GIVE-DUR
 ‘He released himself and played.’ [B10.023.03]

The verb *t-u-k* covers a range of meanings including ‘measure, emerge, stick out’. In (15), it combines with two arguments, a third person actor *di* and a

Table 12. Derivations with contrasting generic verbs in medial position

<i>a</i> (AT)	<i>w-a-k</i>		<i>s-a-k</i>
progressive	LEAVE-AT-THROW 'embrace, throw'		LAY-AT-THROW 'align'
<i>o</i> (POINT)	<i>w-o-k</i>	<i>k-o-l</i>	<i>s-o-k</i>
punctual	LEAVE-POINT-THROW 'throw (once)'	THROW-POINT-GIVE 'tie, bind'	LAY-POINT-THROW 'pass along'
<i>i</i> (PUT)	<i>w-i-k</i>	<i>k-i-l</i>	<i>s-i-k</i>
terminative	LEAVE-PUT-THROW 'carry away'	THROW-PUT-GIVE 'turn over'	LAY-PUT-THROW 'break off, pluck'
<i>u</i> (LEAVE)		<i>k-u-l</i>	
perfect		THROW-LEAVE-GIVE 'throw(n)'	

first person undergoer *no-*. *T-u-k* refers to the event of 'measuring clothes' which is described as clothes that are laid down in a motion away from the actor.

- (15) *di naamang mi no-t-u-k*
 3A cloth take 1S.REC-LIE-LEAVE-THROW
 'He measured me the clothes on.' [B10.021.06]

In (16), *t-u-k* refers to knowing a person, an interpretation that is related to 'measure' in example (15) above.

- (16) *di dara na-t-u-k naha to*
 3A still 1S.PAT-LIE-LEAVE-THROW not PRX.AD
 'He does not know me yet.' [B10.021.04]

Some additional (partial) paradigms are given in Table 12 above. The vocalic generic verbs encode the internal temporal structure of the encoded events.

In sum, in verb lexemes consisting of three segments, the medial vocalic generic verbs encodes the inner aspect of the event denoted by the verb. Abui thus has a set of complex verbs that express aktionsart systematically and overtly in their derivational morphology.

Table 13. Derivations with contrasting generic verbs in initial position

<i>t-a-k</i>	<i>s-a-k</i>	<i>p-a-k</i>	<i>d-a-k</i>
LIE-AT-THROW	LAY-AT-THROW	TOUCH-AT-THROW	HOLD-AT-THROW
‘shoot, put down’	‘align’	‘dump, crash in’	‘clutch, trap’
<i>b-a-k</i>	<i>f-a-k</i>	<i>l-a-k</i>	<i>w-a-k</i>
JOIN-AT-THROW	SEVER-AT-THROW	GIVE-AT-THROW	LEAVE-AT-THROW
‘swallow, snatch’	‘break’	‘mark, count’	‘embrace, throw’

5 The semantic contribution of initial generic verbs

In this section, we analyse the function of the initial generic verb in a complex verb as describing something we refer to as the ‘locus’ of the event, or the spatial constellation of the event. Consider now the derivations in Table 13. Because of their final verb *k*, all these verbs refer to events that are classified as motions, while the medial verb *a* ‘at’ indicates that the event has no internal boundary. The initial verb is variable, and indicates the spatial constellation of the event.

As ‘locus’ marking derivational morphemes, initial generic verbs take a perspective that is quite different from what we are used to in derivational morphology. For example, the verb *t* ‘lie’ indicates a horizontal ‘locus’ that is reached by the motion (expressed by *k*). The verb *s* ‘lay’ indicates a parallel ‘locus’. The verb *p* ‘touch’ indicates a locus that may be described as ‘touching of a surface’, while the root *d* ‘hold’ encodes that the event involves an ‘overall’ type of contact between participants and their location. The root *b* ‘join’ indicates an intersecting locus, while the root *f* ‘sever’ refers to a separated locus. The root *l* ‘give’ refers to a locus that may be described as ‘in reach’. Practically it means that the involved participants are ‘within reach’ of each other, but more distant from each other than e. g. when the generic verb is *p* ‘touch’ or *b* ‘join’. At the same time the participants are still within reach of each other, unlike the participants in cases when the roots *u* ‘leave’ (pronounced and written as *w* in initial position) and *f* ‘sever’ are used. The verb *u* ‘leave’ indicates the remote locus. A summary is given in Table 14 on the following page. (Question marks indicate functions that are unclear to us, asterisks indicate unattested derivations.)

The verb *s-a-k* describes the event of motion that brings about ‘parallel lying, alignment’. In (17a), it refers to the parallel arrangement of bamboo. In (17b), it refers to making an appointment, where the agreement is perhaps

Table 14. Event ‘locus’ encoded by the initial generic verb

Generic verb	Gloss	‘Event locus’
<i>b</i>	JOIN	Intersecting locus
<i>d</i>	HOLD	Overall contact
<i>f</i>	SEVER	Separation
<i>h</i>	LACK	?
<i>k</i>	THROW	Out-of-reach locus
<i>l</i>	GIVE	Within-reach locus
<i>m</i>	IN / TAKE	In-deictic-centre locus
<i>n</i>	SEE	?Visible locus
<i>ng</i>	LOOK	*
<i>p</i>	TOUCH	Contact locus
<i>r</i>	REACH	Within-reach locus
<i>s</i>	LAY	Parallel locus
<i>t</i>	LIE	Horizontal locus
<i>i/y</i>	PUT	*
<i>u/w</i>	LEAVE	Remote locus

considered as some kind of ‘alignment or parallel arrangement’ of suggestions / ideas in the vicinity of the speaker.

- (17) a. *ri ri-mai s-a-k he-te-l tahi-i*
 2P 2P.AL-bamboo LAY-AT-THROW 3II.LOC-DISTR.LOC-GIVE on.CPL-PFV
 ‘you put bamboo in parallel, put it together on top of each other’
 [B02.004.05:20]
- b. *ama el na-pe-i s-a-k-i*
 person before 1S.PAT-near-PFV LAY-AT-THROW-PFV
 ‘people made an appointment with me’ [B02.074.08:16]

In (18), the verb *p-a-k* refers to car passengers crashing down into a ravine. It is a motion event (encoded by *k*), and the initial generic verb *p* ‘touch’ encodes the ‘contact locus’ of the event, the contact between the participants and some location. The motion is oriented downward.

- (18) *oto he-amakaang ba ho-mi mi-a loku yo mi*
 car 3II.AL-person TOP 3II.REC-IN in-DUR PL MD.AD take
p-a-k mahoi-n-i
TOUCH-AT-THROW together-SEE-PFV
 ‘the car passengers that were in the car, they crashed (into a ravine) all together’ [B05.071.05]

In (19a), *d-a-k* refers to holding someone tight. The initial root *d* ‘hold’ indicates that the type of contact between the participants involves the entire surface, i. e. is ‘overall’. In (19b), *d-a-k* refers to leaves applied in traditional healing that are ‘clutched’ on to the speaker. The verb *d-a-k* is also used to refer to the clutching of fish and mice in bamboo traps.

- (19) a. *me na-d-a-k-e!*
 come 1S.PAT-HOLD-AT-THROW-IPFV
 ‘Come and hold me tight!’ [B10.019.05]
- b. *tuli tala mi ba ne-l ha-d-a-k-e*
 tree.sp. leaves take SIM 1S.LOC-GIVE 3II.PAT-HOLD-AT-THROW-IPFV
 ‘Take the leaves of the *tuli* tree and clutch them on me.’ [B10.021.01]

In (20), *b-a-k* indicates a motion event with ‘intersecting locus’. It can be interpreted as ‘snatch’, where one participant intersects the location of another. It can also refer to ‘jaws’, ‘biting’ or ‘snatching food’.

- (20) *na he-l b-a-k-i ba laak-i*
 1S 3II.LOC-GIVE JOIN-AT-THROW-PFV SIM leave.for-PFV
 ‘I snatched him away.’ [B10.024.05]

In (21a), the verb *f-a-k* is used. Here, the motion has a ‘separation locus’, and the verb is interpreted as ‘break’: in a leg that breaks, parts are separated from each other. In (21b), the same verb is used in a transitive construction referring to breaking a banana.

- (21) a. *abui mi-a ne-toku f-a-k-i*
 mountain in-DUR 1S.AL-leg SEVER-AT-THROW-PFV
 ‘my leg broke in the mountains’ [B06.017.07]
- b. *pi=fal baleei f-a-k*
 1PI=together banana SEVER-AT-THROW
 ‘We share a banana’, lit. ‘we break a banana together.’ [B10.007.01]

In (22a), the verb *l-a-k* refers to a motion away from the actor and affecting a house that is ‘in-reach locus’ with respect of the actor. Its meaning corresponds to the English verb ‘demolish’ or ‘take apart’. In (22b), the same verb refers to the event of ‘marking’ or ‘recognizing’. In a serial construction with *iéng* ‘look’, *l-a-k* refers to ‘knowing someone’.

- (22) a. *maama di fala ha-l-a-k*
 father 3A house 3II.PAT-GIVE-AT-THROW
 ‘Father demolished the house.’ [B06.011.02]

- b. *na he-do n-iéng l-a-k naha*
 1S 3II.LOC-HOLD.PNCT 1S.PAT-look GIVE-AT-THROW not
 ‘I don’t know him’ [B10.047.11]

In (23a), *w-a-t* refers to a motion with ‘remote locus’. The verb refers to ‘abandoning’ of parents by their children. In (23b) *w-a-k* refers to the ‘embracing’ of two participants, which is encoded as a motion, with *w* ‘leave’ indicating the remote locus. The distributive prefix *ta-* is used to encode the reciprocal involvement of the participants.

- (23) a. *ni-ya maama o w-a-t he!*
 1PE.AL-mother father MD LEAVE-AT-LIE PROH
 ‘we should not abandon our mother and father’ [B02.158.00:02]
- b. *di=ning ayoku ta-w-a-k ba luuk do=ng*
 3A=in.number two DISTR.PAT-LEAVE-AT-THROW SIM dance PRX=LOOK
we
leave
 ‘they both embraced (each other) to enter the dance’ [B02.087.08:17]

Table 15 lists two paradigms of complex verbs with an alternating initial generic verb. In the first row, the verbs end in *ng*, which encodes that the event is oriented towards a location. The verb *p* encodes ‘contact’ between the participants, *m* encodes that the participants are at the same location, *t* encodes that they are in horizontal position, and *b* encodes that the participant(s) have impact on something else. In the second row, the verbs end in *l* ‘give’ which encodes that the event is externally caused. The medial verb *u* ‘leave’ encodes that the event has reached its final boundary and has persistent effect. The initial verbs encode the event locus: *f* ‘sever’ refers to the separation locus of two participants, *k* ‘throw’ refers to locus that is not within reach (but not too remote either), *t* ‘lie’ refers to the horizontal locus of the participants,

Table 15. Partial derivations with contrasting generic verbs in initial position

<i>p-a-ng</i>	<i>m-a-ng</i>	<i>t-a-ng</i>	<i>b-a-ng</i>
TOUCH-AT-LOOK	IN-AT-LOOK	LIE-AT-LOOK	JOIN-AT-LOOK
‘feel, perceive’	‘possess’	‘release’	‘carry on shoulder’
<i>f-u-l</i>	<i>k-u-l</i>	<i>t-u-l</i>	<i>r-u-l</i>
SEVER-LEAVE-GIVE	THROW-LEAVE-GIVE	LIE-LEAVE-GIVE	REACH-LEAVE-GIVE
‘swallow’	‘throw’	‘stick out, measure’	‘loosen, let slide’

and *r* ‘reach’ refers to the locus where participants were previously ‘in reach’ of each other.

In sum, we have argued that when a generic verb occurs in the initial position of a tripartite complex verb, it describes the ‘event locus’; that is the spatial constellation of the event participants.

6 Conclusions and cross-linguistic perspective

Abui complex verbs behave as single morpho-syntactic units. Heterogeneous complex verbs consist of a non-generic root and a generic verb (generic verb). Homogeneous complex verbs are composed of only generic verbs. In this paper, we have analyzed a number of Abui verbs as homogeneous complex verbs that combine three generic verbs to derive verbal compounds. The semantic composition of such tripartite compounds can be schematically represented as in (24).

(24) [EVENT.LOCUS-INNER.ASPECT]-EVENT.TYPE

The final generic verb of these compounds is the morphological head and encodes the event type. The medial and initial generic verbs are the dependent or modifying elements. In medial position, the generic verb is a vowel and encodes the telic properties of the event (inner aspect). generic verbs in initial position describe the ‘event locus’. For many of the generic verbs we can observe that the different functions attached to the same phoneme in initial, medial and final position are semantically related.

When our Abui consultants translate the meaning of a complex verb, unable to find a fitting Malay equivalent, they regularly resort to hand gestures that express the locus and indicate the trajectory of the event (for instance the verb *d* ‘hold’ is illustrated by putting one hand on top of an object or on a table, the verb *t* ‘lie’ is illustrated by putting both hands on each other. Systematic research into this use of gesture could provide additional insights in the type of event decompositions suggested in this paper.

Abui complex predicates have parallels in constructions found in other languages. A closed set of inflecting verbs with a classificatory function is a common feature in some Northern Australian languages (Schultze-Berndt 2000; McGregor 2002). Classificatory verbs are also found in the some Trans-New-Guinea highland languages of the Papua mainland, combining with nominals and stative verbs (Lang 1975). In Kalam and Kobon (Davies 1984;

Pawley 1987, 1991, 2004, 2005), classificatory verbs can combine with each other in a similar way as the complex predicates in Abui. As stated in Section 1, claims have been made that the highland languages might be related with the non-Austronesian languages of Alor and Pantar Archipelago.

The Kalam verb inventory is smallish, about 100 verbs, and about 25 of these are used frequently. As in Abui, the Kalam verbs have very generic meanings that do not correspond to the meanings of single verbs in more familiar languages like English. (It is thus unfortunate that English must be used for the glossing of such generic verbs.) The Kalam generic verbs often combine with each other. However, in the Kalam system, a root verb may only occur once in a verb sequence, while we have seen that in Abui, the same generic verb can occur more than once in one and the same derivation. Kalam and Kobon verb complexes have been analyzed as serial verb constructions. Consider now the Kalam example in Table 16. The data come from Pawley (1969, 2005) and from Pawley and Lane (1998).

The verb *ag* ‘say’ occurs as either the first or the second member of the complex verb. As the second verb it appears to have a classifying function: it refers to the type of the event. As a first verb, it specifies the event and encodes the means by which the event is carried out. The Kalam verb sequences are argued to be organized temporally, as illustrated in Table 17 on the next page. However, in some cases, one structure can have two readings, as in Table 18 on the facing page. Only one of the readings is sequential, the other suggest a hierarchical structure in the complex verb. The variation in meaning suggest that these sequences have a different underlying structure. When *ap jak am-* means ‘come and go’, the structure can be read as temporally organized sequence; it can be read as a serialization. When *ap jak am-* means

Table 16. Kalam complex verbs derived from *ag* ‘say’

Verb	Gloss	Meaning
<i>ag l-</i>	SAY STABILIZE	‘ask to remain, make an appointment’
<i>ag ñ-</i>	SAY TRANSFER	‘tell, inform’
<i>ag tk-</i>	SAY SEVER	‘interrupt’
<i>ag nŋ-</i>	SAY PERCEIVE	‘request, ask, inquire’
<i>ag weg-</i>	SAY HIDE	‘tell secretly’
<i>ag ask-</i>	SAY AVOID	‘reject, refuse’
<i>ag ay-</i>	SAY PUT	‘ask s.o. to remain, make an appointment’
<i>yn ag-</i>	SEE	‘make a noise while burning’

Table 17. Iconic ordering in Kalam complex verbs

Verb	Gloss	Meaning
<i>d ap-</i>	HOLD COME	'bring'
<i>d am-</i>	HOLD GO	'take'
<i>am d ap-</i>	GO HOLD COME	'fetch'
<i>ap tan jak-</i>	COME ASCEND REACH	'reach the top'

Table 18. Kalam complex verbs with two readings

Verb	Gloss	Meaning
<i>ap jak am-</i>	COME REACH GO	'come and go'
<i>ap jak am-</i>	COME REACH GO	'emerge'

'emerge', we find a right-headed derivation with two non-head elements preceding it, specifying the direction and telic properties of the event. We thus suspect that part of the Kalam serialisations may not be serial verbs but could be analyzed as head final derivations, analogous to the Abui verbs analyzed in this paper. In that case, Abui will not be the only language that forms new verbs by combining generic verbs with each other.

Abbreviations

1 = first person; 1PE = first person plural exclusive; 1PI = first person plural inclusive; 2 = second person; 3A = third person Actor; 3I = third person Undergoer, same referent as Actor; 3II = third person Undergoer, other referent than Actor; AD = addressee-based; AL = alienable; CPL = completive verb stem; CNT = continuative verb stem; DISTR = distributive; DST = distal; DUR = durative; INAL = inalienable; ICP = inceptive verb stem; LOC = undergoer prefix referring to locations, themes and benefactives; MD = medial; P = plural; PAT = undergoer prefix referring to patients; PFV = perfective; PL = plural quantifier; PRF = perfect; PROH = prohibitive; PRX = proximate; REC = undergoer prefix referring to recipient or goal; RED = reduplication; S = singular; SIM = simultaneous linker; SPC = specific; TOP = topic marker

Acknowledgments

We would like to thank Geert Booij and Ger Reesink for their comments on previous drafts of this paper. Michael Cysouw as editor of this volume helped

us further improve the paper. The Abui data presented comes from the northern dialect of Takalelang. The data has been collected during fieldwork in Alor from 2003 to 2009. Our thanks belong to the Abui community and to our consultants. In particular we want to thank the family of Bapak Timoteus Lanma and to Benidiktus Delpada who is our collaborator in documentation of Abui. The research reported here was conducted as part of the NWO Innovative Research Project 'Linguistic variation in Eastern Indonesia', Leiden University, 2002–2007. Research on Abui that took place after 2007 was supported by La Trobe University, Melbourne.

References

- Booij, Geert E.
2005 *The grammar of words: an introduction to linguistic morphology*. Oxford textbooks in linguistics. Oxford/New York: Oxford University Press.
- Davies, J.
1984 *Kobon*. (= Lingua descriptive studies; 3). Amsterdam: North-Holland.
- Holton, Gary, Marian Klamer and František Kratochvíl
2009 The Languages of Alor and Pantar (Eastern Indonesia): A (re)assessment. Paper presented at the Eleventh International Conference on Austronesian Linguistics (11. ICAL), Aussois, France, 22–26 June 2009.
- Kratochvíl, František
2007 *A grammar of Abui: a Papuan language of Alor*. Utrecht: LOT.
- Lang, Adrienne
1975 *The semantics of classificatory verbs in Enga (and other Papua New Guinea languages)*. (= Pacific linguistics; B-39). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- McGregor, William
2002 *Verb classification in Australian languages*. (= Empirical Approaches to Language Typology; 25). Berlin/New York: Mouton de Gruyter.
- Pawley, Andrew
1969 *The structure of Karam: a grammar of a New Guinea Highlands language*. Doctoral dissertation, University of Auckland.
- Pawley, Andrew
1987 Encoding events in Kalam and English: Different Logics for Reporting Experience. In: *Coherence and grounding in discourse: outcome of a symposium, Eugene, Oregon, June 1984*, R. S. Tomlin (ed.), 329–360. Amsterdam/Philadelphia: Benjamins.
- Pawley, Andrew
1991 Saying things in Kalam: Reflections on Language and Translation. In *Man and a half: essays in Pacific anthropology and ethnobiology in honour of Ralph Bulmer*, Andrew Pawley (ed.), 432–444. Auckland: Polynesian Society.

- Pawley, Andrew
 2001 The Proto Trans New Guinea obstruents: arguments from top-down reconstruction. In *The boy from Bundaberg: studies in Melanesian linguistics in honour of Tom Dutton*. Tom Dutton et al. (eds.), 261–300. Canberra: Pacific Linguistics.
- Pawley, Andrew
 2004 Episodic event reports in Kalam: case studies from serial verb and multiclausal constructions. Unpublished manuscript.
- Pawley, Andrew
 2005 A dictionary of Kalam, with ethnographic notes. Unpublished manuscript.
- Pawley, Andrew and Jonathan Lane
 1998 From event sequence to grammar: serial verb constructions in Kalam. In *Case, typology, and grammar: in honor of Barry J. Blake*, Barry J. Blake, Anna Siewierska and Jae Jung Song (eds.), 201–227. (= *Typological studies in language*; 38). Amsterdam/Philadelphia: Benjamins.
- Ross, Malcolm
 2005 Pronouns as a preliminary diagnostics for grouping Papuan languages. In *Papuan pasts: Investigations into the cultural, linguistic and biological history of the Papuan speaking peoples*. Andrew Pawley and R. Attenborough (eds.), 15–66. Canberra: Pacific Linguistics.
- Schultze-Berndt, Eva
 2000 *Simple and complex verbs in Jaminjung: a study of event categorisation in an Australian language*. Nijmegen: Max Planck Institute for Psycholinguistics.
- Spencer, Andrew
 1991 *Morphological Theory: an introduction to word structure in generative grammar*. Cambridge, MA: Blackwell.
- Stokhof, W. A. L.
 1984 Annotations to a text in the Abui language (Alor). *Bijdragen tot de taal-, land- en volkenkunde* 140.1: 106–162.

Ideophones and templatic morphology in Totonac

Teresa McFarland

1 Introduction

Ideophones are a wide-spread but understudied phenomenon, with an early informal definition (see Voeltz and Kilian-Hatz 2001: 2) that continues to be useful:

A vivid representation of an idea in sound. A word, often onomatopoeic, which describes a predicate, qualificative or adverb in respect to manner, colour, sound, smell, action, state or intensity. (Doke 1935: 118).

A more formal definition is difficult because of the great variety ideophones, or expressive speech, exhibit cross-linguistically. Ideophones may belong to different word classes and have varying syntactic and semantic functions in different languages, but they generally share one or more typical phonological characteristics. These include reduplication, sound symbolism, onomatopoeia, violation of the phonotactic constraints of the language, unusual stress or tone patterns and the use of “wild” phonemes not found in the rest of the lexicon (Rhodes 1994). What is typical, then, is a difference from the rest of the grammar, and this different distribution and escape from normal constraints is often also apparent in the morphology, syntax and semantics. We might claim that we know ideophones when we see them; however the inherent expressiveness and “dramaturgic function” (Voeltz and Kilian-Hatz 2001: 3) of ideophones, noted by many researchers, are perceptible only to those with enough knowledge of the grammar of a given language to distinguish their unusualness in context.

Ideophones as single lexical items are found in most if not all languages — words like *smash*, *crunch*, and *whoosh*. Cross-linguistic data also reveals a tendency for ideophones to occur in sets, sometimes quite large, all sharing very similar phonology and semantics, as in English *clink clank clunk* (Rhodes 1994: 283). These sets sometimes give the appearance of being based on a template, sometimes with variations only in sound symbolic segments. Some examples are given in Table 1 on the next page. This paper discusses a rare example of this templatic tendency in ideophonic sets in Filo-

Table 1. Ideophonic sets

(a)	Lahu (Tibeto-Burman): <i>qu-li-chi'-li</i> , <i>ju-li-ja-li</i> , <i>dú-li-dà-li</i> 'dangling down, hanging free' (Matisoff 1994: 120)
(b)	Yoruba (Niger-Congo): <i>rógódó</i> 'large, round', <i>rogbodo</i> 'fine and plump' <i>roboto</i> 'fat' (Tucker Childs 1994: 183)
(c)	Finnish (Finnic): <i>kalkahta-a</i> , <i>kilkahta-a</i> , <i>kolkahta-a</i> 'sound of something hard striking something' (Jarva 2001: 113)
(d)	Ilocano (Malayo-Polynesian): <i>dipag</i> , <i>dipur</i> , <i>lipag</i> , <i>ripug</i> 'crumble' (Rubino 2001: 307)
(e)	Tuvan (Turkic): <i>fylyreef</i> , <i>folyreef</i> , <i>soluraaf</i> , <i>fuluraaf</i> 'river sounds, e. g. water trickling' (Harrison 2004: 204)

meno Mata Totonac¹ (Totonac-Tepehua, Mexico), that goes beyond the narrower linking of related meanings and forms often found in ideophones cross-linguistically. In the examples given in Table 1 and many others, a specific template is associated with a single sound or manner of movement, color or odor. In the Totonac case, an ideophonic morphophonological template exhaustively structures an entire semantic field, that of color terms. Other such templates are closely associated with three other semantic fields, in spite of the nearly complete absence of prosodic templates elsewhere in the language. Relatively little has been published on ideophones, and nothing on their involvement in the emergence of morphophonological templates. Thus I conclude that the Filomeno Mata Totonac case is unique or very rare.

2 Filomeno Mata Totonac ideophones

Filomeno Mata Totonac, one of eight Totonac languages spoken in eastern Mexico, is a head-marking, polysynthetic language with a very complex, mostly agglutinating morphology. The ideophonic categories in Filomeno Mata Totonac besides color terms are manner / sound of motion adverbials; odor / flavor terms; and descriptive adjectives.² Filomeno Mata Totonac ideophones are characterized by partial (in most categories) or total (manner / sound of motion adverbials) reduplication. This is a very salient trait, since reduplication is almost entirely absent outside the realm of expressive speech. The exception is a handful of time expressions such as *kaat-kaat* 'annually' < *kaat* 'year', and *čaali-čaali* 'daily' < *čaali* 'tomorrow'; this process appears not to be synchronically productive. There is no other evidence of any type of prosodic template in Filomeno Mata Totonac. Other charac-

Table 2. Consonant inventory

	Labials	Alveolars	Velars	Uvulars
Plosives	<i>p</i>	<i>t</i>	<i>k</i>	<i>q</i>
Affricates		<i>ts, tl, č</i>		
Fricatives		<i>s, š, ɬ</i>	<i>x</i>	
Nasals	<i>m</i>	<i>n</i>		
Liquids		<i>l, r</i>		
Glides	<i>w</i>		<i>j</i>	

Table 3. *ɬ*-augmentatives

<i>sqawi</i>	‘he bends it’	<i>ɬqawi</i>	‘he bends something thick’
<i>snapu</i>	‘he covers it’	<i>ɬnapu</i>	‘he covers it with s. t. big or heavy’
<i>snat</i>	‘he embraces it’	<i>ɬnat</i>	‘he embraces it brusquely’

teristics of these ideophones include sound symbolism, complex word-final consonant clusters, unusual stress patterns, and restricted distribution.

Sound symbolism is especially prominent in ideophones, but also has sporadic productivity throughout the lexicon. Three sound symbolic intensity series exist in the language: a fricative series: /s, š, ɬ/; a mostly affricate series: /t, ts, tl, č/; and a velar-uvular stop series: /k, q/ (all common series cross-linguistically; see Nichols 1971). For comparison, the full consonant inventory is given in Table 2; the rhotic is a rare phoneme. Note that all of the coronal obstruents are involved in symbolism.

Non-ideophonic examples of sound symbolism are pairs such as *tsiits* ‘warm’ and *čiič* ‘hot’; *suu* ‘peel s. t. (with thin peel)’ and *šuu* ‘peel s. t. with thick peel’; *stunk* ‘straighten’ and *štunq* ‘stretch’; and *čísiti* ‘hair(s)’ and *tsísiti* ‘little hair(s)’. Generally the fronter sound symbolic segments in each series have a neutral value and the more back ones a diminutive or augmentative value, but this is not always true. Semantic variation occurs particularly in the intensity value of back segments; for example, among color terms *ɬ*- usually signals a lighter shade, as will be seen in Section 2.4, while among verb pairs, *ɬ*- often serves as an augmentative. Thus synchronically there is no clear directionality in intensity in any of these series (cf. Table 3).

The next subsections will address in turn each of the ideophonic categories of Filomeno Mata Totonac, in the order of increasing specificity of templatic restriction.

2.1 Sound / manner of motion adverbials

These expressive adverbials form the largest category of ideophones, with over 100 examples collected thus far. They are structurally unusual in their reduplication, stress pattern and restricted distribution in comparison with other adverbs. Most are immediately recognizable due to their total reduplication, rarely found elsewhere in the language, as previously noted. Sound / manner adverbials appear in discourse preceding a general verb of moving or making sounds, such as *'an* 'go', *min* 'come', or *wan* 'say', for example: *ʔqunqʔ ʔqunqʔ kinkawán* 'snore snore he says with his nose', or *pim pim ta'án* 'they go hopping along'. Non-ideophonic adverbs have a freer distribution, normally able to either precede or follow the verb. The ideophones are also recognized by their stress pattern, with each reduplicant (and the verb) receiving primary stress. In contrast, other adverbs in preverbal position generally cliticize to the verb and are thus unstressed. Sound symbolism is occasional in this category; for example, *kapa kapa* is the sound of the hooves of a smaller animal, and *qapa qapa* of a larger animal. Some examples of adverbials are provided in Table 4 on the facing page, beginning with two examples of non-reduplicated adverbials that refer to single abrupt sounds or motions. The final entries in the table exemplify items that are normally used in triplication.

Several generalizations are evident from the table. Most of the adverbials are onomatopoeic and describe a continuing activity. They fall into two shapes, with each reduplicant having either a single closed syllable, or a CVCV disyllable with a (usually) single vowel melody and a voiced final vowel (unstressed final vowels are regularly devoiced in Filomeno Mata Totonac). A prosodic template can be given for the sound and manner adverbials:

(F)CV₁(n)C(F or V₁) where F = fricative, C = any consonant

Figure 1. Template for adverbials

Either template is reduplicated (or triplicated), and primary stress is found on the first syllable of each reduplicant. Among the monosyllabic adverbials, 43 % (33 of 65) have a complex coda of a nasal-stop, stop-fricative, or nasal-stop-fricative cluster. Complex codas are relatively rare outside the expressive vocabulary, being found in approximately 5 % of the non-ideophonic roots.

Table 4. Sound / manner adverbials

<i>piinkš</i>	‘overheated china cracks’
<i>wax</i>	‘a whole bucket of water hits someone’
<i>čaqł čaqł</i>	‘footsteps in a puddle’
<i>čimpš čimpš</i>	‘blinking’
<i>łqunqł łqunqł</i>	‘snoring’
<i>łtank łtank</i>	‘yanking (a rope)’
<i>luqš luqš</i>	‘walking with wet clothing’
<i>lup lup</i>	‘gulping a liquid’
<i>paks paks</i>	‘applause, falling in water’
<i>pim pim</i>	‘hopping’
<i>piqš piqš</i>	‘a leaf burns or rustles’
<i>slank slank</i>	‘water drips’
<i>tił tił</i>	‘water rushes, or grains are poured’
<i>tuks tuks</i>	‘poking’
<i>šlit šlit</i>	‘slipping’
<i>štunq štunq</i>	‘a weight bounces on an elastic thread’
<i>šun šun</i>	‘a car or hummingbird passes by quickly’
<i>lama lama</i>	‘a flame flickers’
<i>kapa kapa (qapa qapa)</i>	‘a burro (large horse) clops along’
<i>muči muči</i>	‘worms squirm around’
<i>pili pili</i>	‘rolling like a stick’
<i>swinki swinki</i>	‘a very thin person walks along’
<i>luqš luqš luqš</i>	‘waves slap against a boat’
<i>qu’ qu’ qu’</i>	‘knocking on a door’
<i>turúnq turúnq turúnq</i>	‘bells ringing’

Note that this template, when all optional positions are filled, as in *łqunqł łqunqł* ‘snoring sound’, represents the maximal syllable for monomorphemes in Filomeno Mata Totonac. This template is less elaborated than that of other ideophones, as will be seen shortly.

2.2 Descriptive adjectives

Another category of ideophones, or perhaps incipient ideophones, is descriptive adjectives. Only a sub-set of such adjectives has ideophonic characteristics, including partial reduplication and mostly single vowel melodies. Sound symbolism is relatively rare in this group too (note *spilili* and *łpilili* ‘spotted’). Some examples are included in Table 5 on the next page, and for comparison, some common non-ideophonic descriptive adjectives are provided

Table 5. Descriptive adjectives

<i>slamáma</i>	‘shiny’
<i>slipípi</i>	‘shiny’
<i>skawáwa</i>	‘dry’
<i>spuxúxu</i>	‘straight’
<i>staqáqa</i>	‘sharp’
<i>lasása</i>	‘thin’
<i>latáta</i>	‘thin’
<i>tsatáta</i>	‘soft’
<i>pamáma</i>	‘soft, yielding’
<i>sqawíwi</i>	‘cool’
<i>šqawíwi</i>	‘insipid’
<i>mulúlu</i>	‘indented’
<i>čimáma</i>	‘fuzzy’
<i>łtulúlu</i>	‘thick’
<i>qulúlu</i>	‘spherical’
<i>šalála</i>	‘full of holes’
<i>skalála</i>	‘intelligent’
<i>spilíli, łpilíli</i>	‘spotted’
<i>smalála</i>	‘dark-skinned’

Table 6. Non-ideophonic adjectives

<i>lánka</i>	‘big’
<i>tináx</i>	‘small’
<i>tsee</i>	‘good’
<i>łmaan</i>	‘long, tall’
<i>aqtsúu</i>	‘short’
<i>qoolúu</i>	‘old’
<i>sáasti</i>	‘new’
<i>túwa</i>	‘difficult’
<i>łapóqo</i>	‘fat’
<i>páta</i>	‘hard’

(F)C₁V₁C₂V₂C₂V₂(V₁ usually the same as V₂)

Figure 2. Template for descriptive adjectives

in Table 6. The partial reduplication of the forms in Table 5 is very salient in Filomeno Mata Totonac. Except for the color terms (see Section 2.4) and these descriptive adjectives, there are only 10 (out of a 2,300-word lexicon) other monomorphemic words with two adjacent identical syllables; only four of these share the adjectival template given in Figure 2.³ This template has three CV syllables, with the second two identical, medial stress, and generally a single vowel melody throughout. It is similar to, but less restricted than the color template, as will be discussed in Section 2.4.

2.3 Odor / flavor terms

Odor and flavor terms make up the third category of Filomeno Mata Totonac ideophones. Table 7 gives a complete listing of all such terms collected to

Table 7. Odor / flavor terms

	Odor / flavor of:
<i>čiiq</i>	‘burnt beans’
<i>čiiik</i>	‘burnt beans, hair, or feathers’
<i>tsiik</i>	‘scented soap’
<i>skunk</i> , <i>ɬkunk</i>	‘egg; fish; dog; blood’
<i>ɬqunq</i>	‘beef or mutton’
<i>haqɬ</i>	‘urine’
<i>hakš</i>	‘spoiled cornmeal; onion; mildewed cloth’
<i>haks</i>	‘burning dry chile; orange peel’
<i>muq</i>	‘mildew or petroleum’
<i>muks</i>	‘flowery’
<i>mukɬ</i>	‘overly flowery’
<i>siqs</i>	‘sweet’
<i>šiḡš</i>	‘sweetish-salty’
<i>squuq</i>	‘salty’
<i>ɬkuukúk</i>	‘taste / irritation of green banana’
<i>sqaaqáq-wa</i>	‘potato peel’
<i>skíixa</i>	‘tortilla or avocado’
<i>škúuta</i>	‘bitter, sour’
<i>skúuta</i>	‘slightly acidic’
<i>šqúuta</i>	‘acidic’
<i>sqúuta</i>	‘a little acidic, taste of a clover plant’
<i>šúuni</i>	‘bitter’

(F)CV₁(n)K(V₁ or F) where F = fricative, C = any consonant, K = *k* or *q*

Figure 3. Template for odor / flavor terms

date, divided into morpho-phonologically and semantically (per the consultants) related sets.

Many of these terms fit a fairly detailed template: single closed syllables with an initial or final segment, or both, drawn from the sound symbolic series *s*, *š*, *ɬ*; *t*, *č*, *ts*, *tl*; or *k*, *q*; most have /*k*/ or /*q*/ in the coda, cf. Figure 3.

The odor / flavor template is similar to the one for the monosyllabic (before reduplication) sound and manner adverbials (see Figure 1 on page 238).

Also similar is the occurrence of complex codas in ten of these terms, a percentage significantly higher than the 5% found in the non-ideophonic lexicon. The final 8 items in Table 7 depart more or less from this template, although *skíixa* ‘taste of tortilla or avocado’ may derive from an earlier form whose final consonant was /q/, and the words for ‘taste/irritation of green banana’ and ‘taste of potato peel’ fit the template except for apparent partial reduplication (-*wa* is an adjectivizing suffix). So, much like with the descriptive adjectives, only part of this semantic field is structured by the template. We will now turn to a semantic field that exhaustively adheres to a morphophonological template, with a single exception.

2.4 Color terms

Color terms are the most strictly templatic semantic area among the ideophones. A complete list of the major terms is included in Table 8.

The language also makes use of a large number of minor color terms, all of them derived from the major terms, usually through a change in the initial sound symbolic consonant. Examples are *ɪnapapa* ‘pale (face)’ < *snapapa* ‘white’; *tlitliqi* ‘gray-black’ < *tsitsiqi* ‘black’; *čučuqu* ‘reddish’ < *tsutsuqu* ‘red’; *ɪmukuku* ‘light yellow’ < *smukuku* ‘yellow’; *xpupuqu* ‘grayish’ < *spupuku* ‘gray/blue’; and *skayiw’a* light green < *škayiw’a* ‘green’. Another term, *ɪmumuqu* meaning ‘fair, non-indigenous’ may have derived from *smukuku* ‘yellow’ with changes in other than the initial consonant that nevertheless allow it to maintain a shape similar to the other color terms. In fact, all of the color terms except *škayiw’a* ‘green’ share a well-elaborated template, cf. Figure 4 on the next page. All consist of three CV syllables, two of

Table 8. Major color terms

Templatic		Non-templatic	
<i>snapapa</i>	‘white’	<i>škayiw’a</i>	‘green’
<i>tsitsiqi</i>	‘black’		
<i>tsutsuqu</i>	‘red’		
<i>saqaqa</i>	‘white’		
<i>špinini</i>	‘red’		
<i>smukuku</i>	‘yellow’		
<i>spupuku</i>	‘gray/blue’		
<i>(s)pupunqu</i>	‘purple’		

(F)C₁V₁C_{1or2}V₁(n)C₂V₁ where F = fricative, C = any consonant

Figure 4. Template for color terms

which are identical, i. e., show partial reduplication. The initial consonant is a member of one of the three sound symbolic “intensity” series; when preceded by a fricative, C₁ is always a nasal or bilabial. The remaining consonants are in most cases limited to labials and velars. A single vowel melody is found throughout. Stress is always medial. It is noteworthy that *no* non-color terms fit this particular morpho-phonological template. There is negative evidence to show that the template not only has a close association with the semantic field, but may restrict lexical additions to it. The borrowed noun for the orange (fruit), *laašuš* (< Spanish *naranjas*)⁴ is widely used for the fruit and has been for a very long time (as revealed by a loan word phonology characteristic of borrowing at an early period of contact with the Spaniards). However *laašuš*, which does not conform to the template, is not used for the color orange, in spite of the absence of a Totonac term.

How does a morpho-phonological template become so closely linked with a broad semantic field that it limits lexical additions to the field both through derivation and borrowing? Since the semantic field in question is color, one that has been the subject of much scrutiny by researchers in historical semantics, we may have a possible answer. Berlin and Kay (1969: 17–22) provide evidence that monomorphemic color terms tend to develop in a fixed order in all languages, with ‘black’ and ‘white’ emerging first, followed by ‘red’, then either ‘yellow’ or ‘green’, then ‘blue’ and finally, in no set order, ‘orange / purple / pink / gray’. In Filomeno Mata Totonac this would predict the order of emergence given in Table 9.

Both ‘red’ and ‘white’ have two terms in Filomeno Mata Totonac. Synchronically in Filomeno Mata Totonac, *tsutsuqu* is the basic term for red,

Table 9. Emergence of color terms

<i>saqaa</i> ‘white’		<i>tsutsuqu</i> ‘red’		<i>smukuku</i> ‘yellow’		<i>spupunqu</i> ‘blue’		<i>pupunqu</i> ‘purple’
<i>snapapa</i> ‘white’	>	<i>špinini</i> ‘red’	>	<i>škayiw’a</i> ‘green’	>	<i>spupuku</i> ‘gray / brown’	>	
<i>tsitsiqi</i> ‘black’								

while it is unclear which ‘white’ term is primary. Across the Totonac-Tepehua family, however, *tsutsuqu* and *snapapa* are the most widespread terms for ‘red’ and ‘white’ respectively. *tsutsuqu* ‘red’ appears likely to have been derived from *tsitsiqi* ‘black’, with a difference only in the vowel.

Let us make the plausible assumption that the first three color words diachronically in Totonac were *tsitsiqi* ‘black’, *tsutsuqu* ‘red’ and *snapapa* ‘white’, all in accord with the templatic structure described above. The accidental phonological and phonotactic similarities, as well as partial reduplication, among the earliest color terms may have somehow constrained the development of the later-emerging terms. The existence of the sound symbolic series, allowing the derivation of minor color terms primarily by the alternation in the initial consonant, would have strengthened the psychological reality of this semantically linked template. Little diachronic work has been done on the Totonac-Tepehua family; much more is needed to pursue this hypothesis. What is clear is that an unusual situation holds within the realm of color terms in Filomeno Mata Totonac, where only a single term escapes the requirement to conform to a highly specified template.

3 Conclusions

Morphophonological templates of greater or lesser specification have been active in the derivation of ideophonic words in four semantic areas in Filomeno Mata Totonac, to a degree not reported previously in other languages. The more typical case cross-linguistically involves a specific template associated only with a single semantic item, whereas in Filomeno Mata Totonac, such a template structured the entire area of color terms. It is unusual that ideophones, usually described as escaping many grammatical norms, in this language impose additional constraints on the phonology and morphology of certain semantic domains.

The hypothesis proposed is that reduplicative templates developed historically into templates with reduplication *and* slots for sound symbolic segments, and then further into morphophonological templates with the major features of all segments fixed or tightly constrained. The development from a reduplicative template to a more elaborated, broadly semantically linked template via sound symbolism is a logical progression, and therefore may not be limited to Filomeno Mata Totonac.

Notes

1. All data collected by the author in Veracruz, Mexico in 2003–2006. Great thanks are due to my consultants José Santiago Francisco, Miguel Jerónimo Laureano, María Agustina García Cortés, and Pascual Cortés López.
2. The categories are similar across the Totonac-Tepehua family; cf. Smythe Kung (2006) for a Tepehua language.
3. These exceptions are the nouns *tuququ* ‘hole’ and *sukuku* ‘small hole’; *tuuxuxu* ‘sack’; and *slukuku* ‘lizard’.
4. The derivation of /laašuš/ < Sp. /(na)ranxas/ may not be transparent. In early loan words, /r/ is borrowed as /l/ and /x/ as /ʃ/. Nouns for objects usually found in plural numbers are often borrowed in the plural, with Spanish final /s/ becoming /ʃ/.

References

- Berlin, Brent and Paul Kay
 1969 *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Childs, G. Tucker
 1994 African ideophones. In *Sound Symbolism*, Leanne Hinton, Johanna Nichols and John J. Ohala (eds.), 178–204. Cambridge: Cambridge University Press.
- Doke, Clement Martin
 1935 *Bantu linguistic terminology*. London: Longmans, Green.
- Harrison, K. David
 2004 South Siberian Sound Symbolism. In *Languages and Prehistory of Central Siberia*, Edward Vajda (ed.), 199–214. Amsterdam: Benjamins.
- Hinton, Leanne, Johanna Nichols and John J. Ohala
 1994 *Sound Symbolism*. Cambridge: Cambridge University Press.
- Inkelas, Sharon
 1998 The theoretical status of morphologically conditioned phonology: a case study from dominance. *Yearbook of Morphology* 1997: 121–155.
- Jarva, Vesa
 2001 Some expressive and borrowed elements in the lexicon of Finnish dialects. In *Ideophones*, Erhard Friedrich Karl Voeltz and Christa Kilian-Hatz (eds.), 111–119. Amsterdam: Benjamins.
- Matisoff, James A
 1994 Tone, intonation and sound symbolism in Lahu: loading the syllable canon. In *Sound Symbolism*, Leanne Hinton, Johanna Nichols and John J. Ohala (eds.), 115–129. Cambridge: Cambridge University Press.
- Nichols, Johanna
 1971 Diminutive consonant symbolism in western North America. *Language* 47: 826–848.
- Rhodes, Richard
 1994 Aural images. In *Sound Symbolism*, Leanne Hinton, Johanna Nichols and John J. Ohala (eds.), 276–292. Cambridge: Cambridge University Press.

Rubino, Carl

- 2001 Iconic morphology and word formation in Ilocano. In *Ideophones*, Erhard Friedrich Karl Voeltz and Christa Kilian-Hatz (eds.), 303-320. Amsterdam: Benjamins.

Smythe Kung, Susan

- 2006 Simbolismo de sonido y el lenguaje expresivo en el tepehua de Huehuetla. In *Memorias del VIII Encuentro Internacional de Lingüística en el Noroeste*, vol. 3. María del Carmen Morúa Leyva (ed.), 331-354. Hermosillo Son.: Universidad de Sonora, Hermosillo, Fonética y Fonología, Lenguas Mexicanas: Tepehua.

Voeltz, Erhard Friedrich Karl and Christa Kilian-Hatz

- 2001 *Ideophones*. Amsterdam: Benjamins.

Subtractive plural morphology in Sinhala

Eike Nitz & Sebastian Nordhoff

1 Introduction

Number marking is a prime example of constructional iconicity (Haiman 1985; Mayerthaler 1988; Newmeyer 1998). The idea is that the plural form of lexemes will consist of more segments than the singular form because the plural also denotes more entities than the singular form.

This paper discusses the Sinhala number system, where this is not the case for [–ANIMATE] lexemes. After giving an overview over the typology of number marking (Section 2), we argue that Sinhala is a language with obligatory number marking (Section 3), before presenting the data (Section 4) and the analysis (Section 5). Section 6 delves into language history and describes how this unfamiliar system came into being, while Section 7 brings up and discusses the notion of iconicity. An overview of similar systems is presented in Section 8, before going more deeply into iconicity matters and the complicating issue of case marking (Section 9), which form the last section before the conclusion.

2 Types of number systems

Corbett (2000) distinguishes several types of number marking according to the presence or absence of a specialized form for *singular reference*, *plural reference* and *general reference*.¹ *Singular reference* means reference to exactly one item, *plural reference* means reference to more than one item and *general reference* means reference to any number of items.

While the terms *singular* and *plural* are widely used everywhere, this is less so with the term *general reference*. This may be due to the fact that the major Western languages do not use this concept. General reference means that reference is made to the entity designated by the lexeme, but that the exact number of referents is not important. General reference can be said to be underspecified for number. An example of this is the following Japanese sentence in (1).

- (1) *Kooen ni wa inu ga iru rasii*
 park in TOPIC dog SUBJ be seem

‘It seems there is a dog / are some dogs in the park.’ (Corbett 2000: 14)

The amount of dogs involved is not specified, and need not be. The sentence is perfectly fine if the speaker has seen one dog, and is none the worse if she has seen several dogs. A form with general reference can thus be used indiscriminately in singular and plural contexts. Note that in English, such a form for general reference does not exist. Speakers of English always have to specify the number of referents.²

Languages differ as to how many forms they have for these three kinds of reference. The most complete system has a separate form for each one of the three functions, Bayso is an example (Corbett and Hayward 1987). There we find *lúban* ‘lion.GENERAL’, *lubán-tit* ‘lion-SG’ and *luban-jool* ‘lion-PL’.³

Other systems conflate general reference and singular reference, which contrast with a form that is used for plural only. Corbett terms this system *general / singular vs. plural system*. Japanese is an example, as we have seen in (1), which is ambiguous. The sentence can refer to one dog or to several dogs. If the plural marker *tati* were present after *inu* ‘dog’, the reading could only be plural.

The mirror image of this system is the *general / plural vs. singular system*. This system contrasts a form which can be used for general or plural reference with a form that can only be used for the singular. The Cushitic language Arbore (Hayward 1984) is an example: *tíise* can mean ‘maize cob’ or ‘maize cobs’, but *tiis-in* can only have singular reference, i. e. ‘a maize cob’. The morpheme that marks the singular is sometimes called *singulative*, so we can call a system making use of such a morpheme a *singulative system*.

The last system, which is the most normal one for speakers of European languages, is a system where number must obligatorily be expressed. There is no form that leaves number unspecified. Let’s call this system *singular vs. plural system*.

3 Sinhala as a plural language

In this section we will discuss the general system of number marking in Sinhala. This will serve as a background for the discussion of the plural marking of inanimates, to be discussed thereafter.

Sinhala is a language of the *singular vs. plural* type. Expression of number is obligatory in Sinhala, as it is in English. There is no way to leave the feature NUMBER unspecified, as the following sentences show.⁴

- (2) *ball-ā inn-ē vatt-ē*
 dog-SG exist.ANIM-EMPH garden-LOC
 ‘The dog is in the garden.’

- (3) *ball-ō inn-ē vatt-ē*
 dog-PL exist.ANIM-EMPH garden-LOC
 ‘The dogs are in the garden.’

Unlike the Japanese sentence in (1), neither of the sentences above can be ambiguous. Each of them has precisely one possible reference, singular in the first case, plural in the second. This is true not only of animates, but also of inanimates.

- (4) *pāra-∅ tiyenn-ē vatt-ē*
 street-SG exist.INANIM-EMPH garden-LOC
 ‘The street is in the garden.’

- (5) *pāra-val tiyenn-ē vatt-ē*
 street-PL exist.INANIM-EMPH garden-LOC
 ‘Streets are in the garden.’

We see that the Sinhala number system is very much like the English system in that number must be expressed and cannot be left unspecified.

It is generally accepted in markedness theory that in such a system the singular is the unmarked form (e. g. Jakobson [1939] 1971). This is also true in Sinhala. Unmarked forms are, among other things, characterized by the fact that they have a wider array of distinctions than marked forms. This is also the case in Sinhala. In Sinhala, animacy and definiteness must be expressed in the singular, but cannot be expressed in the plural. The smaller number of categories expressed in the plural shows that plural is the marked case.

- (6) *ball-ā : ball-ek : ball-ō*
 dog-SG : dog-ANIM.INDEF.SG : dog-PL
 ‘the dog’ : ‘a dog’ : ‘(the) dogs’

- (7) *pāra-∅ : pāra-ak : pāra-val*
 street-SG : street-INAN.INDEF.SG : street-PL
 ‘the street’ : ‘a street’ : ‘(the) streets’

Table 1. Plural marking on Sinhala inanimates

	<i>i</i>	<i>iia</i>	<i>iib</i>	<i>iii</i>	<i>iv</i>	<i>v</i>	
singular	-VCa	-Vya	-Vva	-VCCa	eka	-Xa	(X=d, t, r)
plural	-VC	-V	-V	$-VC \begin{Bmatrix} u \\ i \end{Bmatrix}$	∅	Xa-val	
difference	-a	-ya	-va	-C-	eka	(-val)	
<i>examples:</i>							
singular	<i>pota</i>	<i>kuḍaya</i>	<i>vārtāva</i>	<i>kekka</i>	<i>bas eka</i>	<i>pāra</i>	
plural	<i>pot</i>	<i>kuḍa</i>	<i>vārtā</i>	<i>keki</i>	<i>bas</i>	<i>pāra-val</i>	
gloss	book(s)	basket(s)	report(s)	pole(s)	bus(ses)	street(s)	

In the last column, we see that definiteness is not expressed in the plural, whereas it is expressed in the singular, as the first two columns show. In the middle column, we see the difference in animacy expressed by *-ek* for animates, as opposed to *-ak* for inanimates. This distinction is impossible to make in the plural. We conclude that Sinhala is no exception to the rule that the singular is the unmarked form and the plural the marked form.

4 The plural of inanimates

While the discussion up to now has not revealed any surprising phenomena, we turn now to the main point of this paper, i. e. the plural of inanimates. We will argue that inanimates have the singular as the basic form and that the plural is derived by subtractive morphology.

Sinhala inanimate nouns can be divided into six classes, which can be distinguished on the basis of the phonological shapes of the singular and the plural forms as in (Table 1), based on a synthesis of Matzel (1983), Jayawardena-Moser (2004), and Karunatillake (2004). All of these authors either collapse several classes or leave out class *iv*.

We see that in classes *i* through *iv*, the plural is always shorter than the singular. Only in class *v* do we find the inverse situation. However, the distinctions of animacy and definiteness are still maintained in the singular, and not in the plural.

- (8) *pota* : *potak* : *pot*
 ‘the book’ : ‘a book’ : ‘(the) books’

Let us now discuss these classes in turn.

4.1 Class *i* in CVCa#

This class is characterized by a final *a* preceded by a single consonant. It is the largest class. Number is distinguished by the final *-a*, which is not present in the plural form.

- (9) *pota* : *pot*
book : books

This rule is fully productive for words ending in CVCV such as *liyuma* / *liyum* 'letter(s)', *rupiyala* / *rupiyal* 'rupee(s)', *gamana* / *gaman* 'trip(s)'. Word-final *-h* is not permitted and normally changed to *-s*.

- (10) *gaha* : *gas*
tree : trees

4.2 Class *iia* in CV+va#

Another class consists of the words ending in *-va* in the singular. The plural does not show *-va*.

- (11) *navakatāva* : *navakatā*
novel : novels

- (12) *kantōruva* : *kantōru*
office : offices

4.3 Class *iib* in C+aya# / C+ee#

Similar to the words ending in *-va* are the words ending in *-ya* in the singular. This is not present in the plural

- (13) *satiya* : *sati*
week : weeks

In opposition to *-iya*, the historical ending *-aya* has undergone erosion and is now *-ē*. The plural has not changed, though. It differs from the singular by ending in *-a* instead of *-ē*.

- (14) *pattarē* : *pattara*
newspaper : newspapers

The difference between class *iia* and class *iib* is that the approximant is palatal in one and labiodental in the other. This is fully predictable on phonological grounds: *-va* always follows the back vowels *o*, *u*, and long *ā*, while *-ya* follows the front vowels *e* and *i*, and short *a*. We can say that they are a part of the superclass CVGa#, where G is a glide.

4.4 Class *iii* in CCV#

This group shows a double consonant in the singular and a single consonant in the plural. The last vowel of the plural corresponds to the stem vowel. It is realized as *i* if it follows a front vowel and *u* if it follows a back vowel:

- (15) *kekka* : *keki*
pole : poles

- (16) *avurudda* : *avurudu*
year : years

The most common type in this class are lexemes with geminate consonants in the singular, but the sequence nasal+stop does also occur. In this case, the nasal is reduced to a so-called *semi-nasal* in the plural.

- (17) *kanda* : *kaṇḍu*
[kand̪ə] : [kaṇḍu]
hill : hills

4.5 Class *iv* in C+eka

These words are English loanwords whose traditional British singular form is used for plural reference in Sinhala. The singular is marked by the morpheme *eka*.

- (18) *bas eka* : *bas*
bus : busses

- (19) *travalīṅ bāḡ eka* : *travalīṅ bāḡ*
travelling bag : travelling bags

While *eka* can mean ‘one’ in other contexts, it should not be mistaken for the indefinite article. In fact, *bas eka* can only be definite. For indefinite reference, *bas eka-k* must be used.

4.6 Class *v* in CVXa#

This class is defined by a singular form whose last consonant before the final *a* is a consonant that may not occur word-finally. These consonants are notably the retroflex stops and /ɾ/. The plural form is then the singular form plus the suffix *-val*.

(20) *raṭa* : *raṭaval*
country : countries

(21) *pāra* : *pāraval*
road : roads

5 Analysis

There are two possible ways to analyse these data: either the singular is primary, and the plural is derived, or the plural is primary and the singular derived. We will first discuss these hypotheses and then proceed to evaluating them on the basis of internal consistency, consistency with other phenomena in the number domain, and parsimony.

5.1 Hypothesis 1: the plural is derived from the singular

This hypothesis would be our first guess since we know that animate lexemes behave that way. A set of rules that describes number marking on Sinhala inanimates without ambiguity is as follows:⁵

(22) The plural is formed by stripping the singular of its last segment unless this would result in a phonologically unacceptable form (final double consonant or *r*, [+RETROFLEX]). (*pota* → *potə*, *satiya* → *satiyə*, *kantōruva* → *kantōruvə*, *bas eka* → *bas ekə*)

The last segment in this case is *-a*, *-ya*, *-va* or *eka*, as the case may be.⁶ It is possible that the application of rule (22) yields results that are phonologically not acceptable. We have to distinguish two cases: If rule (22) fails because of the phonological unacceptability of final double consonants, then the following rule applies:

(23) The plural is formed by reducing a double last consonant and replacing the final *a* by *i* for stems in *e* and *i*, or by *u* in the other cases (*kekka* → *keki*).

If rule (22) fails because of the phonological unacceptability of final /*r*/ or [+retroflex], then the following rule applies:

(24) The plural is formed by adding *-val* (*pāra* → *pāraval*).

5.2 Hypothesis 2: the singular is derived from the plural

This is also a plausible hypothesis, especially given that Hypothesis 1 makes use of subtractive morphology, which is very rare in the languages of the world. The starting point of this hypothesis would be that the plural is the base form and that the singular is derived by some sort of morphological process. For classes *i-iv*, this would be an additive process, and for class *v*, a subtractive one.

How can the singular be derived from the plural for classes *i-iv*? The set of rules would look like something close to the following:

- (25) a. If the plural ends in a consonant, add *-a* to form the singular.
 b. If the plural ends in a front vowel, add *-ya* to form the singular.
 c. If the plural ends in a back vowel, add *-va* to form the singular.

This nicely explains classes *i* and *ii*. However, for the following classes, we run into problems. For the plural noun *keki* ‘poles’ we would predict that the singular is *kekiya*, based on rule (25b). But this is not correct, the singular has to be *kekka*. There is also no phonological conditioning involved as can be seen on examining *redda/redi* ‘piece of clothing’ (SG. / PL.) and the phonologically very similar *geḍiya/geḍi* ‘fruit’ (SG. / PL.). In the plural, their phonological shape is very similar, yet they differ in the singular.

In the same vein, the pairs *gaha/gas* ‘tree / trees’ and *bas eka/bas* ‘bus / busses’ illustrate that lexemes in class *i* and *iv* also have to carry specification of the class they are in. This means that we have to restate our rules.

- (26) a. If the plural ends in a consonant and the lexeme is in class *i*, add *-a* to form the singular.
 b. If the plural ends in a front vowel and the lexeme is in class *ii*, add *-ya* to form the singular.
 c. If the plural ends in a back vowel and the lexeme is in class *ii*, add *-va* to form the singular.

We then add

- d. If the lexeme is in class *iii*, the last consonant is doubled and the final vowel is replaced by *a*.
 e. If the lexeme is in class *iv*, the singular is formed by adding *eka*.
 f. If the plural ends in *-val*, then the singular will not have *val*.

As for class *v* ending in *val*, there are two possibilities: a) the plural form is seen as primary and the singular is derived by subtractive morphology. The weakness of this analysis is that there is still subtractive morphology required. In this respect, there is no improvement over Hypothesis 1. Or, b), we have to assume that for lexemes in class *v*, the singular is primary, and not the plural. This means that Hypothesis 2 is internally inconsistent as to the status of singular or plural as primary.

5.3 Discussion

Hypothesis 1 assumes the singular as basic throughout all the classes discussed. It is therefore internally consistent. It is also consistent with number formation in other parts of the grammar. Remember that for animates, we also assumed the singular as basic and the plural as derived (Examples (2)–(5)). This hypothesis also goes well with the markedness effects we observe for definiteness and animacy (Examples (6)–(7)). We observed that the singular shows more categories (definiteness and animacy) than the plural and therefore assumed that the singular was basic. Another advantage of Hypothesis 1 is that it can explain all the forms drawing on phonological information only. The plural can be predicted by simply looking at the phonological shape of the singular. As a drawback, Hypothesis 1 relies on subtractive morphology, which is a very unusual morphological process, especially for number marking (cf. Corbett 2000: 150–151).

Hypothesis 2a on the other hand has the advantage of using no subtractive morphology. But it assumes that the singular is basic in the animate domain, while the plural is basic in inanimate domain (but then again not so in class *v*). It is thus neither internally consistent as far as inanimates are concerned, nor consistent with number marking in the animate domain. It furthermore has to rely on lexical information to produce the correct singular form. For most of the lexemes, their class membership has to be stored in the lexicon. Hypothesis 2b combines the drawbacks of Hypotheses 1 and 2a in having both subtractive morphology and a requirement for lexical specification. For reasons of consistency and parsimony, Hypothesis 1 is to be preferred.

5.4 Sinhala subtractive plural as a *rarum*

If we accept Hypothesis 1, then Sinhala is unique in making use of subtractive morphology for plural formation, a phenomenon explicitly stated by Corbett

(2000) as not attested in the over 250 languages he surveyed. In case we still found Hypothesis 2 more appealing, despite the weaknesses outlined above, Sinhala's number marking strategy would still be a *rarum* because it would be the only *singular vs. plural* language that makes use of a singulative morpheme. All other languages that derive the singular from another form are of the *general / plural vs. singular* type.

6 A look into language history

The discussion has been purely synchronic up to now and has tried to give a good account of the Sinhala number system as it works today. One interesting question that will be discussed in this section is how this system actually came into being. This discussion will focus on the development of number marking on inanimates, where we see the development of a *general / plural vs. singular* system, followed by a reanalysis of this system to the *singular vs. plural* system we find today. The development of number marking with animates on the other hand is unspectacular. Animates were always marked in a *singular vs. plural* way, the only things that changed were the actual suffixes to express this distinction. These suffixes underwent some phonological changes, but these did not jeopardise the fundamental opposition between singular and plural.

6.1 From Middle Indic to Dutch borrowings

Accessible data on Sinhala language history is scarce, the best resource being Geiger (1973), who is not overly explicit on the matter of number marking. It appears that the original Middle Indic number distinction of declension II was lost before a *general / plural vs. singular* system was introduced. This system was eventually reanalyzed as a subtractive *singular vs. plural* system.

In Middle Indic there were affixes that denoted singular and affixes that denoted plural. By attaching these to a stem, the respective reference to number could be established. But in Sinhala the affixes denoting plural were eventually lost in declension II (which is the ancestor of today's inanimate classes *i*, *ii*, *iii* and *v*). It was the bare stems that came to be used for plural reference. As an example, 'hand' had *at-a* for NOM.SG and *at-∅* for NOM.PL⁷ (Geiger 1973: 541). Geiger does not interpret *at* as plural, but rather as collective. He gives *Gehände*⁸ as the literal translation and not *Hände*.

This historical system is characterized by a stem form that can be used as such with a collective interpretation and a NOM.SG morpheme *-a*.⁹ From there, it is only a small step towards the reanalysis of *-a* as a singulative morpheme. And indeed, later, the use of *-a* to mark NOM.SG was generalized for all neuter nouns (Geiger 1973: 509). At this stage, Sinhala showed a *general / plural vs. singular* system, comparable to the Arbore system mentioned in the Introduction.

This new overt marking of the singular can nicely be seen in reborrowings, mostly scholarly, from Sanskrit (*tatsama*). These are borrowed in their Sanskrit singular form, but with collective meaning. To get singular meaning, *-a* (or the phonologically conditioned allomorphs *-ya* or *-va*) had to be used. For instance, Sanskrit *grahaṇa* ‘eclipse’ is borrowed as Sinhala *grahaṇa* ‘eclipses’. The singular is *grahaṇa-ya*. Since the stems are late borrowings from Sanskrit, they have not undergone the deletion of final vowels of Early Sinhala. These borrowings of the archaic form can thus easily be identified by the final vowel in the plural / collective and the intervening glide in the singular.

After the Portuguese (1503) and the Dutch (1656) arrived in Ceylon, Sinhala also borrowed words from the colonial languages. Again, the borrowed forms are singular in the source language, but collective in the borrowing language. So, the Portuguese word *sapato* ‘shoe’ became Sinh. *sapattu* ‘shoes’ with the singular form *sapattu-va*. For a Dutch example, we can again take *kantōru* ‘offices’ (SG. *kantōruva*) from Dutch *kantoor* office or *tē poocci* ‘tea pots’ (SG. *tē poocciya*) from Dutch *theepotje*.

To sum up, until the British ousted the Dutch around 1800, the Sinhala inanimate number system was characterized by a stem form denoting the plural and a singulative morpheme *-a* which is used to derive the singular. This singulative morpheme grammaticalized from the old NOM.SG marker of declension II, which spread to other lexemes, and also loanwords.

6.2 British loanwords

After the Dutch left Ceylon, a change in the number system took place. The meaning of the singulative suffix *-a* became obscure. The evidence for this is that English loanwords no longer take this suffix. When encountering the English word *guitar*, there was no a priori reason that the Sinhalese would not borrow it as **giṭāruva*, along the same pattern as they had done with Dutch *kantoor* some years earlier, which became *kantōruva*. Yet this was not the case. *Guitar* was borrowed as *giṭār eka*.¹⁰

We do of course not know when exactly the English word *guitar* was borrowed. But we know that during Dutch rule, colonial loanword integration still followed the pattern used for Sanskrit *tatsama*, whereas this was no longer the case for the English loanwords borrowed during British rule and afterwards. There is a total absence of Dutch borrowings in class *iv* (*eka*) and a near-total absence of English borrowings in class *ii* (*-ya* and *-va*). This neat separation suggests that the shift in the number marking strategy was not too distant from the shift in political power.

It must have been around the beginning of the 19th century that Sinhala switched from the additive singulative system, which had prevailed at least until the end of the Dutch rule, to the subtractive plural system we find nowadays.

7 Iconicity

7.1 Plural marking as iconic

Having discussed the form and history of the Sinhala system, we now turn to the cognitive factors that might have contributed to its development. One thing that is striking about the Sinhala system is that the singular forms are nearly always longer than the plural forms, as far as inanimates are concerned. This goes against a well-observed opposite tendency in the languages of the world, which has motivated claims about the importance of iconicity in language change.

Plural is one of the most-cited examples of iconicity in morphology (Greenberg [1966] 1976; Mayerthaler 1988; Wurzel 1989). Examples can be found in many textbooks. The idea is that a higher number of specimens of a certain kind in the real world is mirrored by a higher number of sounds in language.

(27) *one dog*

(28) *two dogs*

Dog in example (27) consists of three segments, whereas *dogs* in example (28) consists of four segments. (28) surpasses (27) in both semantics and phonemics. There are cases of English plurals that are less iconic and non-iconic:

(29) *goose : geese*

(30) *sheep* : *sheep*

In (29) the number of segments is equal in singular and plural, three, but the difference in the vowel still indicates a difference in meaning. This is called *minimally iconic*. In (30), there is no difference between the singular form and the plural form. This is called *non-iconic*.

In French, there are some *counter-iconic* plurals, i. e. a more in number of specimens corresponds to a less in the number of phonemic segments:

(31) *œuf* : *œufs*

[œf] : [ø]
'egg' : 'eggs'

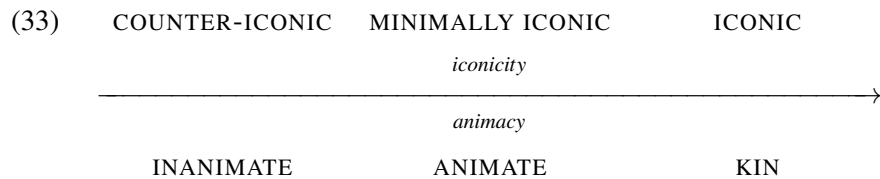
(32) *os* : *os*

[ɔs] : [o]
'bone' : 'bones'

Also in Welsh cases like these are reported (Stolz 2007). However, they seem to be marginal and are in no way a productive process.

7.2 Iconicity and animacy in the Sinhala plural system

We have discussed above that the bulk of the Sinhala inanimate nouns have a singular that is shorter than the plural. This is a counter-iconic way of plural-marking. Animate nouns like *ballā* / *ballō* 'dog / dogs' on the other hand normally show a plural form that is as long as the singular form. That form is minimally iconic in that two things that are semantically different are also different in their phonological representation, but while the difference in semantics is one of *quantity* of specimens, the difference in phonology is only one of *quality* of the vowel. Kin terms finally, like *ammā* / *ammālā* 'mother / mothers' have a plural form that is longer than the singular form (Matzel 1983: 62). This is an iconic form of plural marking. Corbett (2000) notes that the expression of number is sensitive to the Animacy Hierarchy. Often, number is only marked on lexemes high up in the hierarchy, and not marked on items further down. The Sinhala case can also be squared with the Animacy Hierarchy, but in a slightly different vein. Here, the higher up in the hierarchy, the more iconic the plural marking, the further down, the less iconic the plural marking. There is a direct correlation of animacy and iconicity.



8 Similar systems in other parts of the world

Sinhala is not the only language in the world with partly counter-iconic number systems. A clear example are Cushitic languages.¹¹

Discussing the semantics of the lexemes that do indeed have a *general / plural vs. singular* distinction in the Cushitic language Oromo (Borana dialect, Andrzejewski (1960)), Corbett (2000) observes that the singulative marker, if it is ever used at all, only attaches to stems that denote humans.¹² This means that the number marking for persons is counter-iconic. The remainder of the nouns show a *general / singular vs. plural* opposition which is iconic.

Similar to Oromo is the Papuan language Imonda (Seiler 1985). In that language, three kin terms denote plural or dual in their bare form, and a suffix is used for singular reference. This is an even more limited inventory of terms than in Oromo, but again it is remarkable that lexemes *up* the Animacy Hierarchy are marked for number in a counter-iconic way, and not the lexemes further down, just the opposite of Sinhala. We can say that Oromo and Imonda show an inverse correlation of animacy and iconicity. The more animate a lexeme, the less iconic the plural marking and *vice versa*.

As shown above, Sinhala has a direct correlation of animacy and iconicity. A language that parallels Sinhala in having this direct correlation is Kiowa (Watkins 1984). Kiowa has a very special system of marking the unexpected number:

- (34) *cê* : *cê.gò*
 horse : horses

Horses are ‘expected’ to be singular, the plural is unexpected, and this is marked by *gò*. Poles on the other hand are ‘expected’ to occur in higher quantities, this is why the unexpected number is the singular, which in turn is marked with *gò* (or an allomorph thereof, like *dò* in the following example)

- (35) *á.dò* : *á*
 pole : poles

We see that the marking that Kiowa uses for ‘horse’ is iconic, and the marking that it uses for ‘pole’ is counter-iconic. Virtually all animates behave like ‘horse’, whereas the behavior of inanimates is somewhat less clear, but there is still a majority that behaves like ‘pole’.

To sum up, there are two languages that have an inverse correlation of animacy and iconicity, Oromo and Imonda. In these languages, only a few lexemes take counter-iconic marking at all, and number marking is optional. These contrast with Sinhala and Kiowa, where there is a direct correlation between animacy and iconicity. These languages share the further characteristics that the counter-iconic plural marking is not marginal at all, and that number marking is obligatory. The higher importance of marking number in Sinhala and Kiowa and the higher total number of counter-iconic plural marking in these languages mean that a good case can be made for the direct correlation between animacy and iconicity. The optionality of number marking in Kiowa and Imonda on the other hand, paired with the very low frequency of counter-iconic marking in these languages mean that the inverse correlation between animacy and iconicity does not find strong support.

We conclude that we have found two languages where iconicity and animacy solidly align, whereas the two possible exceptions do not have the same weight.¹³ The aim of this paper is to present a *rarum* and not to speculate on the cognitive and functional forces that brought this about. Still, an explanation for the alignment of animacy and iconicity in Sinhala and Kiowa (and the absence of a strong example of the contrary) could be that human beings are much more likely to individuate things up the Animacy Hierarchy. Hence singular reference to FATHER is much more likely than singular reference to STONE. The reverse is true for reference to more than one exemplar. Since “grammars code best what speakers do most” (DuBois 1985: 362–363), it would be preferable to have a shorter coding for the form most commonly used in discourse, i. e. the singular form for animate things and the plural for inanimate things. As a correlation, the form less often used would be longer (plural for animate, singular for inanimate). This boils down to saying that animates should have a plural longer than the singular, which is iconic marking, and inanimates should not, which is non-iconic. As a final reformulation: animates should get iconic marking, while inanimates should not.¹⁴ This is what we find in Sinhala and Kiowa.

There are obvious shortcomings to the above explanations like animates most often talked about with plural reference (e. g. mosquitoes) or inanimates most often talked about with singular reference (e. g. home). We do not think

Table 2. Case suffixes in Sinhala

	Singular	Plural
Nominative	∅	∅
Genitive	-ē	-vala
Dative	-ṭa	-valaṭa
Instrumental	-eṁ / -iṁ	-valiṁ

that the coarse explanation for the correlation between animacy and iconicity we have outlined above can explain everything, but it seems an acceptable first approach to a *rarum*.

9 Iconicity through the back door

In the preceding section we have seen that Sinhala has counter-iconic marking for inanimate lexemes. This conclusion was arrived at by looking at the nominative. If we take into account nominal forms inflected for other cases, a different picture emerges.

Four cases are distinguished on Sinhala inanimates: Nominative, Genitive, Dative and Instrumental.¹⁵ The case suffixes are different for singular bases and plural bases.

The nominative (the citation form) is not overtly marked. Hence, there is no difference in the length of the suffix in the singular and the plural, both being zero. In this case, the counter-iconic pattern that the stems show is not changed by the suffix. In the other cases, however, the oblique case suffixes are longer in the plural than in the singular. As an example, the dative of *pota* ‘book’ is *pota-ṭa*, the plural is *pot* in the nominative and *pot-valaṭa* in the dative.¹⁶ In the nominative, the plural is one segment shorter in the singular, but in the dative, the plural is three segments longer than the dative singular form.

We noted in Table 1 on page 250 that the plural is normally 1–3 segments shorter than the singular in the citation form. Table 2 shows that the case suffixes are 1–4 segments longer in the plural than in the singular. On forming an oblique case in the plural, the extra length of the plural case suffixes thus compensates for the lack of iconicity in the stems. We see that for inanimates the number marking in the oblique cases is iconic, while this is not true for the nominative case. The change is brought about by material added to the suffix (and not to the stem). This is why we can speak about iconicity sneaking in

through the back door of the suffix, instead of properly taking its place on the base (inflected for number).

While it was possible to make an educated guess about the discourse motivations of counter-iconic number marking in the citation form, we seriously doubt that a good explanation can be found why this cognitive motivation would not hold in the oblique cases.

10 Conclusion

This paper has shown that Sinhala has a counter-iconic way of marking number on inanimates which is best analyzed as a subtractive morphological process. No other known language exhibits subtractive morphology as a regular means for number marking. This makes the Sinhala number marking system a *rarum*.

This system arose diachronically through a temporary loss of number distinction in declension II, which was later compensated by the exaptation of the nominative singular marker *-a* as a singulative marker. That singulative marker was reanalyzed as pertaining to the stem later on, most probably in the 19th century when the British had taken over the rule over Ceylon from the Dutch.

The lack of iconicity in Sinhala number marking correlates with the Animacy Hierarchy, which seems to be true for Kiowa as well. Some other languages (Cushitic languages, Imonda) seem to contradict this tendency, but in these languages the counter-iconic number marking is a marginal phenomenon and not as characteristic of the number system as in Sinhala or Kiowa. However, this tendency is cancelled for Sinhala if we take into account not only the citation form but also the inflected nominal forms. In these cases, we find that number marking is close to iconic.

The Sinhala number marking system defies easy description and explanation. Markedness phenomena, iconicity, animacy, and the case system are in a complex relation. This multifaceted influence on the number system is a challenge for cognitive linguistics and only time will tell whether good cognitive motivations for such a system can be found.

Abbreviations

ANIM = animate; EMPH = emphatic; INAN(IM) = inanimate; INDEF = indefinite; LOC = locative; PL = plural; SG = singular; SUBJ = subject

Acknowledgments

We would like to thank Umberto Ansaldo and Robert Cirillo for comments on an earlier version of this paper. All remaining mistakes are our own responsibility and should not be attributed to them. We are also grateful to the numerous linguists that participated in the discussion after Nordhoff's talk at the *Rara & Rarissima* Conference. This discussion has helped us enormously to come to grips with the motivations of the Sinhala number system.

Notes

1. General reference is sometimes called 'transnumeral' (Biermann 1982) or 'common number' (Jespersen 1924).
2. It is important not to confuse *general* reference and *generic* reference. The above example is of general reference, but not of generic reference, since nothing is said about the properties of 'dog' as such. On the other hand, the sentence *Lions live in Africa* is generic, but not of general reference because it would be incorrect if there were only one living lion left in Africa.
3. There is also a paucal *luban-jaa* 'lion-PAUCAL', which is not important in the remaining discussion.
4. Technically speaking, the Sinhala sentences have an argument focus on *vattē* 'garden'. In Sinhala, the focus construction is much more common than, say, a cleft sentence in English. This is why a plain English sentence was chosen as the translation.
5. Cf. also Matzel (1983: 59–60).
6. Theoretically, lexemes ending in *eka* in the singular and in *ek* in the plural would pose a problem, because one would not know whether the segment to subtract should be *eka* or just *a*, but no such lexemes happen to exist. Lexemes that end in *-v* or *-y* in the plural do not exist either, so one is sure to subtract the glide as well.
7. Only where NOM.SG and NOM.PL were the same, was a periphrastic construction involving **vara* 'mass' used. This eventually yielded the *-val* of the present class *v*.
8. This word is a neologism coined by Geiger in analogy to *Gebirge* 'mountain range' from *Berg* 'mountain' and several other *Ge-* words that denote collectives.
9. Unfortunately, there is no information available on class *iii* (*kekka/keki*), and the development of this class remains unclear to the present. *Kekka* might be an assimilation of **kekiya*, a process that is well attested elsewhere in the history of Sinhala, for instance in the development of the past forms of the second conjugation, where we get *däkkā* < **däkiyā* 'see.PST'.
10. One could assume that *giṭār eka* was already borrowed from Dutch *gitaar*, but this is not possible since a retroflex stop never occurs in words of Dutch origin. The Dutch dental stop is rendered as dental in Sinhala, whereas the English alveolar stop is rendered as retroflex.
11. As noted above, number marking is obligatory in Sinhala but not in the Cushitic languages.
12. With one exception, the word for 'bull'.

13. There are of course hundreds of languages where the iconicity of plural marking is constant and these languages are therefore not very interesting for the investigation of correlation between animacy and iconicity.
14. Whether the marking chosen for inanimates is counter-iconic as in Sinhala, or just simply non-iconic as in all the languages that do not mark number at all on inanimates (e. g. Japanese (Corbett 2000: 74)), is just a matter of degree, but not of kind.
15. Animates additionally have Accusative and Locative.
16. It is important not to confuse the plural case formative *val* with the plural suffix *vala*. That they are in fact different can be seen from class *v* obliques in the plural, e. g. *pāra-val-vala-ta* 'to the streets' with an effective twofold occurrence of the sequence *val*.

References

- Andrzejewski, Bogumil W.
 1960 The categories of number in noun forms in the Borana dialect of Galla. *Africa* 30: 62–75.
- Biermann, Anna
 1982 Die grammatische Kategorie Numerus. In *Apprehension: Das sprachliche Erfassen von Gegenständen I: Bereich und Ordnung der Phänomene*. Hansjakob Seiler and Christian Lehman (eds.), 229–243. Tübingen: Narr.
- Corbett, Greville G.
 2000 *Number* Cambridge: Cambridge University Press.
- Corbett, Greville G. and Richard Hayward
 1987 Gender and number in Bayso. *Lingua* 73: 1–28.
- DuBois, John
 1985 Competing motivations. In *Iconicity in Syntax*, John Haiman (ed.), 343–365. Amsterdam/Philadelphia: Benjamins.
- Geiger, Wilhelm
 1973 *Kleine Schriften zur Indologie und Buddhismuskunde*. Wiesbaden: Steiner.
- Greenberg, Joseph H.
 1976 *Language Universals – with special reference to feature hierarchies*. Reprint. The Hague/Paris: Mouton. [original print: 1966].
- Haiman, John (ed.)
 1985 *Iconicity in Syntax*. Amsterdam/Philadelphia: Benjamins.
- Hayward, Richard
 1984 *The Arbore language: a First Investigation: Including a Vocabulary*. Hamburg: Buske.
- Jakobson, Roman (ed.)
 1971 *Word and language. Selected writings. Vol. 2*. The Hague/Paris: Mouton.
- Jakobson, Roman
 1971 Signe zéro. In *Word and language. Selected writings*. Roman Jakobson (ed.), (Vol. 2.) Reprint, 59–61. The Hague/Paris: Mouton. [original print: 1939].
- Jayawardena-Moser, Premalatha
 2004 *Grundwortschatz Singhalesisch – Deutsch*. (3rd edition) Wiesbaden: Harasowitz.

- Jespersen, Otto
 1924 *The Philosophy of Grammar*. London: Allen and Unwin.
- Karunatilake, W. S.
 2004 *An Introduction to Spoken Sinhala*. 3rd edition. Colombo: Gunasena.
- Matzel, Klaus
 1983 *Einführung in die singhalesische Sprache*. 2nd edition. Wiesbaden: Harras-sowitz,
- Mayerthaler, Willi
 1988 *Morphological Naturalness*. Ann Arbor: Karoma.
- Newmeyer, Frederick J.
 1998 *Language Form and Language Function*. Cambridge, MA/London: M.I.T. Press.
- Seiler, Walter
 1985 *Imonda, a Papuan language*. (= Pacific Linguistics; B-93). Canberra: Depart-ment of Linguistics.
- Stolz, Thomas
 2007 Re: duplication. Iconic vs. counter-iconic principles (and their areal correlates) In *Europe and the Mediterranean as linguistic areas*, Paolo Ramat and Elisa Roma (eds.), 317–350. Amsterdam: Benjamins.
- Watkins, Laurel J.
 1984 *A grammar of Kiowa*. Lincoln/London: University of Nebraska Press.
- Wurzel, Wolfgang U.
 1989 *Inflectional Morphology and Naturalness*. Dordrecht/Boston/London: Klu-
 wer.

The Dravidian zero negative: diachronic context of its morphogenesis and conceptualisation

Christiane Pilot-Raichoor

1 Introduction

The Dravidian ‘zero negative’, as Master (1946) called it, is a clear structural *rarum* (see Section 2) which has failed to be properly acknowledged, even by the Dravidianists who generally reduced it to a phonetic accident ($*\bar{a} > \emptyset$), that leaves the negative forms <verb root + person endings> without any sign of negation. The existence of this strange construction, attested in several languages over a period of two thousand years, cannot be questioned. The debate surrounding it bears entirely on its origin.

A critical examination of the traditional explanation (Section 3) will show its weaknesses. Moreover, a closer look at the earliest records of a Dravidian language (Section 4) will reveal that neither $*\bar{a}$ nor \emptyset are likely to be ‘Proto-Dravidian’ negators. Both of them emerged from an important grammatical reshuffling which occurred pre-historically and can be perceived only in the earliest stages of Tamil (from 3rd century BC onwards). An analogy with the paradigms of inherently negative verbs (‘to be not’) triggered the genesis and the stability of the zero negative construction. Finally, its conceptualisation (Section 5) will be explained by working out the systemic semantics of the contrastive ‘tenses’ (Past / Non-Past) and ‘negatives’ ($*\bar{a}$ and \emptyset) morphemes. This analysis will be supported by theoretical and typological arguments.

2 A robust linguistic fact

Though its existence has often been denied by Dravidianists and therefore generally ignored by other linguists, the zero negative construction is an obvious and common linguistic fact in most South Dravidian languages, old and modern, written and oral.

2.1 The morphological evidence

The existence in Dravidian of finite verb forms constructed on the schema <verb + person suffix> with a negative meaning, in contrast to positive forms

Table 1. Tamil finite paradigms

	Past	Future	Negative
1S	<i>paṭi-tt-ēṇ</i>	<i>paṭi-pp-ēṇ</i>	<i>paṭiy-ēṇ</i>
2S	<i>paṭi-tt-āy</i>	<i>paṭi-pp-āy</i>	<i>paṭiy-āy</i>
3SM	<i>paṭi-tt-āṇ</i>	<i>paṭi-pp-āṇ</i>	<i>paṭiy-āṇ</i>
3SF	<i>paṭi-tt-āl</i>	<i>paṭi-pp-āl</i>	<i>paṭiy-āl</i>
3SN	<i>paṭi-tt-atu</i>	<i>paṭi-kkum</i>	<i>paṭiy-ātu</i>
1P	<i>paṭi-tt-ōm</i>	<i>paṭi-pp-ōm</i>	<i>paṭiy-ōm</i>
2P	<i>paṭi-tt-īr</i>	<i>paṭi-pp-īr</i>	<i>paṭiy-īr</i>
2P (honor.)	<i>paṭi-tt-īrkal</i>	<i>paṭi-pp-īrkal</i>	<i>paṭiy-īrkal</i>
3PM/F	<i>paṭi-tt-ār</i>	<i>paṭi-pp-ār</i>	<i>paṭiy-ār</i>
3PM/F (honor.)	<i>paṭi-tt-ārkal</i>	<i>paṭi-pp-ārkal</i>	<i>paṭiy-ārkal</i>
3PN	<i>paṭi-tt-ana</i>	<i>paṭi-kkum</i>	<i>paṭiy-ā</i>

Transliterated and adapted from Arden ([1891] 1976: 141, 228).

NB: y- occurring in *paṭiy-* is a glide after front vowel.

constructed on the schema <verb + tense + person suffix>, is a descriptive fact that even the most reluctant authors have to accept: “in several languages the negative is signalled by a zero” (Krishnamurti 2003: 348).

What distinguishes this zero negative paradigm from other “negatives without negators” studied by Miestamo (2010) is the simplicity of the construction. This is evident when comparing the three basic paradigms found in the description of literary Tamil or Kannada.

In Tamil, for the verb *paṭi-* ‘to learn’, the following paradigms can be set for the Past (*paṭittēṇ* ‘I learned’), the Future (*paṭippēṇ* ‘I will learn’) and the Negative (*paṭiyēṇ* ‘I do / did / will not learn’), as illustrated in Table 1.

In Old Kannada, the corresponding forms for the verb *nōḍu* ‘to see’ are: *nōḍidem* ‘I saw’, *nōḍuvem* ‘I will see’ and *nōḍem* ‘I do / did / will not see’, see Table 2 on the facing page.

What is meant by ‘zero negative’ clearly appears in the systemic description: the analysis of the positive forms in Tamil and Kannada makes it possible to identify a morphological \emptyset - in place of the “tense” affixes in the negative forms as shown in Table 3 on the next page.

These three paradigms actually reflect the fundamental ternary opposition system of ‘Past’ vs. ‘Non-Past’ vs. ‘Negative’ which has been reconstructed as a common basis for Proto-Dravidian verbal morphology (see Section 5.2.2).

Table 2. Old Kannada finite paradigms

	Past	Future	Negative
1S	<i>nōḍ-id-eṃ</i>	<i>nōḍ-uv-eṃ</i>	<i>nōḍ-eṃ</i>
2S	<i>nōḍ-id-ai</i>	<i>nōḍ-uv-ai</i>	<i>nōḍ-ai</i>
3SM	<i>nōḍ-id-aṃ</i>	<i>nōḍ-uv-aṃ</i>	<i>nōḍ-aṃ</i>
3SF	<i>nōḍ-id-aḷ</i>	<i>nōḍ-uv-aḷ</i>	<i>nōḍ-aḷ</i>
3SN	<i>nōḍ-id-udu</i>	<i>nōḍ-uv-udu</i>	<i>nōḍ-adu</i>
1P	<i>nōḍ-id-em</i>	<i>nōḍ-uv-em</i>	<i>nōḍ-evu</i>
2P	<i>nōḍ-id-ir</i>	<i>nōḍ-uv-ir</i>	<i>nōḍ-ir</i>
3PM/F	<i>nōḍ-id-ar</i>	<i>nōḍ-uv-ar</i>	<i>nōḍ-ar</i>
3PN	<i>nōḍ-id-uvu</i>	<i>nōḍ-uv-uvu</i>	<i>nōḍ-avu</i>

Transliterated and adapted from Kittel (1903: 138, 143, 158)

Table 3. ‘Tense’ morphs in Literary Tamil and Old Kannada

	Root	Tense	Person	Root	Tense	Person
Past	<i>paṭi</i>	<i>-tt-</i>	<i>ēn, āy, etc.</i>	<i>nōḍ</i>	<i>-id-</i>	<i>eṃ, ai, etc.</i>
Future	<i>paṭi</i>	<i>-pp-</i>	<i>ēn, āy, etc.</i>	<i>nōḍ</i>	<i>-uv-</i>	<i>eṃ, ai, etc.</i>
Negative	<i>paṭi</i>	<i>-ḷ-</i>	<i>ēn, āy, etc.</i>	<i>nōḍ</i>	<i>-ḷ-</i>	<i>eṃ, ai, etc.</i>

2.2 Meaning of the negative paradigm

The meaning of this basically tenseless zero negative paradigm is as simple as its construction: it is, as Arden (1976: 228) called it, a “pure negative” which can refer “to all time”. It usually expresses a strong, categorical negation.

Examples can be found in the oldest Tamil literary corpus, the ‘Sangam’¹ (examples (1) through (3), from Rajam 1992) as well as – two thousand years later – in a contemporary unwritten minority language, Badaga (examples (4) through (7) on the following page, personal data) with very little change.

- (1) Old Tamil (Rajam 1992: 640) [pura.399:14]

cell-ēn cell-ēn pīraṛ mukam nōkk-ēn

go.0.1S go.0-1S others face see.0-1S

‘I will not go. I will not go. I will not look at the faces of others.’

- (2) Old Tamil (Rajam 1992: 638) [kuru.178:3-4]

ivaḷ iṭai mulai kiṭantum naṭuṇkal āṇ-īr
 she middle breast lie.CVB shiver.NMLZ be_content.0-2P

‘Although you are lying between her breasts you do not stop trembling.’

- (3) Old Tamil (Rajam 1992: 635) [aka.33:18]

anru nam ari.y-āy
 then us know.0-2s

‘You did not know us then’

- (4) Badaga (personal data)

nā avva gaṇḍana ari.y-e
 I her.GEN husband.ACC know.0-1s

‘I don’t know her husband’

- (5) Badaga (personal data)

enga iduga bele ari.y-o
 we this.DAT price know.0-1p

‘We do not know the price for this’

- (6) Badaga (personal data)

ē gaṇḍa.n-ū iduna māḍ-a
 which chief.any this.ACC do.0-3M

‘No chief will do this’; lit. *ē... - ū* ‘whichever / whatever’

- (7) Badaga (personal data)

ama munduga bar-a
 he forward come.0-3M

‘He won’t come up in life’

2.3 A widespread and enduring negative construction

The zero negative paradigm is well attested:

- in the *old stages of the literary languages*
- in several *modern languages spoken by minorities*

As it is more frequent in South India, Pederson (1993: 236) suggests that “there was areal influence spreading the zero negative from South Dravidian to some Central languages”.

In Tamil, these negative forms originated during the Sangam period (cf. Section 4), developed fully in post-Sangam or medieval Tamil and vanished as a regular paradigm in modern Tamil.

In Kannada, which was attested centuries after Tamil, the negative paradigm is well recorded in the old stages of the language corresponding roughly to medieval Tamil and subsequently tends to vanish. It is also well attested in literary Telugu, but rare in Malayalam which uses the **ā* marker except in a few instances in the old literary records, still influenced by Tamil.

More importantly, the zero negative paradigm is still alive in several unwritten languages: dialectal variants of the literary languages (Havyaka Kannada) or languages spoken by several minorities, notably in the Nilgiri hills of South India (Kota, Toda, Badaga). The zero negative construction is also attested partially in Gondi (South-Central Drav.) and Brahui (North Drav., cf. Krishnamurti 2003: 353).

Due to the general aversion of linguists to this negative construction (cf. Section 3.1.2) and to language specific developments,² it is difficult to know exactly the spread of the phenomenon.

2.4 Functionality of the zero negative construction

As shown in Section 2.1, the zero negative construction is basically a plain declarative finite paradigm. A few words should be added concerning its relation to other negative finite constructions (Section 2.4.1) and to other positive finite paradigms (Section 2.4.2) which fulfill similar functions.

2.4.1 *Other negative finite constructions*

At all historical periods and in all languages that have zero negatives, the zero negative construction co-exists with other ways of expressing negation. Two main types of negation can be distinguished in Dravidian: morphological (different types of negative finite paradigms) and syntactic (several types of construction with an inherently negative verb or a negatively conjugated verb as auxiliary). Both types exist in nearly all Dravidian languages.³

In Tamil for instance, the zero negative paradigm occurs along with various other Negative periphrastic constructions, most of them using a fossilized

form of the old negative root⁴ *il* (< **cil-*) ‘not to be’ as an auxiliary. The auxiliary *māṭṭ-* is also used conjugated in the zero negative paradigm⁵ in the *Negative Future*. Arden (1976: 228–230) identifies the following main negative constructions:

- the pure *Negative Tense* {root of the verb + personal terminations} co-exists with
- the *Negative present* {Nominative Singular of the Present Neuter Participial (or Verbal) Noun in *-tu + illai*}
- the *Negative Past Tense* {infinitive + *illai*}
- the *Negative Future* {infinitive + *māṭṭ-* + personal terminations}

The presence of other constructions often does restrict the use of the negative paradigm to some specific uses. For instance in Tamil: “it is commonly used only as an habitual Present Tense. Thus [...] *paṭiyēṇ* ‘I learn not’” (Arden 1976: 228).

In short, the zero negative construction is not retained as a ‘standard’ negation in any of the *modern* languages. However it still survives in the *literary* languages:

1. as an archaic feature: such is the case in Kannada, for which Sridhar (1990: 228) explicitly states that the Negative ‘tense’ is “found mostly in older writing”.
2. as part of complex/periphrastic negative constructions⁶, as stated for the Negative Future in Modern Tamil: *colla māṭṭ.ēṇ* ‘I will not say’, *colla māṭṭ.āy* ‘You will not say’, etc.
3. in some expressions in Modern Tamil, such as: *avanai kāṇōm* ‘He is not to be seen, lit., we do not see him’ (Subrahmanyam 1971: 333), or, as reported by Kittel in Kannada:

“The English verb ‘to will’, ‘to be pleased’, ‘to desire’ is expressed by *ol*, of which (*ollem*), *ollenu*, *olle*, vulgarly *valle*, ‘I will not’, and other persons of the negative are *in every body’s mouth*, e. g. [...] *uṇṇa olle*, ‘I will not eat’.” (1903: 340–341) [emphasis mine].

These expressions, as noticed by Kittel, are quite common, though the speakers are unaware of the oddity of the negative form they use!

Table 4. Negative paradigm versus positive paradigms

Tamil (medieval):				
Negative	Future	Past	Present	
<i>var-ēn</i>	vs. <i>kāṇ-p-ēn</i>	<i>va-nt-ēn</i>	<i>cey-kinr-ēn</i>	
‘I will not come’	‘I will see’	‘I came’	‘I do’	
V+pers.	vs. V+Fut+pers.	V+Past+pers.	V+Pres+pers.	
Toda:				
<i>pī·x</i> (S2: <i>pī·-</i>) ‘to go’, <i>po·r</i> (S2: <i>pod-</i>) ‘to come’				
Negative	Present / future	Future	Past	Tenseless (aorist)
<i>pī·x-en</i>	vs. <i>pī·-p-en</i>	<i>pī·-k-in</i>	<i>pī·-šp-en</i>	<i>pī·-n-en</i>
<i>po·r-en</i>	<i>pod-p-en</i>	<i>paš-k-in</i>	<i>pod-šp-en</i>	<i>pod-en</i>
‘I do / did / will not V’	‘I (shall) V’	‘I will V’	‘I did V’	‘I (did / will) V’
V+pers.	vs. V+Prs-Fut+pers.	V+Fut+pers.	V+Past+pers.	V+Aor+pers

2.4.2 Other positive finite paradigms

In modern Dravidian languages, the zero negative paradigm is usually opposed to several positive tenses, but this does not affect its functional contrastive meaning as all the positive paradigms are overtly marked, as illustrated in Table 4 (Toda data from Emeneau 1984). In other words, in all the languages where this form of negation exists, the positive tenses⁷ of the declarative paradigms can always be analysed in (at least) three morphemes: {verb root – “tense” – person} while the zero negative paradigm is reduced to two segmental units: {verb root – person}.

Presented in isolation, the negative meaning of the construction {verb root – person} thus appears as a sheer nonsense; however if we consider it in the bigger picture of the verbal systems, throughout the languages and the centuries, it becomes clear that the zero negative construction acquires its identity and maintains its linguistic and communicative distinctiveness on the basis of constant contrasting system in the marking of *tense*:

Negative paradigm [- tense] vs. Positive paradigms [+ tense].
no marker vs. overt marker

Does this contradict the universal of polarity marking, according to which the positive forms are less marked than the negative ones? Looking at the raw data, a trivial answer would be “yes”, but further analysis shows that the negative forms are in fact not less marked than the positive ones. Thus, from a morphological point of view, a positional zero can be considered a true marker, as in numbers (e.g. 1, 10, 100...); from a categorial point of view, Miestamo (2010) argues that in both cases it is tense-aspect-mood which is marked and not polarity (positive or negative) as such. In the same line, I shall argue (Section 5) that it is not polarity which is encoded but the existence / non-existence of the verb as a process.

3 Controversy on the origin of the zero negative

In spite of all these pieces of evidence, the very existence of the zero negative paradigm is still hard to accept not only for general linguists or typologists (cf. Forest 1993: 19–20: “ce n’est qu’une apparence” [it just appears (to exist)]) but also for most Dravidianists, even the most renowned, such as Caldwell, Kittel, Bloch or, more recently, Krishnamurti, who writes in a major reference book on the Dravidian languages (2003: 348 [emphasis mine]):

“The negative allomorphs occurring in inflected verbs have *abnormal* phonology [i. e. zero allomorph] and are, therefore, of uncertain origin. *The notion of a zero negative in Dravidian is a myth.*”

Thus, nothing has really changed since Pederson (1993) reminded us of Master’s complain:

“Ninety years have passed since Gundert and Caldwell first discussed the origin of the Dravidian Zero Negative as it is now called and the problem remains unsolved.” (Master 1946: 137)

3.1 Matter of the dispute: the negative allomorphs $-\emptyset$ - and $*\{\bar{a}\}$

A proper account of the zero negative has been constantly biased by a debate about its origin and its relation to another morphological negative marker $*\bar{a}$.

3.1.1 *An uneven distribution*

Outside of the South subgroup, most Dravidian languages present, in place of the zero, a morph $*\bar{a}$ (or a variant of it). Table 5 gives a sample of the languages having a reflex of $*\bar{a}$ (group I) in contrast to those presenting a $-\emptyset$ -

Table 5. Comparison of selected Dravidian negative finite paradigms

	Past			Non-Past			Negative		
	vb.	sfx.	pers.	vb.	sfx.	pers.	vb.	sfx.	pers.
Group I	<*ā								
Tulu	'learn'	<i>kal</i>	-t- <i>e</i>	'learn'	<i>kal</i>	-p- <i>e</i>	'listen'	<i>kēn</i>	ay- <i>e</i>
Gondi	'beat'	<i>pā</i>	-t- <i>ōn</i>	'see'	<i>sūr</i>	ānt- <i>ōn</i>	'wash'	<i>nor</i>	vō <i>nu</i>
Konda	'see'	<i>sūr</i>	-t- <i>a</i>	'bring'	<i>ta</i>	-n- <i>a</i>	'do'	<i>ki</i>	-?- <i>e</i>
Group II	∅								
Tamil	'cry'	<i>aḷu</i>	-t- <i>ēn</i>	'make'	<i>cey</i>	-v- <i>ēn</i>	'see'	<i>kāṇ</i>	-∅- <i>ēn</i>
Kannada	'listen'	<i>kēḷ</i>	-d- <i>eṃ</i>	'give'	<i>kuḍ</i>	-uv- <i>eṃ</i>	'see'	<i>nōḍ</i>	-∅- <i>eṃ</i>
Toda	'come'	<i>po</i>	-d- <i>en</i>	'come'	<i>pod</i>	-p- <i>en</i>	'come'	<i>po·r</i>	-∅- <i>en</i>

(group II) in the negative finite paradigms. Furthermore, in all the languages which have a zero in the finite paradigm, we find a marker *ā (or a variant of it) in all the non-finite forms of the verb. Comparing the first person forms of the three paradigms of Kannada given in Table 2 on page 269 with the corresponding participial forms (verb + 'tense' + *a* 'adjectival marker'), shows that the negative participial form is overtly marked with a morph -*ad*- in place of the zero morph of the finite paradigm:

	Past	Future	Negative
Finite form (1 SG.)	<i>nōḍ-id-eṃ</i>	<i>nōḍ-uv-eṃ</i>	<i>nōḍ-∅-eṃ</i>
Participial form	<i>nōḍ-id-a</i>	<i>nōḍ-uv-a</i>	<i>nōḍ-ad-a</i>

The restricted distribution of the -∅- negative allomorph has obviously legitimated the question of its origin.

On the one hand, the allomorph -∅- is restricted (i) to only one paradigm of the indicative finite forms and (ii) to a limited group of Dravidian languages, on the other, the *ā allomorph (i) is found in most of the negative morphological constructions (several finite paradigms as well as most of the non-finite forms (participial and nominal forms) and (ii) is attested in all languages.

Comparative works usually reconstruct three negative morphs. Subrahmanyam gives: *-ā- (1971: 383), -∅- (1971: 385) and *-vā- (1971: 386, for some South-Central Dravidian languages), but in detail the situation is much more complicated.

Cautiously, Subrahmanyam (1971: 383) admits that "it is very difficult to decide as to which of the two types [-∅- or *-ā-] represents the Proto-Dravidian situation".

3.1.2 *Misleading analyses overrepresenting *-ā-*

At present, it is impossible to have a precise view of the expression of negation and its historical developments in the Dravidian family of languages. Preconceptions about the need to identify or reconstruct a negative marker have biased many analyses of individual languages. In particular, minority languages of Central India present a great variety of negative constructions which have been insufficiently or inaccurately analysed.⁸

The non-specialist must be aware that there is a strong tendency among Dravidianists to favour the **ā* allomorph, leading to some dubious analyses even for the most well-known languages. In this view, it is interesting to compare the paradigm of *nōḍu* ‘to see’ as given by Kittel for Medieval Kannada and the paradigm of *tinnu* ‘to eat’ as given by Sridhar (1990) for (modern) Kannada, see Table 6.

Table 6. Two alternative analyses compared

	Negative (medieval Kannada in Kittel 1903: 159), <i>nōḍu</i> ‘see’	Negative “tense” (in Sridhar 1990: 227–228), <i>tinnu</i> ‘eat’
1S	<i>nōḍ-em̐</i> , <i>nōḍ-enu</i> , <i>nōḍ-e</i>	<i>tinn-e(-nu)</i>
2S	<i>nōḍ-e</i>	<i>tinn-e</i>
3SM	<i>nōḍ-am̐</i> , <i>nōḍ-anu</i> (<i>nōḍ-a</i>)	<i>tinn-a(-nu)</i>
3SF	(<i>nōḍ-aḷ</i>), <i>nōḍ-aḷu</i>	<i>tinn-a-ḷu</i>
3SN	<i>nōḍ-adu</i>	<i>tinn-a-du</i>
1P	<i>nōḍ-evu</i>	<i>tinn-e-vu</i>
2P	<i>nōḍ-iri</i>	[? missing in Sridhar]
3PM/F	(<i>nōḍ-ar</i>), <i>nōḍ-arū</i>	<i>tinn-a-ru</i>
3PN	<i>nōḍ-avu</i>	<i>tinn-a-vu</i>
NB.: <i>-nu</i> is the regular phonetic evolution of final <i>-ṃ</i>		

Sridhar analyses the vowel which follows the verb root as a marker of negation and justifies his segmentation by the following comment:

“a special verb paradigm in which the negative tense marker (*-e* in first and second person and *-a* in the third person) is placed immediately to the right of the verb stem (i. e., in place of the tense suffix) and before the agreement marker. e. g., *hōg-a-ḷu* go-neg-3sf, ‘she doesn’t/won’t go’.” (1990: 227)

However, this analysis appears to be biased when one takes in consideration the data presented in Table 2 above. Contrary to Sridhar’s statement, the

Table 7. Krishnamurti's analysis of Modern Telugu

'Habitual-future or non-past':	-t-	<i>ammu-t-ānu</i>	'I (habitually) sell, I shall sell'
'Past tense':	-i-	<i>amm-i-ānu</i>	'I sold'
'Negative tense':	-a-	<i>amm-a-nu</i>	'I (do, did, and shall) not sell'

Table 8. Verbal and predicative nominal person endings in Telugu

Verb: ‘to sell’	1S	1P	2S	2P	3SM	3SF/N	3PM/F	3PN	
Non-Past	<i>ammu t-</i>	<i>-ānu</i>	<i>_-ām</i>	<i>_-āw</i>	<i>_-āru</i>	<i>_-āḍu</i>	<i>_-un-di</i>	<i>-āru</i>	<i>-āy</i>
Past	<i>amm i-</i>	<i>ānu</i>	<i>_-ām</i>	<i>_-āw</i>	<i>_-āru</i>	<i>_-āḍu</i>	<i>_-in-di</i>	<i>-āru</i>	<i>-āy</i>
Negative	<i>amm-</i>	<i>anu</i>	<i>_-am</i>	<i>_-aw</i>	<i>_-aru</i>	<i>_-aḍu</i>	<i>_-adu</i>	<i>-aru</i>	<i>-aw</i>
Noun: ‘the person’	1S	1P	2S	2P	3SM	3SF/N	3PM/F	3PN	
	<i>vāḍ-</i>	<i>anu</i>	<i>vār-amu</i>	<i>vāḍ-avu</i>	<i>vār-arū</i>	<i>vāḍ-u</i>	<i>vā-ru</i>		

endings: *-e(nu)*, *-e*, *-a(nu)*, *-aḷu*, *-adu*, *-evu*, *-iri*, *-aru*, *-avu* are (as shown in Kittel's analysis) unsegmentable person suffixes (direct reflexes of the person suffixes used in the positive and negative paradigms in Ancient Kannada) and there is no trace of a "negative tense marker" in them.

In fact, in his improper segmentation of a negative *-e/-a* tense marker, Sridhar follows Krishnamurti's analysis of the Telugu, which itself, on closer examination, also appears unlikely. In his *Telugu verbal bases*, Krishnamurti (1961: 213) gives the segmentation reproduced in Table 7 for the three basic paradigms of the verb *ammu* 'to sell' in Modern Telugu, which sets, next to the tense markers *-t-* and *-i-*, a negative marker *-a-*.

However, the proposed segmentation of an *-a-* negative marker in *amm-a-nu*, *amm-a-m*, *amm-a-w* etc. cannot be maintained when we tabulate, along with the verbal paradigms, the predicative paradigm of the noun *vāḍu* 'person' (Krishnamurti, 1961: 233, note 8, *vāḍ-anu* lit. 'the person, I', *vār-amu* lit. 'the persons, we') meaning 'I am the person', 'we are the persons', etc. Table 8 shows beyond any doubt that the same set of (positive unsegmentable) personal endings *-anu*, *-amu*, *-avu*, *-aru* etc. is used in both the Negative paradigm and the Nominal paradigm. In this table, it can also be noticed that, contrary to what would be expected if there had been a fusion of a negative marker with the person endings (as usually suggested, see Section 3.2.1), it is the positive forms (Past and Non-Past) of the person suffixes which contain a long vowel, while the negative ones – like the nominal ones – have

Table 9. A systemic analysis of Telugu

	Root	“Tense”	Person
Non-Past	<i>ammu</i>	+ <i>-t-</i>	+ <i>ānu, āw, ām, āru etc.</i>
Past	<i>amm</i>	+ <i>-i-</i>	+ <i>ānu, āw, ām, āru etc.</i>
Negative	<i>amm</i>	+ <i>-∅-</i>	+ <i>anu, aw, am, aru etc.</i>

only short vowels. Therefore, a systemic analysis will identify a morphological zero negative marker in Telugu (cf. Table 9), as in Tamil and Kannada⁹ (cf. Table 3 on page 269).

A similar distortion of the basic facts was noticed by Pederson (1993: 234) with regard to another important Dravidianist:

“Andronov (1970) doesn’t try to explain the phenomenon and provides only lists of cognate negative suffixes (which err on the side of *overrepresenting* ā as richer and more widespread than would appear to be the case on examination of individual languages)” [emphasis mine].

3.2 The traditional account: a phonetic accident $*\bar{a} > \emptyset$

According to most of the linguists who questioned its existence, the zero negative paradigm can only be explained by a phonetic accident.

3.2.1 A standard phonetic rule in historical linguistics: $V + V > V$

The hypothesis of a phonetic fusion between the vocalic negative marker and the following vocalic person suffixes appears in Caldwell (1856):

“We have thus arrived at the conclusion that *a* is the sign of negation which is most systematically used by the Dravidian languages in the formation of the negative voice of the verb. It has, it is true, disappeared from the conjugated forms of Tamil and Canarese [= Kannada]; [...]. The negative *a*, being succeeded in Tamil and Canarese by the initial vowel of the pronominal suffix, appears gradually to have got incorporated with it; and an evidence of this incorporation survives in the euphonic lengthening of the pronominal vowel in Tamil.” (Caldwell 1856 [1976]: 473)

and has been repeated by most of the linguists who have discussed the question from Bloch (1935) to Pederson (1993).

The phonetic fusion of two vowels is a very common phenomenon in historical linguistics and could be readily accepted as a plausible explanation.

3.2.2 An “abnormal” phonetic rule in Dravidian

However, in the present case, the fusion is phonetically completely “abnormal”, as Krishnamurti (2003: 348) admits. Bloch (1935) was the first to recognize that the morphophonemic (sandhi) rules that are required do not exist in any of the attested forms of the languages:

“[...] le verbe dravidien n’exprime que par accident la négation au moyen d’un suffixe zéro, mais qu’il possède réellement un suffixe négatif [-a-]” (1935: 162), “là où le négatif est apparemment sans marque spéciale, c’est qu’il y a eu une voyelle -a- élidée ; la seule difficulté, qui n’a pas été soulignée, mais qui a sans doute inconsciemment empêché l’adhésion des autres savants, c’est qu’il faut admettre que les règles de sandhi ancien admettaient, notamment en tamoul, des élisions ou plutôt des contractions que déjà la langue classique ne permettait plus.” (1935: 159)

[“Dravidian verbs only accidentally express negation through a zero suffix but [...] do possess a true negative suffix [-a-]” (1935: 162), “in places where the negative apparently does not have any special marking, the -a- vowel has actually been elided ; the only difficulty which has not been emphasized but which surely has unconsciously prevented other scholars from supporting the idea is that the old sandhi rules, especially in Tamil, allowed elision, or rather contractions that the classical language already denied”]

3.2.3 Shortcomings of the phonetic accident hypothesis

Consonant epenthesis is the normal rule

The phonetic explanation requires sandhi rules that run against what is usually observed in the ancient stages of the Dravidian languages:

“Typically, even the older South Dravidian languages invoked sandhi consonant epenthesis pre-vocalically, but in the case of zero negation there is vowel deletion.” (Pederson 1993: 236)

What is expected at the junction of two morphemic vowels is consonant epenthesis as seen in the Tamil Negative paradigm with the insertion of the glide -y- after the front vowel of *paṭi* ‘to learn’ and the person suffixes *paṭi-y-ēn* or the glide -v- after back vowel of *ō* ‘to end, cease’ in *ō-v-ār*. The same epenthesis can also be seen in the Tulu negative form *kēn-ay-e* (cf. Table 5 on page 275), which actually contains a negative marker -a- before the person suffix.

Both the negative marker and the person suffix contain long vowels

The reconstructed morphemes for the negative marker as well as for the person suffixes have long vowels. It is true, however, that both of them present

short variants, variously distributed in the individual languages, and many authors take advantage of this allomorphy to avoid unlikely phonetic processes.

(i) Some authors start with a reduction of the negative marker $*\bar{a}- > *a-$ (explicitly or implicitly, cf. the preceding citations of Caldwell and Bloch) so that the reader will more readily accept $*a- > \emptyset$. This is the strategy adopted by Krishnamurti (2003) who also argues – after Bloch (1935) – for the presence of a laryngeal (H) in the reconstructed negative marker:

“Taking the total scenario into account, we can reconstruct $*aHa-$ (or $*aHaH- > -\bar{a}y-/ay-$) for Early Proto-Dravidian or Pre-Dravidian which developed into a long grade $-\bar{a}$ (by contraction), a short grade $*a-$ by loss of $*H$ and the zero grade by the loss of this vowel before personal suffixes beginning with vowels.” (Krishnamurti 2003: 353)

However this laryngeal hypothesis (a) does not escape the reconstruction of a long $*\bar{a}$ for the negative marker: to account for the “five allomorphs $ay/\bar{a}y\sim\bar{a}/a\sim\emptyset$ ” (Krishnamurti 2003: 351) identified in South Dravidian I, “we need to set up [...] $*aH$ (or more legitimately $*\bar{a}H$)” (Krishnamurti 2003: 351) and (b) does not offer a better solution for the “loss” of $-a-$ (even in its reduced form) before the personal suffixes.

According to his hypothesis $k\bar{a}\bar{n}\text{-}\emptyset\text{-}\bar{i}rka\bar{l}$ ‘you (pl.) do not see’ (2003: 349) should come from $*k\bar{a}\bar{n}a\text{-}\bar{i}rka\bar{l} < *k\bar{a}\bar{n}\bar{a}H\text{-}\bar{i}rka\bar{l}$. The reduction of $-a + \bar{i} > \bar{i}$ being phonetically difficult to accept, Krishnamurti does not speak of fusion but mentions only the “loss” of $-a-$ in a vocalic context ($/_V\ldots$). However, the Kota form $tin\text{-}\emptyset\text{-}k\bar{o}$ (2003: 350) ‘(3RD SG/PL) do not eat’ cannot be accounted for by the same rule, as the personal suffix begins with a consonant.

(ii) Other linguists choose to start with reduced variants of the personal suffixes. Thus Pederson (1993) suggests the following development:

“[...] in Proto-South Dravidian the personal suffixes all originally had short vowels. Subsequently the long \bar{a} before the previously short vowel initial personal suffixes was lost with compensatory lengthening of the short vowel.”, then “[...] the long vowel personal suffixes [of the negative] subsequently and gradually extended to positive polarity uses as well.” (Pederson 1993: 238–239)

Unfortunately, it is difficult to accept his hypothesis as (a) no historical data support it: the personal suffixes with long vowels were already attested in all types of constructions in the Sangam corpus, not only with verbs in past $t\bar{o}nriy\bar{e}n$ ‘I appeared’ [pura 397:11] (Rajam 1992: 600) or non-past $ariv\bar{e}n$ ‘I realize’ [kuru 352:5–6] (Rajam 1992: 626), but also with noun-based constructions $p\bar{a}viy\bar{e}n$ ‘I [who is] a sinner’ ($< p\bar{a}vi$ ‘sin’) [tiru 5–6.58] and (b) the

motivation for extension of the negative suffixes to the positive and nominal constructions remains unclear.

Finally, however one tries to handle the problem, the phonetic fusion of the negative marker with the personal suffixes remains “abnormal”, fitting neither with the South Dravidian sandhi rules, nor with the general phonetic rules. Despite all these inconsistencies, the phonetic explanation has rarely been questioned.

3.3 Constraints of the Comparative Grammar theoretical framework

Why have most of these expert philologists so easily accepted the reduction of $\bar{a} > \emptyset$ which is obviously an *ad hoc* solution, devoid of any basic phonetic justification? Among the factors which allowed the phonetic accident hypothesis to persist is the weight of the theoretical framework in which they work.

3.3.1 Theoretical support to the phonetic explanation

From its beginnings (Caldwell 1856), Dravidian linguistics has been dominated by the Comparative methodology and the hypothesis of a phonetic accident fits well in this framework.

First, in the long and successful practice of the Comparative Grammar framework, it has been repeatedly observed that when two linguistic units differ only by the presence vs. absence of one segment, it is usually due to phonetic erosion and finally the disappearance of the segment present in the original construction. Therefore, when the negative constructions (a) Root- \bar{a} -pers. and (b) Root-...-pers. are compared, it is naturally presumed that some phonetic material has weakened and disappeared in (b).

Second, the method has to account for the relationship between all the languages of the family. As $*\bar{a}$ is attested in all the subgroups, but not \emptyset , it is presumed that $*\bar{a}$ should be reconstructed as Proto-Dravidian (cf. Subrahmanyam 1971: 383).

An independent factor that has probably allowed the phonetic accident hypothesis to survive so long is the functional equivalence between the two constructions whatever its actual structure: Root- \bar{a} -pers. or Root- \emptyset -pers., the ‘Negative Tense’ fulfills roughly the same function inside the verbal system and therefore there is not too much harm in accepting that the \bar{a} negative marker has simply disappeared phonetically. See Section 5.1.2 for a justification of this alternance.

3.3.2 *Causes for the failure of the comparative methodology*

At least two causes may be identified which led to the failure of the comparative method in this particular case.

The first one is linked to the goals of the theoretical framework. Basically, the method has to ‘reconstruct’ the original morpheme: it must have some ‘substance’ (even if it is an abstract one, cf. the laryngeal proposed by Bloch and Krishnamurti), $*\emptyset > *\bar{a}$ is not imaginable, therefore it has to be $*\bar{a} > \emptyset$.

Flaws in its application should also be noticed. The strength of the comparative method relies primarily on the regularity of phonetic ‘laws’, but, in this particular case, the authors found only ‘abnormal’ rules to justify the loss of a communicatively important marker, the negation morph.

The second cause is linked to the particular structure of the Dravidian data available for the comparison. The method relies primarily on the comparison of all the languages of a given family. In the case of the Dravidian family, for nearly one millennium only one language, Tamil, is attested (partially), while the vast majority of the languages of the family – unwritten languages – were recorded more than two thousand years later... Therefore, the ‘weight’ of the Early Old Tamil data in the whole family is relatively reduced.¹⁰ However, as will be developed in the next section, it is only in the earliest records that one can find clues to the genesis of the zero negative construction.

To sum up the key elements of this section, it may be stated that, *phonetically*, the zero-type and \bar{a} -type paradigms are fundamentally different and irreducible (i.e. they cannot be linked together: no evolution from one marker to the other is possible), but *functionally*, they are equivalent (in complementary distribution in two different groups of languages).

4 **Historical context of the morphogenesis of the negative markers $*\bar{a}$ - and $-\emptyset$ -**

The debate, as stated at the beginning, concerns the origin of the zero morph expressing negation. If, on close examination, the traditional account of a phonetic accident has to be rejected, we are left with the initial question: what is its origin? The comparative approach comes to a stop with the reconstruction of three distinct negative markers ($*\bar{a}$ -, $*\nu\bar{a}$ - and $-\emptyset$ -) and fails to shed light on the origin of the zero morph.

As an alternative to this approach, it may be worthwhile to scrutinize the oldest historical records. Fortunately, the zero negative paradigm is a

characteristic feature of South Dravidian, the subgroup for which we have the longest history. In fact, as briefly mentioned above, the Early Old Tamil records predate by eight long centuries any other Dravidian records.¹¹ This corpus offers many peculiar linguistic features which are not found in any other language attested later. It is precisely in these oldest data that clues to the morphogenesis of the zero-negative paradigm can be found.

4.1 Distinctive features of Early Old Tamil: evidence for a typological shift

The linguistic analysis of the earliest Tamil records is extremely difficult. The monks' cave inscriptions in Tamil-Brāhmī script are fairly short and the Sangam corpus is grammatically a real 'chaos', as Pederson noticed about the personal suffix (1993: 238). The nature of the corpus – composite and poetic – may be partially responsible for the unusual features observed in it, but the main reason is most probably that the Tamil language underwent an important typological shift in the preceding period.

The hypothesis of a typological shift from an isolating language with no grammatical distinction between the lexical categories (noun, verb, adjective. . .) was proposed by earlier Dravidianists such as Caldwell (1856), Bloch (1946), Meenakshisundaram (1965), and strangely forgotten in the most recent works. Zvelebil et al. state it in the following lines:

“It can be supposed that the period preceding the EOT [Early Old Tamil] stage was in a state similar to ‘isolation’ while the EOT period [the first historically attested stage of development] had a transient character with disappearing traces of isolation, typical features of agglutination and nascent features of inflection” (Zvelebil et al. 1967: 37)

There are clear indications in the earliest records that the grammatical typology of the language was significantly different not only from the grammar of the modern languages but also from the grammar reconstructed by the comparative method. The main distinctive features of Early Old Tamil are:

– little morphology

“One prominent feature of Old Tamil is the frequent omission of morphological or syntactic material which marks semantic relations, such as case markers, coordinators, attributive markers, etc. This often gives rise to strings of uninflected words . . .” (Lehmann 1998: 98).

“The vast majority of nouns in the Corpus are in the nominative case” (Mahadevan 2003: 289)

– no verb / noun distinction

“It would appear that originally there was no difference in any instance between the verbal and the nominal form of the root” (Caldwell 1976: 195)

– polyfunctionality

“Many bare stems function as nouns and verbs in classical Tamil” (Rajam 1992: 54).

“[...] each root may be said to be capable of a threefold use (1) as a noun, (2) as an adjective and (3) as a verb.” (Caldwell 1976: 195)

In current terminology, it can be said that there are obvious traces in the oldest records of the polycategoriality and invariability of the lexemes. Detailed studies of the Tamil-Brāhmī inscriptions (mainly 2nd century BC – 2nd century AD) by Mahadevan (2003) and of the early literature by Rajam (1992) also show that numerous elements that appear (later or in the same corpus) as components of morphological constructions or as derivatives of lexemes are attested as independent words.¹² These features indicate that in the pre-history of this group of languages, the complex, categorially distinctive morphology of the modern Dravidian languages was not yet established. Crucially, this entails that the morphological type of negation cannot be reconstructed as Proto-Dravidian, but is a phenomenon which emerges at the dawn of Tamil history, as we shall see below.

4.2 Development of distinctive categorial morphology

In the Sangam corpus we are probably witnessing the last traces of isolation available for any Dravidian language. The data recorded in the Sangam compilation do not give any precise date or localization of the poems. However, in scrutinizing the plethoric morphological variations attested in this corpus we can sketch different phases of evolution leading the language from an isolating stage to the fully developed morphology of the modern languages.

(i) The development of agglutination

During this early transitory period, simple constructions appeared: root + suffix, particle or another root. Then agglutination generalized and a great diversity of constructions were formed — all possible morphemic combinations seem to have been tried,¹³ most of which disappeared in later periods. The Sangam corpus concurrently holds all these forms from invariable root strings to complex morphological constructions.

(ii) The noun-verb distinction

The distinction between verb and noun emerged slowly. For the verb, Mee-nakshisundaran (1965: 27) precisely states that “the Tamil Finite Verb structure reveals two strata, the earlier one which does not possess the pronominal suffixes and the later one which has the fully developed pronominal suffixes”. One of the earliest attested form is *ceyy-um* ‘does’¹⁴ and “a number of forms which end in *al* or *a* or *ka* [*k+a*] which are later on considered as the suffixes of the verbal nouns. These forms were used as implied command” (1965: 30). In the process of agglutination, various ‘particles’ were thus added to the root (*cey-in*, *cey-t-u*, *cey-k-u* etc.) and the first constructions on indistinct <root + (‘particles’) + person> appeared. The noun-verb distinction started with the use of slightly different variants of originally identical morphs in the nominal and verbal constructions (Murugaiyan and Pilot-Raichoor 2004). Notably, some elements inserted on the ‘verbs’ between the root and the personal affix were used to specify spatio-temporal determinations of the event¹⁵. These affixes later became the tense-aspect-mood markers of the verb, while similar or identical elements became components of the case morphology for the noun.

(iii) The development of regular nominal and verbal paradigms

At some point in time, the exuberant morphology which resulted from the generalization of agglutination tended to become organized into regular paradigms, giving rise roughly to what we find in the modern languages, with distinctive nominal and verbal morphology.

The old polyfunctional constructions of <root -(xxx)- person> became specialized either into predicative function or into argument function. The development of the ‘finite verb’ forms (earlier undistinguished from the participial nouns¹⁶), specialized in the predicative function would have offered the best context for a new distribution of the verbal constructions and have opened the way to systematic (re-)interpretation of their values.

The zero negative paradigm probably emerged during this last phase.

4.3 Emergence of morphological negation

In the perspective of the typological shift sketched above, it can be presumed that, in the pre-historic stage, negation was expressed only with the inherently negative roots (several roots having roughly the meaning of ‘to be not’ **cil*, **al*, **mal*, ...). The idea that the morphological negation developed in the wave of this major grammatical evolution finds support in some surpris-

ing data from the oldest records: simple constructions (two elements) which appear later as well-established *negative* forms are also attested in the Sangam corpus with *positive* meanings.

Interestingly, the polarity ambiguity concerns both types of constructions: the zero and the **ā* constructions.

4.3.1 *Polarity shift of personal forms in Early Old Tamil*

Notably, the construction <Lexical root + person suffix> is attested with both positive and negative meanings.

- (8) Root+pers: positive
kāṇ-ām ‘will see-we’ [aka 110.19] (Agesthalingom 1977: 186)
kēḷ-ām ‘shall we listen to’ [kali 144.7-8] (Rajam 1992: 612)
- (9) Root+pers: negative
ceyy-ām ‘will not do-we’ [aka 106-7] (Agesthalingom 1977: 186)

To understand this fact, one should remember the nature of the corpus (collections of poems) and its time span over centuries (at least from 150 BC to 500 AD, maybe including earlier poems). It could be presumed that the *positive* meaning traces to the very earliest phase of agglutination (with no noun-verb distinction) and the original (“universal”) meaning carried by the attachment of person suffixes to a lexical root.¹⁷ The *negative* meaning must be assumed to be a secondary development. At first, the simple original construction would have become more or less obsolete during the flourishing phase of agglutinative morphology (inserting many specifiers, the proto-voice, tense, mood markers, between the root and the person suffix). Later on, for the lexemes categorially fixed as ‘verbs’, the construction was reinstated as part of the verbal system. The old – obsolete –, inherently bi-morphemic, construction was free to be reanalysed. In contrast with the newly developed positive forms of the structure <verb - ‘tense’ - pers.> and in analogy with the paradigm of inherently negative verbs (cf. Section 4.3.3) it was viewed as ‘missing some specification’ in the medial position and reanalysed into a three morph structure <verb - Ø- pers.>. This lack of specifier could be interpreted, in analogy with the lexical negative paradigm, as signifying the non-existence of the process denoted by the verb root (cf. Section 5 for a more precise analysis).

In this view, the polarity ambiguity of the construction root+person observed in the Sangam is assumed to be due mainly to the composite nature

of the corpus. The two meanings represent two distinct points (initially: positive, finally: negative) of a grammatical evolution. There is no reason to assume that in any real speech form the two opposite values co-existed. It was not a gradual process but a shift of meaning due to the reanalysis of the structure¹⁸ which came to be interpreted as a negative conjugation.

4.3.2 *Polarity ambiguity of the *-ā morpheme*

Similarly, strong evidence that *-ā cannot be accepted as an original Proto-Dravidian negative is found in these Old Tamil data. Suffixed to a bare stem, *-ā fills various modifying functions where it is attested with both positive and negative meanings. This point was clearly stated in Rajam (1992):

“As a suffix, it always provides an unaccomplished sense in adjectival or adverbial participles but an incomplete sense in an infinitive. Specifically, it provides a **negative** sense in an adjectival or adverbial participle but a **positive** one in an infinitive.” (1992: 855; [emphasis mine])

e. g. *ōr-ā* ‘listening’ vs. *āk-ā* ‘that which does not become’ (1992: 865–866)

“[...] there is no structural indicator to separate the two kinds of *ā* [...]. Both are followed by another verb. Context determines the meaning in such cases” (1992: 859)

An additional argument comes from Agesthalingom (1977: 120) who notes that in the Sangam corpus “the pattern *ceyyā* [‘to do’ + *ā*] denotes both negative and positive, though the latter is far predominant”, confirming the orientation of the shift from positive towards negative, which became regular later on.

It should also be noted that the suffix *-ā* appears with inherently negative verbs without changing their polarity: ¹*illā* ²*nam* ‘we ¹who cannot exist’ (Rajam 1992: 864).

As explained above in the quote by Rajam, the polarity ambiguity of the *-ā* morph seems to stem from two distinct developments of a common value of unaccomplishment/incompleteness. In contrast to what we proposed for the negative paradigm, it is not impossible that the two opposite polar values co-existed for some time at the beginning of the functional differentiation (participle vs. infinitive).

4.3.3 *Structural analogy between negative lexical verbs and the zero negative paradigm*

A crucial step in the construction of the zero negative paradigm was the alignment of forms of the general verb class with those of inherently neg-

ative verbs. While positive forms of finite verbs stabilized their medial morphological material into regular sets of tense / aspect morphs, negative lexical verbs, *il-*, *al-* which remained basically ‘tenseless’ did not acquire these tense / aspect specifications and continued to be built on the simple schema: <root + pers.> *all-ēn*, *all-āi*, ... This structural model favoured the *negative* interpretation of the *finite construction*: <root + pers.> for the other verbs.

(10) *ill-ēn*, ‘I am not’, *ill-āi* ‘you are not’, etc.

(11) *ceyy-ēn*, ‘I do / shall / did not do’ *ceyy-āi* ‘you do / will / did not do’, etc.

The negative conjugation was established, identical to that of negative lexical verbs. Nouns, built on the same model (root+person, cf. Table 8 on page 277) were not affected by this ‘negative’ interpretation as the noun-verb distinction was already well established.

The analogy extended also to the *non-finite forms*, but in this case it seems that it was the negative lexical verbs which were aligned on other verbs, with the generalization of *-āt-* (from **ā* in Section 4.3.2 + *-t-* regularly present in the old aorist positive participle) morph:

(12) *ceyy-āt-a* ‘not doing’ (adjectival participle)

(13) *ill-āt-a* ‘not being’ (id.).

The stability of these negative forms (negative lexical or conjugated negatively) lasted for centuries. The loss of person inflection on negative lexical verbs (cf. Section 2.4) weakened the zero negative paradigm which became marginal and tended to disappear.

This detailed historical investigation reveals the complex situation which allowed the zero negative paradigm to emerge in the context of a major typological shift which developed in several stages until the morphological conditions were set for distinct nominal and verbal morphologies, in addition to a particular feature of the Dravidian languages: the existence of inherently negative verbs conjugated in persons. The change of polarity was triggered by *analogy* – a common process in historical linguistics – with the paradigm of negative lexical verbs.

These complex conditions fit well with a recent proposal made by Harris (2008: 76) claiming that

“[t]ypologically unusual constructions can be explained in terms of their origin. [...] It is the fact that so many specific factors or changes must co-occur or occur sequentially in an appropriate order that explains the infrequency of these constructions”.

The Dravidian zero negative paradigm is indeed a “historical accident”, but not a phonetic accident as assumed by most Dravidianists.

Scrutinizing the Early Old Tamil data also revealed that the $-\bar{a}$ marker which became the main negative marker in Dravidian did not originally have a negative meaning. Therefore we end with new questions:

- (i) If neither of the two markers \bar{a} and \emptyset are true negators, then how is negation expressed in Dravidian?
- (ii) If \bar{a} and \emptyset are not related phonetically, then why can they commute in the expression of morphological negation?
- (iii) If $-\bar{a}$ is not originally a negator, what is its origin?

Many more detailed questions could be added!

5 Construction and conceptualisation of the negation in Dravidian

In this section, I would like to show, without denying the originality of the Dravidian zero negative paradigm, that it can nevertheless be explained by some general theoretical frameworks and be supported by typological parallels. Two distinct aspects of the conceptualisation of negation in Dravidian will be addressed: (i) the grammatical strategy used to construct the negative meaning and (ii) the semantics of the negative markers.

To answer these questions we shall take support from diverse theoretical proposals on verb systems developments (Guillaume 1929) and on negation (Cuilioli 1990) as well as from typological studies on negation (Forest 1993; Miestamo 2005, 2010).

The main theoretical framework I will use to explain the zero negative is an old one, developed by Guillaume in *Temps et verbe* (1929). This work is part of a more general – and non main stream – linguistic theory called ‘*Psychomecanique du langage*’¹⁹ which encompasses our modern historical, typological and cognitive approaches to language. This work, read by chance, gave me the initial hints to explore the question (Pilot-Raichoor 1998) and later appeared to provide the most comprehensive explanation to the Dravidian zero negative.

Typological studies (Forest 1993; Miestamo 2005) abundantly illustrate the fact that the linguistic expression of negation is far more diversified and complex than the simple addition of a negator to a positive sentence. From the preceding section, it was established that neither the zero morph nor the $-\bar{a}$ morph (cf. Section 4.3.2) involved in the morphological negative

forms can be reconstructed as Proto-Dravidian negators. This fact indicates that negation in Dravidian should be analysed in light of the various indirect strategies used in languages to express negation. More precisely, as argued by Miestamo (2010), Dravidian negation belongs to a subgroup of languages using strategies which apparently lack an overt negator.

5.1 Conceptual unity of zero-type and \bar{a} -type paradigms

As stressed by Master (1946), whatever hypothesis we admit for the origin of the zero negative, the very existence of this construction is a linguistic question which should be addressed in the scope of “General, Universal, or Philosophical Grammar” (1946: 136), adding further, against the phonetic accident hypothesis, that “[g]enerally speaking, an hypothesis which assumes a persistent tendency of a negative particle to eradicate itself, is logically obnoxious” (1946: 146).

The general problematics of the zero negative is complex. Miestamo (2010: 171) identifies “two basic aspects that need to be discussed: that the marking of tense or some other categories is absent and that there is no overt marker of negation”. In addition to these, a Dravidianist would add the question of the alternation between $*\bar{a}$ and \emptyset as morphological negative markers. These three questions are, in fact, tightly connected, but the inaccurate descriptions of negation proposed by most Dravidianists prevented typologists from perceiving their fundamental relation. The following lines will try to elucidate this point.

5.1.1 *The construction of morphological negation in Dravidian*

The typologies mentioned above distinguish two main types of negation: ‘symmetric’ and ‘asymmetric’ in Miestamo (2005), ‘recusative’ and ‘suspensive-reassertive’ in Forest (1993). The distinctions made by these authors do not exactly match (cf. Miestamo 2005: 22–24), but what is important here is that both distinguish a type called ‘symmetric’ or ‘recusative’, where negative expressions differ from the positive ones only by the addition of an overt negator, from another type, ‘asymmetric’ or ‘recusative-suspensive’, which implies other differences between positive and negative expressions. The characteristic Dravidian constructions of negation are assigned to this latter type by both authors.²⁰ However, the explanations proposed by the two authors differ slightly and bring distinct insights.

Miestamo's typological study (2005) rightly assigns the Dravidian negative constructions to the general type of constructional and paradigmatic asymmetry affecting the tense-aspect-mood categories. More precisely, in Miestamo (2010), he introduces a 'subtracting zero negative construction', a sub-type not attested in the 2005 sample of languages. Based on the lack of a 'tense' morph in the zero negative paradigm, his analysis is in the line of the one originally proposed by Caldwell (1856) who also starts from the systemic lack of tense in the construction to develop his brilliant explanation of the 'logic' of the zero negative:

"What is the rationale of this negative? The absence of signs of tense appears to contribute to the expression of the idea of negation: it may at least be said that it precludes the signification of the affirmative. In consequence of the absence of tense-signs, the idea expressed by the verb is abstracted from the realities of the past, the present and the future: it leaves the region of the actual events, and passes into that of abstractions" (Caldwell 1976: 471).

Forest's approach (1993: 49–51) is based on an analysis of **-ā* as a marker of infinitive (an analysis inexact for the Telugu example given, but true for other languages/constructions) and emphasizes the double movement of the Dravidian construction: suspensive (infinitive) and reassertive (through the person endings). Suspensivity is due to "the proper value of the infinitive, which does not imply the speaker's commitment to the real factual aspect of the process" (1993: 49) and reassertion is accurately described as "the proper function of the person endings in Telugu: the simple attachment of the process to the 'sphere' of a participant or a non-participant to the speech-act, without assuming that the process is factual/real" (1993: 50). His analysis echoes an earlier detailed proposal made by Kittel (1903). Also starting from an inexact reconstruction of an **-ā* marker in Kannada, Kittel gave us precious indications on the way the negative meaning is obtained in Dravidian constructions.

"There can be no doubt that the origin of the conjugated negative is based on the so-called infinitive ending in *-a* in the same manner as that of the negative participle is. That infinitive originally was a verbal noun and only in course of time came to get its specific meaning. Thus, *nōḍa* at first meant 'seeing', 'a seeing' and thereupon 'to see', 'about to see', 'yet to see'. *nōḍem* (*nōḍa+em*) therefore signifies 'a yet to see-I', i. e. my seeing (is or was) yet to be or (will be) yet to be, or my seeing (is) not actually existing, (was) not so, or (will) not be so, whence we arrive at the meaning 'I do not see', 'I did not see', '(I have not seen)', 'I shall not see'." (1903: 161)

These analyses, proposed independently by Kittel and Forest, are congruent and offer crucial clues to the semantic construction of negation in Dravidian. The major insight of these latter analyses is the relative independence of the two components of the construction: the verb and the person endings. There is a simple ‘attachment’ of a verbal notion to the ‘sphere’ of a given entity represented in the ‘person ending’. Most importantly, as stressed in Kittel’s translations, there is a distinct encoding of time in each of the components.

For the verb, the time value is fixed: ‘to see’, ‘about to see’, ‘yet to see’ and clearly refers to specific *phase* in the time-line of a verb (see Section 5.2.1 regarding Guillaume’s ‘verb image’): the process denoted by the verb has not yet started. The verb component carries the representation of an abstract, non-evolutive verbal notion, not that of an actual temporal event.

The person endings bring a different – and implicit – time indication: these suffixes represent entities which are located in some temporal period viewed from the speech time. Thus it is in fact these suffixes which give the ‘tense’ interpretation to the whole construction according to the positioning of the subject (‘I/you/he...’) in the present (‘is’), the past (‘was’) or the future (‘will/shall’).

All the finite forms with person distinction will in fact carry this double time reference. Let us refer to the time encoded in the negative/tense morphemes as *aspectual-time* and the time implicitly emerging from the person affixes as *localizing-time*.

5.1.2 Relation between **-ā-* and *-∅-* in the construction of negation

This clarification about the construction of the negative meaning in the Dravidian finite forms allows us to solve the major question of the relation between **-ā-* and *-∅-* which has hindered Dravidianists for so long. As stated earlier, the marker **-ā-* was not originally a sign of negation, but a kind of ‘suspensive’ specifier as is confirmed by historical data (see Section 4.3.2). Initially the form *ceyy-ā* contrasted with *ceyy-um*, one denoting a potential event, the other a real event (cf. Section 4.2). When the personal constructions developed on the schema <verb - ‘tense’ - person>, this contrast crystallized into the ‘tensed’ – or more precisely, aspectual – stems (Past/Non-Past) denoting the verbal representation of a *process* and the verbal noun/infinite stem carrying only the meaning of the verbal *notion*.

At this point, we may answer Pederson’s first question (1993: 234): “Why is there an acceptable alternation between *-ā* and *∅*?” The *∅* can be substituted

so easily for *-ā-* because the strategy of negation used in Dravidian does not require the presence of a negator, but the indication that the verb form inserted in the construction does not represent a real process. This indication is given explicitly by the use of the *infinitive*, a nominal form of the verb and implicitly by the use of the *bare stem*. Both are functionally and semantically close: neither of them imply the reality/factuality of the process. The bare stem is totally underdetermined while the infinitive/verbal noun is more marked, signaling the *notional* value of the lexeme. In the reasoning developed above both function equally well: what is important is that they indicate the non-factual/non-real character of the process denoted by the lexical verb.

Pederson's second question (1993: 234): "Why not always just { \emptyset }?" may also be answered. Morphologically, the zero marked negative can operate only in one finite construction where it contrasts with all the other structurally identical constructions which are overtly marked as positive by a segmental morpheme. Semantically, it is the internal conflict between two distinct time components that gives the construction its negative meaning: (1) the stative/abstract – temporally null – aspect of the verb conflicts with (2) the assertion (indicative mood) of a predicative relation between this verb representation and a subject (encoded in the person suffix) situated in an actual temporal context (present/future or past, as in Kittel's analysis). When these conditions are not met, in all other finite (injunctive or devoid of person endings) and non-finite (preasserted) negative constructions a more overt expression of the non processive status of the verb is required and thus the **-ā* marker became identified as an explicit sign of negation.

The two lines of analysis sketched above, Miestamo/Caldwell and Forest/Kittel converge on the fact that what is involved in negative constructions is a fundamental relation of the verb to the representation of time. The relation between time and negation is not so straightforward and requires further theoretical justification.

5.2 A unified approach to Dravidian morphological negation in Guillaume's framework

Most of the singular aspects of the Dravidian data discussed so far find theoretical support in some fundamental ideas proposed by Guillaume in *Temps et verbe* (1929) and *L'architectonique du temps dans les langues classiques* (1945), notably the alignment of negative and 'tense' (aspectual) markers and the diachronic evolution of the systems.

5.2.1 Guillaume's key ideas of 'verb image' and 'chronogenesis'

The first important idea for our explanation is Guillaume's basic representation of the evolution of the 'verb image' in time. He expresses the 'career' of a verb in terms of tension / detension: at the initial phase (I)/ t^0 , 'the verb has its entire tension in front of it. Nothing of this tension has yet been spent' and, at the final phase (F)/ t^n , all its tension is exhausted. The medial phase (M), represent the timespan (instantaneous or enduring) during which the process is active. The following simplified schematization (Figure 1) is derived from Guillaume's (1929: 18) original illustration with the French verb *marcher*.

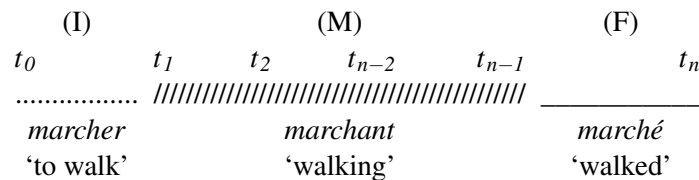


Figure 1.

A second fundamental idea proposed by Guillaume is that of *chronogenesis*. Roughly speaking we may say that for Guillaume chronogenesis is a mental activity which brings the verb from its basic image of a process (as above) to the achieved characterisation of the verb.

“Aspect, mood, tense do not refer [...] to phenomena of different nature, but to internal phases of a phenomenon of a unique nature: the chonogenesis: in a word, aspect, mood, tense represent a single thing considered at different moments of its own characterisation’ (Guillaume 1929: 11).

This idea is linked to another, that of *chronothesis* identifying specific moments in the development of the characterisation of the verb. For French, Guillaume distinguishes three developmental positions: initial (tenses *in posse*), median (tenses *in fieri*) and final (tense *in esse*). “The realisation of the verb in the tenses *in posse* gives rise to the nominal moods (infinitive and participle)” (Guillaume 1929: 11) and the last position (*in esse*) to the indicative mood. The middle position (*in fieri*) gives rise in French to the subjunctive forms.

In a diachronic perspective, Guillaume aimed to develop, for the verb systems, “means of representation [...] fit to show the historical successivity, evolving or revolving” (1945: 65). In this view one of the interpretation of

chronogenesis is to identify verb systems as ‘stratified structures’, i.e. ‘complex mental constructions’ in which “some elements [are] acquired only after others [have] been acquired” (Hewson and Bubeník 1997:2).

5.2.2 Interpretation of the Dravidian data in the light of Guillaume’s ideas

The stratified development of verb forms adequately matches Meenakshisundaran’s view (1965: 27; cf. Section 4.2 (ii)) that the Tamil verb develops in two strata. Traces of the typological shift left in the earliest Dravidian data offer the rare opportunity to watch the birth and growth of a new system.

Original stage: no morphological variation, lexical root

First stratum: (Section 4.2) *ceyy-um* (?aorist) vs. *ceyy-a/-al/...* (?optative)

Development of agglutination

Second stratum: attachment of the person suffix and elaboration of the ternary contrast

The historical data also help to understand the development of the verb categories. Adopting the previous schematization, it can be proposed that from the first stratum emerged a primary distinction between a kind of realis (*ceyy-um*) and a kind of irrealis (*ceyy-a*). The next distinction emerged from the splitting of the realis form in two values Non-Past and Past and the reinterpretation of the irrealis form as an abstract/notional representation of the verb. Adopting the previous schematization (Figure 1 on the preceding page), the evolution can be illustrated in Figure 2.

One key element in this early development of the verb system is the contrast between a non-processive representation of the verb (*ceyy-/ā-*) and the representation of the verb as a process associated with a marker (*-v-*, *-t-*) encoding some aspects of the time it requires for its realization.²¹

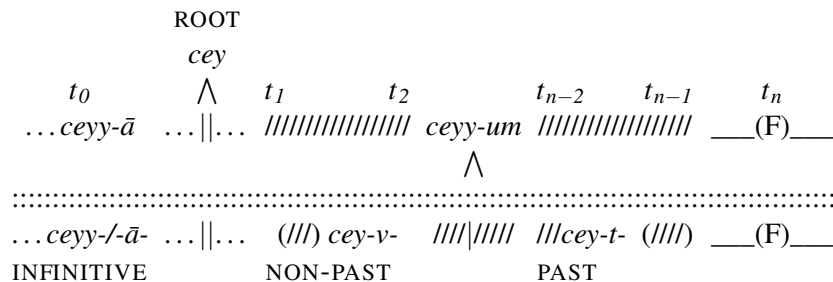


Figure 2.

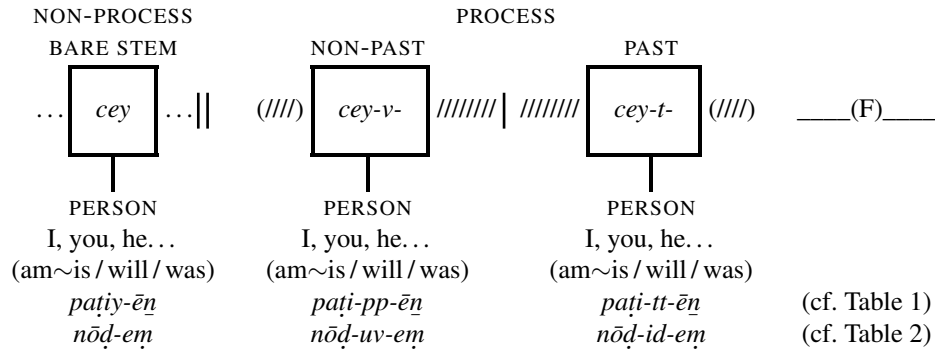


Figure 3.

The last schema (Figure 3) represents the stage of categorial distinctions obtained at the end of the agglutinative phase and verb-noun distinction. In line with the Kittel / Forest proposal detailed above (Section 5.1.1), each stem form (Bare stem / Infinitive, Non-Past and Past) is given a new ‘verbal’ interpretation when a person ending is ‘attached’ to it.

The construction of their original meaning can be modelled on the one proposed above for the negative: ‘I (am / was / will) + V (in a state of non-processivity) = *Negative* ‘I do / did / will not V’, similarly, for the *Non-Past tense*: ‘I (am / was / will) + V (in a state of unaccomplishment) = ‘I am / was / will be V-ing’, often ‘I will / can V’ and for the *Past tense*: ‘I (am / was / will) + V (in a state of accomplishment, not necessarily achieved) = ‘I did V / I (will) have V-ed’, often ‘I did / have done V’. In Guillaume’s chronogenetic view, the indicative finite forms represented in these schemata offer the most characterised forms at this historical moment.

This stage of development is the one reconstructed by the comparative works which set at the base of the Dravidian verbal systems a ternary contrast between two positive tenses ‘Past’ and ‘Non-Past’ and a ‘Negative’ one (cf. Subrahmanyam 1971; Zvelebil et al. 1990: 31; Krishnamurti 2003: 291) as illustrated by the finite forms in Tables 1, 2 and 5.

From this sketchy presentation, we can see that Guillaume’s framework offers a fairly simple explanation to most of the questions raised by the Dravidian negative construction.

His dynamic approach to the construction of the verb systems (‘evolving and revolving’) allows for the *creation of new strata*, with sometimes a drastic change in the meaning of the forms / components.

The basic ‘verb image’ provides a crucial clue to justify the positioning of the so-called ‘negative’, but actually ‘abstracting’ morphemes (*-ā* ‘infinitive’ / *-∅* ‘bare stem’) in paradigm with the ‘tense’ morphemes.

The idea of chronogenesis, assuming that the verbs forms attain their full and final characterisation in the tenses *in esse*, i. e. the finite personal forms in Dravidian, gives support to Kittel/Forest analysis which views the person endings, representing entities situated in a given spatio-temporal ‘sphere’ as the carriers of the real ‘tense’ specification. In addition, the layered constructions of the forms allows for the *double encoding of time*: aspectual in the verb stem vs. localizing in the person suffix. A key element in the construction of the negative meaning as we have seen in Section 5.1.

5.3 Additional theoretical and typological supports to the Dravidian negative construction

Among all the languages analysed so far through various typologies, the morphological construction of the Dravidian zero negative appears to be unique. However, the semantics of the Dravidian negation do point to a few similar constructs.

Before presenting these language-specific similarities, it is worth mentioning a brief study on negation where Culioli (1990) presents an unusual approach to the question. Though he does not mention Dravidian languages, his model offers more interesting theoretical support to the zero negative construction. Among other crucial ideas, Culioli (1990: 95) distinguishes two basic types of negation, one of which is “the negation of localisation and of existence”²², “a grammatical process which shows that there is an absence, a void”, in reference to “a mode of existence (localized occurrence / no occurrence for a given localization [. . .])” (1990: 94). The linking of the two crucial notions of ‘existence’ and ‘localization’ closely matches the semantics of the Dravidian zero negative. The existential component of negation, also noticed in Miestamo (2005: 96) under the correlation with the ‘reality status’, is encoded in Dravidian, in the ‘aspectual’ stem of the verb, for which the *-∅*- allomorph is certainly the most exact grammatical coding of the idea of ‘absence’ or ‘void’ one could imagine! The locational component, encoded indirectly in the person suffixes, is also fundamental as it is only by its (conflicting) presence that the negative interpretation of the construction can be construed. The Dravidian negative construction precisely expresses that at a given localization there is no occurrence of the event denoted by the verb lexeme.

A few more specific similarities can be found in various languages.

Regarding the **ā* negative construction, two distinct types of correlation may be invoked: one – fairly frequent – with the interrogative markers (Miestamo 1995: 100–101; Forest 1993: 48–49); the other with various means expressing a ‘distance’ (‘distance marker’ Miestamo 1995: 136–137; ‘distanciateur’ [‘distancing marker’] Forest 1993: 48, 51, 142, ‘métaphore spatiale’ [‘spatial metaphor’], 1993: 62). In Dravidian, there is a formal identity between the negative morpheme **-ā-*, the interrogative particle **-ā* and the remote deictic lexical base **ā* that may lend support to Gundert’s hypothesis (given in Caldwell [1976: 478]) proposing to

“derive the negative meaning itself from the interrogative, and ultimately from the [remote] demonstrative [**ā*]²³. [...] From the meaning of a question comes the meaning of negation. *adu varum-ā?* will it happen? = it will never happen”.

The occurrence of a similar chaining between *localisation* – *interrogation* – *negation* in a variety of Nahuatl (San Miguel Tzinacapan, Mexico) is described in Pury-Toumi (1982).

Regarding the *∅*-negation, though it is not sure that a similar morphological encoding may ever be found, some lexical parallels around the notion of a spatio-temporal ‘void’, may exist. Thus Culioli (1990: 95) calls attention to the Vietnamese negative marker, actually the word *không* signifying ‘void, empty, zero’.

6 Conclusion

I hope to have demonstrated that the Dravidian zero negative construction, a well attested linguistic fact in several languages, emerged under complex and specific historical circumstances which can be summarized as follows.

In the broad context of a major typological shift, distinct grammatical categories and paradigms were developed.

The finite positive forms of the verbs were all specified by some ‘actualizing’ material (tense, aspect, mood). The only forms which were devoid of this material were those of the negative lexical verbs.

The zero negative {root-pers.} paradigm can be explained only as a secondary development in a phase of systemic reorganization. It was built analogically to the lexical negative verbs which favoured the negative interpretation of the zero, the ‘missing’ material between the root and the person.

The simultaneous occurrence of this set of conditioning factors being rare, the Dravidian zero negative remains unique.

The long and erratic linguistic story of the Dravidian zero negative is interesting *per se* as it alerts us also on the complex relationship between theory and data. A well-established and efficient theory such as comparative grammar in this case, may have a blinding effect on the observable data: all leading, sharp-minded Dravidianists in some way denied the zero negative construction. Master's voice calling for a more general approach to the question was not heard. But it is indeed from less prevalent, though insightful and general theories (Guillaume, Culioli) that we can draw the best interpretative arguments.

The recent developments of typology have broadened our views on the negative strategies and allow us to include the Dravidian case into a particular class of negative strategies which do not require the presence of a true negator, see Miestamo (2010).

Due to these analyses, the zero negative, rejected as a morphological monster by most Dravidianists, may well appear as a perfectly logic product of the human mind, with a zero iconically encoding the temporal void as a pure structural expression of the idea of a non-event.

Abbreviations

∅ = zero negative marker; 1P = 1st person plural; 1S = 1st person singular; 2P = 2nd person plural; 2S = 2nd person singular; 3PM/F = 3rd person plural masculine or feminine; 3PN = 3rd person plural neuter; 3SF = 3rd person singular feminine; 3SF/M = 3rd person singular feminine or neuter; 3SM = 3rd person singular masculine; 3SN = 3rd person singular neuter; ACC = accusative case marker; CVB = converb; DAT = dative case marker; Drav. = Dravidian; GEN = genitive case marker; honor. = honorific; NMLZ = nominalized verb; pers. = person; sfx. = suffix; vb. = verb

Notes

1. 'Sangam' is the name of an important compilation of more than two thousand poems (grouped in two collections of anthologies) which were composed approximately between 150 BC to 500 AD, cf. Rajam (1992). Abbreviations such as [pura.399:14], [kuru.178:3-4], etc. refer to these anthologies.

2. For instance in Gondi, comparison of the Present-Future and Negative paradigms (Subrahmanyam 1971: 268, 354) shows zero morphs for some persons: vb+*ānt*-(Pres.-Fut) versus. vb+∅ (Neg.)+ *-ōn* (1S), *-ōm* (1P), *-ōr* (3M), *-ēr* (3MP), e. g. *sūr-ānt-ōn* 'I see' vs. *tinn-ōn* 'I do not eat'. However, in 3rd pers. non-masculine, the person suffixes differ: *-ā* (sg.); *-ān* (pl.) in Pres.-Fut vs. *-ō*, *-ōn* in Negative and in 2nd pers., a segment *-v-*, representing a variant of the negative marker, is inserted before the person suffix, e. g. *tin-v-ī* (2S), *tin-v-īt* (2P).
3. Except two North-Dravidian languages: Kuḍux and Malto which have lost the morphological negation under influence of the Indo-Aryan languages (cf. Subrahmanyam 1971: 381).
4. The negative roots **il* 'not to be', **al* 'not to be so' were inflected for person in Old Tamil, *il-ēn* 'I am not', *il-āy* 'you are not', etc.
5. From a verb *māṭṭu* 'to fasten, tackle, be able... ', it takes the meaning 'cannot, will not' as a negative auxiliary.
6. There may be also traces of it in the use of variants of the negative auxiliary **tō*, *to?*, *tōt* (from a verb root *tōr* 'to be visible, ... , to exist') found in several Central Dravidian languages, e. g. Kolami, Naiki, Konda.
7. In Toda, as in several other Nilgiri languages, a different verb stem, called S2 by Emeneau is used the Indicative or 'realis' mood. This S2 is derived from the so-called 'Past' stem, reconstructed by Comparative methodology, e. g. Aorist: *pod-en* < *po-r+t+en*.
8. For instance, the glottal segment */-ʔ-/*, as in Konda *ki-ʔ-e* 'I do not do', is usually considered as a reflex of **v* (< negative marker **vā*), but it also occurs with the inherently negative verbs, e. g. Kuvi *pāyi hil-ʔ-o* (< **cil* 'to be not') 'I am not beating'.
9. Master (1946: 143) gives the same analysis for the three languages.
10. Meenakshisundaram stressed the limitations of the comparative approach in Dravidian, noticing that it 'is more like the comparative study of the Romance languages, where the splitting of the Proto-language could not be carried into the hoary past as one could with the Indo-European (1965: 11–12)
11. Approximate dates are: Tamil inscriptions 3rd century BC, literature 150 BC, Kannada inscriptions 450 AD literature 900 AD, Telugu inscriptions 650 AD, literature 11th century AD Malayalam records became distinct from Tamil between the 9th and the 11th centuries AD. Most of the unwritten languages (all the Central and North Dravidian languages) were recorded from the 19th century onwards.
12. Mahadevan's comment on *antai*: "a frequent honorific affix (masc.) [...] occurs as an independent word in T[amil]B[rahmi] inscriptions, but only as a bound suffix in Literary Tamil. *antai* is arguably the most significant of the 'lost' words in Old Tamil recovered through the decipherment of the Tamil-Brāhmī inscriptions"
13. Meenakshisundaran (1965: 2) notes "the addition of new morphemes in the place or in addition to older morphs, or changes in the arrangements of the morphs."
14. remnant of an earlier distinction sg.: *ceyyun*, pl.: *ceyyum*
15. "[...] these formatives, later on coming to be utilised as tense sign, at the earlier stage should have had some differences in meaning, but unfortunately that could not be explained as matters stand at present" (Meenakshisundaran 1965: 31–32).
16. In the inscriptional corpus studied by Mahadevan (2003) there are no 'finite verbs': only the polyfunctional 'participial nouns' are attested.

17. Meenakshisundaram (1965: 3) notes that “the subject followed the predicate very frequently in ancient times” — allowing for the person subjects to become suffixed to the lexical root.
18. It should also be noticed that there is absolutely no trace in these oldest records of a transitory stage as it would be expected in the case of the phonetic fusion of **ā* with the person suffixes (from a conjectural ***kāṇ-ā-īr* > ***kāṇ-a-īr* > ***kāṇ-e-īr* **‘see-ā-2p’ towards the attested *kāṇ-īr*).
19. See Hirtle (2007) for an English introduction to Guillaume’s theory.
20. Exceptions are found in some Central/North Dravidian languages which developed a recusative type of negation under the influence of Indo-Aryan/Iranian languages.
21. This schematisation justifies and emphasizes the particular aspectual values encoded in the two stems. Non-Past and Past differ only slightly, the first one asserting the non-accomplished part of the process, the latter one its accomplished part, without fixing any precise boundaries to the process. Contrary to the French word *marché* the Dravidian Past stem does not represent the process as ‘exhausted’.
22. the other one is “la négation qualitative de rejet par le sujet de ce qui est mauvais” (Culioli 1990: 95).
23. In Dravidian, the deictic origin of the *ā* marker may be worth exploring as in parallel with two other vocalic ‘tense’ markers occurring in various formations -*i*- (Past) and -*uv*- (Non-Past), -*utt*- (Present in Kannada), may be also -*u*- in the earliest aorist forms (sg. *ceyy-u-n* / pl. *ceyy-u-m* ?) there are two deictic bases *i*- ‘proximate’ and *u*- ‘intermediate’. In the perspective of the typological shift, re-use of older deictic spatio-temporal ‘particles’ in the construction of the new verb forms may not be ruled out.

References

- Agesthalingom, S.
 1977 *A Grammar of Old Tamil with special reference to Patirrupattu: Phonology and verb morphology*. Annamalainagar: Annamalai University.
- Arden, Albert H.
 [1976] *A Progressive Grammar of the Tamil Language*. [repr. 5th ed. 1942, rev. by A. C. Clayton]. Madras. The Christian Literature Society. [original publication: 1891].
- Bloch, Jules
 1946 *Structure grammaticale des langues dravidiennes*. Paris: Adrien-Maisonneuve.
- Bloch, Jules
 1935 La forme négative du verbe dravidien. *Bulletin de la Société de Linguistique de Paris* 36: 155–162.
- Caldwell, Robert
 [1976] *A comparative grammar of the Dravidian or South-Indian family of languages*. [3rd ed. 1913] revised by J. L. Wyatt and T. Ramakrishna Pillai. Madras: University of Madras. [original publication: 1856].
- Culioli, Antoine
 1990 La Négation: marqueurs et opérations. In *Pour une linguistique de l'énonciation*; Tome 1, Antoine Culioli (ed.), 91–113. Paris: Ophrys.

- Emeneau, Murray B.
 1984 *Toda Grammar and Texts*. (= Memoirs of the American Philosophical Society; 155). Philadelphia: American Philosophical Society.
- Forest, Robert
 1993 *Négations. Essai de syntaxe et de typologie linguistique*. Paris: Klincksieck.
- Guillaume, Gustave
 1929 *Temps et verbe*, Paris, Champion. [2nd ed. 1965].
- Guillaume, Gustave
 1945 *L'architectonique du temps dans les langues classiques*. Copenhagen: Einar Munksgaard.
- Harris, Alice C.
 2008 On the Explanation of Typologically Unusual Structures. In *Linguistic Universals and Language Change*, Jeff Good (ed.), 54–76. Oxford: Oxford University Press.
- Hewson, John and Vít Bubeník
 1997 *Tense and aspect in Indo-European languages: Theory, typology, diachrony*. Amsterdam: Benjamins.
- Hirtle, Walter H.
 2007 *Language in the mind: An introduction to Guillaume's theory*. Montreal: McGill-Queen's University Press.
- Kittel, Ferdinand
 1903 *A Grammar of the Kannada Language in English comprising the three dialects of the language (ancient, mediaeval and modern)*. Mangalore: Basel Mission Book and Tract Depository.
- Krishnamurti, Bhadriraju
 1961 *Telugu verbal bases. A Comparative and Descriptive Study*. Delhi/Vanarasi/Patna: Motilal Banarsidass.
- Krishnamurti, Bhadriraju
 2003 *The Dravidian languages*. Cambridge: Cambridge University Press.
- Lehmann, Thomas
 1998 Old Tamil. In *The Dravidian languages*, Sanford B. Steever (ed.), 75–99. London/New York: Routledge.
- Mahadevan, Iravatham
 2003 *Early Tamil Epigraphy. From the Earliest Times to the Sixth Century A.D.* Chennai, India: Cre-A; Cambridge, MA: Department of Sanskrit and Indian Studies, Harvard University.
- Master, Alfred
 1946 The zero negative in Dravidian. *Transactions of the Philological Society* (London) 45: 137–155.
- Meenakshisundaram, T. P.
 1965 *A history of Tamil language*. (= Building Centenary and Silver Jubilee Series; 22). Poona: Deccan College.
- Miestamo, Matti
 2005 *Standard Negation: The Negation of Declarative Verbal Main Clauses in a Typological Perspective*. (= Empirical Approaches to Language Typology; 31). Berlin/New York: Mouton de Gruyter.

- Miestamo, Matti
 2010 Negatives without negators. In *Rethinking Universals: How rarities affect linguistic theory*. Jan Wohlgemuth and Michael Cysouw (eds.), 169–194. (= Empirical Approaches to Language Typology; 45). Berlin / New York: Mouton de Gruyter.
- Murugaiyan, Appasamy and Christiane Pilot-Raichoor
 2004 Les prédictions indifférenciées en dravidien: témoins d'une évolution typologique archaïque. In *Mémoires de la Société de Linguistique de Paris. Les constituants prédictifs et la diversité des langues*. (= Mémoires de la Société de linguistique de Paris. Nouvelle Série; 14), 155–177. Leuven: Peeters.
- Pederson, Eric
 1993 Zero negation in South Dravidian. In *CLS 27. Part Two: The Parasession on Negation*, Lise M. Dobrin, Lynn Nichols and Rosa M. Rodriguez (eds.), 233–245. Chicago: Chicago Linguistic Society.
- Pilot-Raichoor, Christiane
 1998 Temps et négation en Dravidien: Passé / Non-Passé / Négatif. Sémantique d'un système morphologique. In *Proceedings of the 16th International Congress of Linguists*, Société de Linguistique de Paris (ed.). CD-ROM. Oxford: Pergamon. Paper N° 0204.
- Pury-Toumi Sybille de
 1982 Quand oui c'est non et non c'est où. *Amérindia* 7: 23.–38.
- Rajam, V. S.
 1992 *A Reference Grammar of Classical Tamil Poetry (150 B.C.–pre-fifth/sixth century A.D.)*. Philadelphia: American Philosophical Society.
- Sridhar, S. N.
 1990 *Kannada*. London: Routledge.
- Subrahmanyam, P. S.
 1971 *Dravidian Verb Morphology: A Comparative Study*. Annamalainagar (Tamil Nadu): Annamalai University.
- Zvelebil, Kamil, Jurij Glasov and Michail Andronov
 1967 *Introduction to the Historical Grammar of the Tamil Language. Pt. 1: Preliminary Textual Analysis*. Moscow, Nauka.

Relative root complement: a unique grammatical relation in Algonquian syntax

Richard A. Rhodes

1 Introduction

It is very probably the case that every human language contains some pattern of applicative syntax, i. e., that pairs of synonymous clauses can be found in which one contains an object that corresponds to an oblique nominal in the other. For example,

- (1) a. *He bought a latté for me.* (oblique syntax)
b. *He bought me a latté.* (applicative syntax)

Most of the syntactic complexity in applicative constructions centers around the nature of the grammatical relation borne by the nominal corresponding to the object in the oblique syntax. In some languages that question is further clouded by the fact that applicative syntax is obligatory, i. e., no oblique syntax is possible with certain classes of notional cases.

Algonquian languages show this complexity. They all have garden variety applicative constructions, as in the Ottawa (Ojibwe) example in (2).

- (2) *Wdoozhtamwaan waa-gwinid.*
od_i-ozhit-amaw-aa_j-an_j change \ wii=agwi-ini_j-d_j
3.ERG_i-make-APPL-3.AN.P.OBJ_j-OBV_j REL\FUT=wear-OBV_j-3.SBJ.CONJ_j
'She is making it for her so she can wear it.' (Bl S 630)¹

The syntactic properties in Algonquian languages for the type of construction exemplified in (2) are:

1. that the applicative construction is obligatory, i. e., there is no non-applicative option, and
2. that the displaced object shows the syntax of a secondary object.

Algonquian languages are all primary-secondary object languages in the sense of Dryer (1986). An extensive discussion of the syntax of secondary objects in Ojibwe is given in Rhodes (1990).

What makes Algonquian languages unique is that they all have a second, and much more common, construction which brings notionally oblique nominals into a closer grammatical relation with the verb, but leaves original objects, if any, unchanged in grammatical relations. An Ojibwe example is given in (3).²

- (3) *Niniing-sh go naa wgii-naabmaan niwi mnidoon.*

*aniniw[-ing]*_{RRC=sh go naa o_i-gii=[iN³]}_{RR-aabam-aa_j-an}
 [man_i-LOC]_{RRC=PTCL} PTCL PTCL 3.ERG_i-PST=[like]_{RR-see-3.AN.P.OBJ_j-OBV}
 [aniwi manidoo_j-an.]_{P.OBJ}
 [this.OBV spirit_j-OBV]_{P.OBJ}

‘He saw the spirit in the form of a man.’ (2R.3.20)

Algonquianists have terminology for describing this phenomenon. The verbal morpheme *iN-* in (3) is called a RELATIVE ROOT.⁴ Relative roots constitute a small, closed class of bound morphemes, listed in (4).

- (4) *akw-* ‘a certain length, so long’
apiit- ‘a certain extent, so much’
daN- ‘in a certain place’ (short gloss: ‘at’)
daSw- ‘of a certain number, so many’
iN- ‘in a certain direction, in a certain way’ (short gloss: ‘to’, ‘like’)
ond- ‘from a certain place, for a certain reason’ (short gloss: ‘from’, ‘because of’)

The nominal or adverbial that fills the slot licensed by a relative root is called a RELATIVE ROOT COMPLEMENT.⁵

The syntax of relative root complements (abbreviated: RRCs) is unique within the range of arguments and adjuncts in Algonquian languages. It can be shown that they are neither objects nor obliques. I will spend the rest of this paper demonstrating this fact. We will discuss the behavior of objects, relative root complements, and obliques with respect to six patterns in Algonquian morphosyntax and syntax. The patterns are 1) verb agreement, 2) the distribution of non-third person pronominals, 3) obviation, 4) understood definite readings, 5) targets of advancement, and 6) accessibility for relative clause formation. By looking at those patterns it will become clear that there are four distinct types of non-subject entities in Algonquian (two types of objects, relative root complements, and obliques). We will look at the patterns one by one, starting with the most complex, verb agreement.

2 Agreement

Algonquian languages are prototypical head-marking languages. Verbs show agreement for subjects, as in (5), primary objects, as in (6), and secondary objects as in (7).

(5) Subject

Mwi binoojiinh

mawi-w_i abinoojiinh_i

cry-3.SBJ_i child_i

‘The child is crying.’ (Bl S108)

(6) Primary object

Ngii-gndaan wiyaas.

ni-gii=gond-am_i-n_i wiyaas_i.

1.SBJ-PST=swallow-INAN.P.OBJ_i-OBJ_i meat_i

‘I swallowed the meat.’ (Bl S395)

(7) Secondary object

Miinwaa mitigoons wgii-nokaazon gii-bshanzhehang ...

miinawaa mitigoons_i o_j-gii=anokaazo-n_i

and stick_i 3.ERG_j-PST=put.to.use-OBJ_i

gii=bashanzheh-amk-g_j

PST=whip-INAN.OBJk-3.SBJ.CONJ_j...

‘And she used a stick to whip it ...’ (Bl T20.16)

But not, as expected, for nominal obliques, as exemplified in (8) and not for relative root complements, as in (9) [= (3) above].

(8) Nominal obliques

a. Locative

..., gii-bbaa-yaamwaad iw bgoji-mnoomin widi waabishkokiing.

gii=bbaa=yaa-m-waa-d iw

PST=around=have-INAN.OBJ_j-3.PL-3.SBJ.CONJ_j that.INAN

bgoji=mnoomin iwidi waabishkokiw-ing

wild=rice_j there swamp-LOC

‘..., and they went around getting wild rice in the swamp.’

(Bl T5.2)

b. Temporal

Ngo-bboon ngii-yaa widi.

ningo=biboon ni-gii=ayaa iwidi

one=winter *1.SBJ-PST=be.at there*

‘I went there **for one year.**’ (Bl S11)

c. Instrumentals

Mii dash gii-nsaabaawdood wewiib iw moozhwegaans iw mshkiki.

mii dash gii=nisaabaawad-ooj-dj wewiib iw

CLEFT PTCL PST=wet-1INAN.OBJi-3.SBJ.CONJi quickly that.1INAN

moozhwegaans; [iw mashkikiw]

handkerchief; [that.1INAN medicine]INS

‘Then she hastily wet the handkerchief **with the medicine.**’

(Bl T30.28)

(9) *Niniung-sh go naa wgii-naabmaan niwi mnidoon.*

[aniniw-ing]RRC=sh go naa oj-gii=[iN]RR-aabam-aa-an

[mani-LOC]RRC=PTCL PTCL PTCL 3.ERGj-PST=[like]RR-see-3.AN.P.OBJj-OBV

niwi manidooj-an.

this.OBV spiritj-OBV

‘He saw the spirit **in the form of a man.**’ (2R.3.20)

The first reaction is to think that the relative root in the verb (*in-* in example (9)) might be just another kind of agreement marker. But although its presence licenses the occurrence of the relative root complement, the morphology of relative roots shows no sensitivity to any property of the nominals whose presence they license. In Algonquian, every agreement marker is sensitive to animacy whether or not it is sensitive to person and/or number, but relative roots do not even register animacy. Compare the inanimate relative root complement in (10) with the animate relative root complement in (9). The relative root has the same form in both cases.

(10) *Ngii-nkwetwaag dash da-nankiiyaan iw.*

ni-gii=nakwetaw-aa-ag dash da=[iN]RR-anokii-yaan

1.SBJ-PST=respond-3.AN.P.OBJ-PL PTCL MODAL=[like]RR-work-1.SG.SBJ.CONJ

[iw.]RRC

[that.1INAN]RRC

‘I told them that I would do **that kind** of work.’ (Bl T9.21)

A full exposition of agreement in Ottawa, as in any Algonquian language, is quite complex. For our purposes, it is enough to observe that in subjects and primary objects, the agreement morphology cross-references gender, person, and number, but in secondary objects, only gender and number are reflected. This property of secondary objects is discussed in Rhodes (1990). Since the focus in this paper is on relative root complements, we will refer the reader to the earlier paper for details on secondary object syntax.

The agreement facts show that subjects and objects are syntactically distinct from relative root complements and obliques. This is summarized in Table 1.

Table 1.

	Primary Subject	Primary Object	Secondary Object	Relative Root Complement	Relative Root Oblique
Full verb agreement	+	+	-	-	-
Gender / number agreement	+	+	+	-	-

3 Restrictions on first and second person

First and second person pronominals are limited in their distribution. They cannot appear as secondary objects, as discussed in Rhodes (1990), and exemplified in (11).

$$(11) \quad Mii \text{ wi } pii \left\{ \begin{array}{l} *giin \\ *niin \\ niwi \end{array} \right\} gaa-ggweji-daawed.$$

*mii iwi apii { *giin, *niin, aniwi }*
 CLEFT that.INAN time { *1, *2, that-OBV }
 CHANGE \gii=gagweji=adaawe-d
 REL\PST=try=sell-3.SBJ.CONJ

‘That’s when he tried to sell $\left\{ \begin{array}{l} \textit{you} \\ \textit{me} \\ \textit{him} \end{array} \right\}$.’

Similarly, first and second person pronominals cannot appear as relative root complements. The corresponding grammatical forms involve either

clausal paraphrases (12b) or obligatory advancements (12c) which we will discuss in Section 6.

- (12) a. * *Giin da-zhaa.*
 [giin]_{RRC} da=[iN]_{RR-yaa-w}
 [2]_{RRC} FUT=[to]_{RR-go-3.SBJ}
 ‘He will come to you.’
- b. *Eyaayin da-zhaa.*
 [CHANGE\ayaa-yin]_{RRC} da=[iN]_{RR-yaa-w}
 [REL\be.at-2.SG.SBJ.CONJ]_{RRC} FUT=[to]_{RR-go-3.SBJ}
 ‘He will come to you.’ (lit. ‘... where you are’)
- c. *Gga-zhaamig.*
 gi-ga=[iN]_{RR-yaa-m-igo}
 2.SBJ-FUT=[to]_{RR-go-TR-INVERSE}
 ‘He will come to you.’

Neither first nor second person pronouns ever appear as obliques. If a first or second person is the goal or source or general location of an action, a specific body part or a description of the location must be given as shown in (13a). Sentence (13b) is so thoroughly ungrammatical, that it is hard to figure out what the right morphology should be.

- (13) a. *Bneshiinh dash gii-boonii maa ndinmaangning.*
 bineshiinh dash gii=boonii-w [imaa ni-dinimaangan-ing]_{OBL}
 bird PTCL PST=alight-3.SBJ [there 1.POSS-shoulder-LOC]_{OBL}
 ‘Then a bird landed on my shoulder.’
- b. * *Bneshiinh dash gii-boonii maa niin(ing).*
 bineshiinh dash gii=boonii-w [imaa niin(-ing)]_{OBL}
 bird PTCL PST=alight-3.SBJ [there 1(-LOC)]_{OBL}
 ‘Then a bird landed on me.’

First and second person can, of course, be subjects and primary objects. Because of limitations of time and space, I will exemplify this only for primary objects, as in (14).

- (14) a. Second person primary object
 Mii wi pii gaa-ndomnaan.
 mii iwi apii CHANGE\gii=nandom-in-aanh
 CLEFT that.INAN time REL\PST=call-2.P.OBJ-1.SG.SBJ.CONJ
 ‘That’s when I called you.’

- b. First person primary object
Ggii-baashkzwimi na wii gii?
gi-gii=baashkizw-i-min na wiin giin
 2.SBJ-PST=shoot-1.P.OBJ-1.PL ques ptcl 2
 ‘Were you shooting at **us excl.**?’
- c. Third person primary object
Nzhiingenmaa oosan.
ni-zhiingenim-aa o-oos-an
 1.SBJ-hate-3.AN.P.OBJ 3.POSS-father-OBV
 ‘I dislike his father.’

In summary the facts regarding the distribution of first and second person augment Table 1 on page 309 as in Table 2:

Table 2.

	Primary		Secondary	RRC	Oblique
	Subject	Object	Object		
Full verb agreement	+	+	-	-	-
Allow 1 / 2 person pronominals	+	+	-	-	-
Gender / number agreement	+	+	+	-	-

4 Obviation

The Algonquian language family is famous for having a marker of disjoint reference, the obviative. The syntax of obviation is complex. It has been widely discussed, including by the present author (Rhodes 1991, in particular). In single clauses there can only be one non-obviative nominal among third person core arguments. This is exemplified for all possibilities in (15).

- (15) a. Subject and primary object
Wgii-waabmaan niwi semaan.
o-gii=waabam-aa_i-an aniwi asemaa_i-an
 3.ERG-PST=see-3.AN.P.OBJ_i-OBV **that.OBV** tobacco_i-OBV
 ‘He saw the tobacco.’

b. Subject and secondary object

Wgii-daawenan niwi semaan.

o-gii=adaawe-ni-an aniwi asemaa_i-an

3.ERG-PST=sell-OB_j-OBV that.OBV tobacco_i-OBV

‘He sold the tobacco.’

c. Primary and secondary object

Ngii-miinaag semaan giwi kiwenziinyag.

ni-gii=miin-aa_i-ag asemaa_j-an agiwi kiwenziiny_i-ag

1.SBJ-PST=give-3.AN.P.OB_j-PL tobacco_j-OBV those.AN old.man_i-PL

‘I gave the old men tobacco.’

d. Subject, primary, and secondary object

Wgii-miinaan semaan niwi kiwenziinyan.

o_i-gii=miin-aa_j-an asemaa_k-an aniwi

3.ERG_i-PST=give-3.AN.P.OB_j-OBV tobacco_k-OBV those.OBV

kiwenziiny_j-an

old.man_j-OBV

‘He gave the old men tobacco.’

This principle does not, however, apply in the case either of relative root complements, as shown in (16) (= example (9) on page 308) or of obliques, as shown in (17) (= example (13a) on page 310).

(16) *Niniing-sh go naa wgii-naabmaan niwi mnidoon.*

[aniniw-ing]_{RRC}=sh go naa o_i-gii=[iN]_{RR}-aabam-aa_j-an

[man_i-LOC]_{RRC}=PTCL PTCL PTCL 3.ERG_i-PST=[like]_{RR}-see-3.AN.P.OB_j-OBV

niwi manidoo_j-an.

this.OBV spirit_j-OBV

‘He saw the spirit in the form of a man.’ (2R.3.20)

**Niniining-sh go naa ...*

*[*aniniw-ini-ing*]_{RRC}=sh go naa

*[man-OBV-LOC]_{RRC} PTCL PTCL PTCL

(17) *Bneshiinh dash gii-boonii maa ndinmaangning.*

bineshiinh dash gii=boonii-w [imaa ni-dinimaangan-ing]_{OBL}

bird PTCL PST=alight-3.SBJ [there 1.POSS-shoulder-LOC]_{OBL}

‘Then a bird landed on my shoulder.’

*...*maa ndinmaangnining*

*[*imaa ni-dinimaangan-ini-ing*]_{OBL}

*[*there* 1.POSS-shoulder-OBV-LOC]_{OBL}

Adding the facts regarding obviation to Table 2 gives us Table 3.

Table 3.

	Subject	Primary Object	Secondary Object	RRC	Oblique
Full verb agreement	+	+	-	-	-
Allow 1 / 2 person pronominals	+	+	-	-	-
Gender / number agreement	+	+	+	-	-
In the scope of obviation	+	+	+	-	-

5 Null definite readings

The phenomenon of null definite readings is well known for pro-drop and head-marking languages. If only the agreement marker is present in a clause, with no external coreferential nominal, the reading is definite. Examples for Ojibwe subjects and both types of objects are given in (18).

- (18) a. Subject of intransitive
Mii dash gii-bmi-dkajid.
Mii dash gii=bimi=dakaji-d
 CLEFT PTCL PST=along=catch.cold-3.SBJ.CONJ.
 ‘And **he** caught cold (during his travels).’ (Bl T19.2)
- b. Subject of transitive
Mii dash gii-nkwetwid.
Mii dash gii=nakwetaw-i-d
 CLEFT PTCL PST=answer-1.P.OBJ-3.SBJ.CONJ.
 ‘And **she** answered me.’ (Bl T8.36)
- c. Primary object
Aapji dash go nzaagtoon.
aapji dash go ni-gii=zaagit-oo-n
 very PTCL PTCL 1.SBJ-PST=hold.dear-INAN.P.OBJ-OBJ.
 ‘I’m very stingy with **it**.’ (Bl T15.12)

d. Secondary object

Nga-zhaagooskmaagenan, ...

ni-ga=zhaagooskamaage-nan

1.SBJ-FUT=win.a.competition.(for)-AN.OBJ

‘I’m gonna be the first one to **her**, ...’ (1R1.12)

Verbs containing relative roots also have the property of null definite reference. If there is no explicit relative root complement, then the reading is one of definite pronominal / deictic reference.⁶ An example is given in (19).

(19) Null relative root complement

Baamaa go ga-zhaami.

baamaa go [Ø]_{RRC} gi-ga=[iN]_{RR}-yaa-min

afterwards PTCL [**there**]_{RRC} 2.SBJ-FUT=[to]_{RR}-go-1.PL

‘Afterwards we’ll *inc* go **there**.’ (B1 S45)

Just to be clear, this is not a property shared by any type of oblique in these languages, even those that are licensed by a verb when that verb does not contain a relative root. Obliques in Algonquian are of two syntactic types. One type is part of the syntactic frame of particular verbs, the other type expresses part of the general semantic properties of states and events and is independent of the verb of the clause. Examples contrasting a frame oblique with a general oblique are given in (20).

(20) a. frame locative

Jiimaaning ngii-booosmi miinwaa bezhig kwe, ...

[jiimaan-ing]_{OBL} ni-gii=boozi-min miinawaa bezhig kwe

[boat-LOC]_{OBL} 1.SBJ-PST=embark-1.PL and one woman

‘She and I and another woman got into **a canoe**, ...’ (B1 T5.2)

b. general locative (= [8a])

..., gii-bbaa-yaamwaaad iw bgoji-mnoomin widi waabshkokiing.

gii=bibaa=ayaa-mo-waa-d iw

PST=around=have-INAN.OBJ_i-3.PL-3.SBJ.CONJ_j that.INAN

bagoji=manoomin [iwidi waabishkoki-w-ing]_{OBL}

wild=rice_i [there swamp-LOC]_{OBL}

‘..., and they went around getting wild rice **in the swamp**.’ (B1 T5.2)

Of these two kinds of obliques, those that arise out of general semantic considerations, like those in (19), are not comparable to relative root complements, in that they are related to clauses as a whole and not directly to verbs.

But those obliques that are in the frame of a verb are generally held to be optionally licensed by that verb as in (20a). Such cases are comparable to relative root complements. They differ from relative root complements, however, in that they do not show understood definite reference, even optionally. For example, *boozid* means ‘get in (a vehicle), embark’ and can license an oblique locative specifying the vehicle, as in (20a), but it is not a relative root. When *boozid* is used without explicit mention of the vehicle, the vehicle cannot receive an understood definite reading, as shown in (21).

- (21) No explicit locative with non-relative root

Mii dash ge nii gii-boozyaan.

Mii dash ge niin gii=booz-i-yaan

PTCL PTCL also 1 PST=embark-1.SG.SBJ.CONJ

(a) ‘Then I, too, boarded.’⁷ (Bl T8.9)

(b) *‘Then I, too, boarded **it**.’

That the vehicle implicit in the meaning of *boozid* is not syntactically available can also be shown by the fact that it passes tests for anaphoric islands. Thus the sentence in (22) can only have the temporal (b) reading and not the relative clause reading (a).

- (22) *Gii-booz-i, gaa-boozyaanh.*

gii=booz-i-w CHANGE\gii=booz-i-yaanh

PST=embark-3.SBJ REL\PST=embark-1.SG.SBJ.CONJ

(a) *‘He got in the boat / car / train / plane that I got in.’

(b) ‘He got in the boat / car / train / plane after I had gotten in.’

These facts hold for all verbs that license optional obliques, in contrast to verbs containing relative roots. Returning to our tables, this set of facts distinguishes relative root complements from obliques (Table 4).

Table 4.

	Primary		Secondary	RRC	Oblique
	Subject	Object	Object		
Full verb agreement	+	+	-	-	-
Allow 1 / 2 person pronominals	+	+	-	-	-
Gender / number agreement	+	+	+	-	-
In the scope of obviation	+	+	+	-	-
Null definite reading	+	+	+	+	-

6 Change of grammatical relations

Relative root complements share with the core grammatical relations the property that they can be the target of a change in grammatical relations. In Ottawa relative clauses cannot be formed on oblique locatives and instrumentals directly. The oblique must be advanced to relative root complement. Examples are given in (23). The relative root that is used in the formation of locatives is *daN-* ‘at’, the one that is used with instruments is *ond-* ‘from’.

(23) a. Locative head of relative clause

... [ADV [REL CL *waa-dzhi-mkadekewaad*] *wiigwaaming*]
 CHANGE\wii=[**dazhi**]_{RR}=makadeke-waa-d *wiigiwaam-ing*
 REL\FUT=[**at**]_{RR}=fast-3.PL-3.SBJ.CONJ house-LOC
 ‘... in the shelter where they would fast.’ (B1 T23.3)
 (cf. *Wii-mkadewag wiigwaaming*.⁸ ‘They will fast in the house.’)

b. Instrumental head of relative clause

... [NP *iw zhoon’yaa* [RELCL *waa-wnji-mno-yaawaad*]]
iw zhooniyaa CHANGE\wii=[**onji**]_{RR}=mino=ayaa-waa-d
 that money REL\FUT=[**from**]_{RR}=good=be.at-3.PL-3.SBJ.CONJ
 ‘... the money with which they can live comfortably’ (B1 T15.1)
 (cf. *Wii-mno-yaawag iw zhoon’yaa*. ‘They will live comfortably with that money.’)

For completeness’ sake I will also show that subjects and both types of objects can be the targets of change in grammatical relations.

6.1 Passive

Advancement to subject is passive. An example is given in (24).

(24) Passive

Gdakiimnaa gwii-mkamgoomi, ...
gid-aki-im-inaan gi-wii=makam-igoo-min
 2.POSS-land-POSS-1.PL 2.SBJ-FUT=rob-PASS-1.PL
 ‘We will be robbed of our land, ...’ (2R1.14)

6.2 Advancement to primary object

Primary objects can also be targets of advancement in Ojibwe. There are plenty of instances of surface primary objects that are notional obliques appearing with verbs that contain an applicative suffix, e. g. benefactives or recipients; two examples are given in (25).

(25) a. Benefactive

Aanii-sh mii sa gii-gwaabhamwaad, ...

Aanii=sh mii sa

PTCL=PTCL CLEFT PTCL

gii=gwaabah-amaw-aa-d

PST=scoop.with.instrument-APPL-3.AN.P.OBJ-3.SBJ.CONJ

‘So then he scooped some of it out for him, ...’ (1R3.19)

b. Recipient

... ninmawaad, ...

inin-amaw-aa-d

hand-APPL-3.AN.P.OBJ-3.SBJ.CONJ

‘... [and] handed it to him, ...’ (1R3.19)

Relative root complements can show up as primary objects as well. If the relative root complement is notionally animate⁹, there is a decided preference for it to appear either as a primary or secondary object. For example, the verb *zhaad* ‘go to a certain place’ licenses a relative root complement. If the location is inanimate as in (26), a garden-variety relative root construction appears.

(26) *Gii-zhaa shpimsagong.*

gii=[iN]_{RR}-yaa-w [ishpimisagw-ing]_{RRC}

PST=[to]_{RR}-go-3.SBJ [upstairs-LOC]_{RRC}

‘He went upstairs.’ (B1 S54)

But when the goal of the motion is notionally animate (either literally or by metonymy), then the goal appears as the primary object, as in (27).

(27) *Zaasbaakdokenjin gii-zhaamaad.*

change\ziisabaakodoke-ni-d-in gii=[iN]_{RR}-yaa-m-aa-d

REL\make.sugar-OBV-3.SBJ.CONJ-OBV PST=[to]_{RR}-go-TR-3.AN.P.OBJ-3.SBJ.CONJ

‘And he went to the sugarers.’ (1R 4.50)

6.3 Advancement to secondary object

Secondary objects can likewise be targets of advancement in Ojibwe. In Southwest Ojibwe the advancement of the animate relative root complement of *izhaad* is to secondary object rather than to primary object as in Ottawa. The argument is parallel to that just given. The key evidence is given in (28).¹⁰ Factoring these data into our summary table yields Table 5.

- (28) *Wemitigoozhii ayaakozi-d nindizhaanan.*
wemitigoozhii CHANGE\akozi-d *nind-[iN]_{RR}-yaa-nan*
 Frenchman REL\be.sick-3.SBJ.CONJ 1.SBJ-[to]_{RR}-go-AN.OBJ
 ‘I went to a sick Frenchman.’ (Baraga 1878: 251)

Table 5.

	Primary Subject	Primary Object	Secondary Object	RRC	Oblique
Full verb agreement	+	+	-	-	-
Allow 1 / 2 person pronominals	+	+	-	-	-
Gender / number agreement	+	+	+	-	-
In the scope of obviation	+	+	+	-	-
Null definite reading	+	+	+	+	-
Target of advancement	+	+	+	+	-

7 Accessibility for relative clause formation.

Every grammatical relation in Ojibwe is freely available for relative clause formation except for obliques. As discussed above in Section 6, locative and instrumental obliques must be advanced to relative root complement to be relativized. Only temporal obliques may relativize directly. Examples of each type of relative are given in (29).

- (29) a. Subject relatives
 i. *Gaa [NP waya [REL CL ekwaabid]] nwaabmaasii.*
 ... waya CHANGE\akawaabi-d ...
 ... someone REL\watch-3.SBJ.CONJ
 ‘I didn’t see anyone watching’ (Bl T8:33)

- ii. ... [_{NP} *aw gchi-gwaabgesi* [_{CL} *gegaa go naa gaa-bkobiidaabaazhid*]]
aw gichi=gwaabgesiw gegaa go naa
that.AN big=dogfish almost PTCL PTCL
change-gii=bakobiidaabaaN-i-d
REL\PST=drag.into.water-1.P.OBJ-3.SBJ.CONJ
 ‘... a big dogfish that nearly dragged me into the water’ (BI T7:3)

b. Primary object relative

- ... [_{NP} *kina* ... *giw* [_{REL CL} *waajnokiiimgig*]]
akina agiw CHANGE\wiji=anokii-m-Ø-ag-ig
all those.AN REL\with=work-TR-3.AN.P.OBJ-1.SBJ.CONJ-3.PL
 ‘... all of those I work with ...’ (BI T11:6)

c. Secondary object relative

- Eshkam znagad wii-deb naming* [_{NP} *iw* [_{REL CL} *waa-boodweng*]].
 ... *iw CHANGE\wii=boodwe-m-g*
 ... *that.INAN REL\FUT=burn-INDF-3.SBJ.CONJ*
 ‘It’s getting increasingly difficult to get firewood.’ (BI T15:15)
 [lit. that which one burns (as fuel)]

d. Relative root complement relative

- Gye go mii gii-miin’gooyaang* [_{NP} *rooms* [_{REL CL} *waa-dnizyaang*]].
 ... *rooms CHANGE\wii=[daN]_{RR}-izi-yaang*
rooms REL\FUT=[at]_{RR}-be-1EXCL.PL.SBJ.CONJ
 ‘And we were then given rooms to stay in.’ (BI T8:14)

e. Oblique relative (temporal only)

- ... [_{ADV} *pii* [_{REL CL} *nebaanid*]] ...
 ... *apii CHANGE\nibaa-ni-d*
time REL\sleep-OBV-3.SBJ.CONJ
 ‘... when she fell asleep ...’ (1R2:30)
 [lit. at the time at which she fell asleep]

Adding this last type of data to our summary of the syntax of Ojibwe grammatical relations gives us Table 6 on the following page. From this table it is easy to see that relative root complement is a distinct type of grammatical relation.

Table 6.

	Subject	Primary Object	Secondary Object	RRC	Temporal Oblique	Other Obliques
Full verb agreement	+	+	-	-	-	-
Allow 1 / 2 person pronominals	+	+	-	-	-	-
Gender / number agreement	+	+	+	-	-	-
In the scope of obviation	+	+	+	-	-	-
Null definite reading	+	+	+	+	-	
Target of advancement	+	+	+	+	-	
Accessible for relative clause formation	+	+	+	+	+	-

8 Conclusion

The extensive literature on grammatical relations developed by proponents of Relational Grammar consistently shows only three core relations, subject and two types of objects (or 1s, 2s, and 3s), opposed to obliques. The key distinction between core relations and obliques is that obliques can never be targets of advancement. What we have shown is that in Ojibwe there is a grammatical relation intermediate between the core relations and obliques. It splits the key properties of core relations and obliqueness, but it has the crucial property that it can be the target of advancement. There have been other suggestions in the Relational Grammar literature for grammatical relations intermediate between core relations and obliques (in particular, Postal 1990). However, unlike the Algonquian construction, the data behind such proposals are marginal and not very robust. By contrast, the Algonquian relative root construction is extremely common, much more common in running text than the widely discussed inversion construction, and the data are very robust. The relative root construction is family-wide in Algonquian. The specific details are slightly different in different Algonquian languages, but the basic syntax is identical.

Finally, there remains the question of whether the relative root construction is actually a *rarum*. Since no similar construction has been reported, the possibility exists that there might be other languages with parallel constructions that simply haven't been reported. However, there is good evidence that

such is not the case. In Rhodes (2005) I argued that the historical source of the relative root construction is the incorporation of postpositions into the verb. Many otherwise typologically and genetically diverse languages, including some Indo-European languages, have similar historically incorporated adpositional material in verbs. See for example, Craig and Hale (1988), who present evidence from three Western Hemisphere language families, one in North America, one in Central America, and one in South America. The difference between Algonquian languages and the languages in other families showing adpositional incorporation is that in the non-Algonquian languages the stranded NPs either become objects, as shown in (30), or they remain as obliques, as shown in (31).

- (30) Winnebago ([17b] in Craig and Hale 1988: 327)

fiá-nan-nín-zhín-je-enán.
 on-by.foot-2.OBJ-stand-AUX-DECL
 'It is standing on you.'

- (31) Homeric Greek (from [25] in Garrett 1990: 196)

stómat-os d eks-éptusen hálmēn
 mouth-GEN PTCL out-spit.AOR.3.SG seawater.ACC
 'He spat seawater from his mouth.' (Odyssey 5.322)

pân d hup-ethermánthē ksíphos haímat-i
 all.NOM ptcl under-heat.AOR.PASS.3.SG sword.NOM blood-DAT
 'The whole sword was heated under the blood.' (Illiad 16.333)

The fact that in Algonquian the NPs associated with relative roots are neither reanalyzed as objects nor are left as stranded obliques, but are instead restructured as a distinct grammatical relation clearly qualifies the Algonquian relative root construction as a *rarum*.

Abbreviations

The abbreviations used in glossing generally follow the Leipzig Glossing conventions. For example, 1.SBJ, 2.SBJ, 3.SBJ, and 3.ERG represent first, second, third person subject, and third person subject of transitive (i.e., ergative) respectively. Deviations from the Leipzig conventions are the following:

AN = animate; AOR = aorist; CHANGE = a process morpheme; CLEFT = marker of clefting; CONJ = conjunct; INAN = inanimate; MODAL = modal; NP = noun phrase; OBV = obviative; P.OBJ = primary object; PTCL = discourse or emphatic particle; QUES = yes-no question particle; RR = relative root; RRC = relative root complement

Notes

1. The apparatus for citing examples from published sources is as follows: The sources are: 1R = Kaye, Piggot, and Tokaichi (1971), 2R = Piggot and Kaye (1973), B1 = Bloomfield (1958). Bloomfield contains both texts and sentences; these are distinguished using 'T' and 'S'. The next number is the number of the text or example sentence. If the citation is from a text, the number of the sentence following the punctuation of the published version is given preceded by a period. In the case of 1R and 2R, I have cited the taped version if there is a difference between it and the published version. So (B1 T8.20) means Bloomfield (1958) text number 8, sentence 20. All sentences not cited are from my personal field notes.
2. Most of the data in this paper is drawn from the Ottawa dialect of Ojibwe. While the details are different, all the dialects of Ojibwe, and, in fact, all Algonquian languages show analogous syntactic phenomena.
3. In underlying forms the *n* that alternates with *zh* in palatalizing environments is indicated by *N*. Similarly the *s* that palatalizes to *sh* is indicated by *S*.
4. These morphemes are called *relative roots* for reasons of terminological congruence in Algonquian studies. The prototypical Algonquian stem has two parts, an initial and a final. Examples of both a noun and a verb are given in (32).

- (32) a. (a) *ndomaad*
 [nandaw-im]-aa-d
 [seek-speak]-3.AN.P.OBJ-3.SBJ.CONJ
 [INITIAL-FINAL]
 'call' (citation form)
- b. *mskomin*
 miskw-imin
 red-berry
 [INITIAL-FINAL]
 'raspberry'

The class of initials that are monomorphemic are called *roots*. The term *relative* is also an Algonquianist term, referring to the fact that these morphemes regularly license an adjunct. To be clear, it has nothing to do with relative clauses. More recent Americanist studies have called analogous morphemes in other language families *relational preverbs* (Craig and Hale 1988).

5. An earlier treatment of the relative root construction and its implications for transitivity can be found in Rhodes (2005).
6. Lest some Algonquianist take exception on this point, let me point out a wrinkle in this general property of relative roots. In his discussion of relative roots in Ottawa, Bloomfield (1958) noted that there are some verbs containing relative roots which seem to lack the property of requiring a complement.

"... some words with relative roots occur in special meanings requiring no [relative root complement]: [*dzhimaad*] 'he talks about him', root [*daN-*], requiring no [relative root complement]." (Bloomfield 1958: §20.5, 130)

However, Bloomfield only understood half the story. The crucial property of these words is that they *cannot* take a relative root complement. Rather, they are all idioms and

cannot have their pronominal (which happens to be a zero) replaced by an NP; this is a crucial part of their idiomaticity. In the case of *dzhimaad*, the word that Bloomfield cites, it is decidedly negative in implication, as shown in (33).

- (33) *dzhimaad* 'talk about s.o., esp. talk about s.o. behind their back, say something bad about s.o., gossip about s.o.', lit. 'talk about him there', cf. neutral *dbaajmaad* 'talk about s.o., esp. tell a story about s.o., give a report about s.o.'

stem: *daN- im-*

at- act.with.speech.(on)-

In being idiomatic and having what amounts to pronominals as part of their idiomaticity, Ojibwe words like *dzhimaad* are analogous to idioms in better known languages which may also have pronominals as a crucial part of their idiomaticity, e. g. English *run for it*, *Dig it!*, and the like. The point here is that these idiomatic uses of relative roots are the exceptions that prove the rule that relative roots have obligatory complements.

7. In Bloomfield's translation this sentence is rendered "So then I, too, got on board the train." But 'train' is actually only inferable. It's an inaccurate translation.
8. Not all contemporary Ottawa speakers like sentences like this. The bare instrumental construction is being replaced by clausal adjuncts.
9. All the Ojibwe languages (and to the best of my knowledge all the Algonquian languages) syntactically distinguish notional animacy from grammatical animacy. For example, grammatically animate but notionally inanimate referents cannot function as possessors or as the subjects of direct clauses.
10. Although one might object that the evidence for advancement of a relative root complement to secondary object in Southwestern Ojibwe is not relevant to Ottawa, there is also a pan-Ojibwe construction that can be argued to have obligatory advancement from relative root complement to secondary object. Outside of relative roots, no roots in Ottawa affect the argument structure of the verbs they occur in with two exceptions, the root *wiid-* 'with' and the root, *ap-* '(up)on'. Verbs containing *wiid-* take primary objects. Verbs containing *ap-* take secondary objects. Some examples are listed in (34).

- (34) *pabid* 'sit on' (expected transitive: **patood*)
penmod 'have faith in' (expected transitive: **pendang*)
pikweshmod 'lay one's head on' (expected transitive: **pikwesdood*)
 (It is worth noting that the form *patood* is accidentally homophonous with a good form meaning 'run to a certain place, run along a certain path', which takes a relative root complement. The pairing of syntax and meaning is enough to make it certain that the two would be distinct verbs.)

This syntax of the forms in (34) is entirely due to the initial *ap-* '(up)on'. The parts of these verbs that set the transitivity all have regular ways to form transitives (listed in the parentheses) which cannot be used in these cases. Since *ap-* '(up)on' is the only non-relative root in modern Ottawa that licenses an argument slot in the verb constructions in this way, and in view of its relative root-like meaning, and its consistency with relative root morphosyntax, I argue that it is a relative root that requires its complement to advance to secondary object. The *ap-* construction is old in Algonquian. Cognates of

Ojibwe words containing *ap-* exist in other Algonquian languages. In these languages the morphology and syntax of *ap-* forms is as quirky as it is in Ojibwe and in the same ways.

References

- Baraga, Frederic I.
1878 *A theoretical and practical grammar of the Ojibwe language*. Montreal: Beauchemin and Valois.
- Bloomfield, Leonard
1958 *Eastern Ojibwa: grammar, texts, and word list*. C. Hockett (ed.) Ann Arbor: University of Michigan Press.
- Craig, Colette G. and Kenneth L. Hale
1988 Relational Preverbs in Some Languages of the Americas: Typological and Historical Perspectives. *Language* 64.2: 312–344.
- Dryer, Matthew S.
1986 Primary and secondary objects, and antitativity. *Language* 62: 808–845.
- Garrett, Andrew
1990 Applicatives and Preposition Incorporation. In: *Grammatical Relations: A Cross Theoretical Perspective* K. Dziwirek, P. Farrell, and E. Mejías-Bikandi (eds.), 183–198. Palo Alto: CSLI Stanford.
- Kaye, Jonathan D., Glynn L. Piggott and Kensuke Tokaichi
1971 *Odawa Language Project: First Report*. Toronto: Department of Anthropology, University of Toronto.
- Piggott, Glynn L. and Jonathan D. Kaye
1973 *Odawa Language Project: Second Report*. Toronto: Centre for Linguistic Studies, University of Toronto.
- Postal, Paul M.
1990 French Indirect Object Demotion. In: *Studies in Relational Grammar 3*. P. Postal and B. Joseph (eds.), 104–200. Chicago: University of Chicago Press.
- Rhodes, Richard A.
1990 Ojibwa Secondary Objects. In: *Grammatical Relations: A Cross Theoretical Perspective* K. Dziwirek, P. Farrell, and E. Mejías-Bikandi (eds.), 401–414. Palo Alto: CSLI Stanford.
- Rhodes, Richard A.
1991 Obviation, Inversion, and Topic Rank in Ojibwa. *Berkeley Linguistic Society Parasession to the 16th Annual Meeting*, D. Costa (ed.), 101–115. Berkeley: BLS.
- Rhodes, Richard A.
2005 *Clause structure, core arguments, and the Algonquian relative root construction*. The 1998 Belcourt Lecture. Winnipeg: Voices of Rupert's Land.

Mawng lexicalised agreement in typological perspective

Ruth Singer

1 Introduction

1.1 What is lexicalised agreement?

Lexicalised agreement is the lexicalisation of otherwise productive verbal agreement morphology. Although relatively undescribed in the typological literature, it is found in a number of unrelated languages of the Australian, North American and Asian continents. Lexicalised agreement is found most often in languages with pronominal verbal agreement for subject and object. In certain expressions, the agreement features are fixed for one or both arguments. In fact, the lexicalised agreement index does not correspond to a real argument but is a dummy argument. An example of such an expression in the Australian language Mawng is shown in (1).

- (1) *Ngani-ngartpanpu-n*
3MA/1SG-have.headache-NP
'I have a bad headache.' (Hewett et al. 1990)

Note that the gloss given for the verb in (1) shows the features of the subject followed by the features of the object after a forward slash. The pronominal prefix is treated as a portmanteau morpheme so the order of subject and object glosses do not reflect ordering of subject and object morphemes. The expression in (1) has lexicalised transitive subject agreement. The verb appears to index a third person Masculine gender subject but this argument cannot be expressed by an NP. Neither can it be interpreted as coreferent with any participant available in the context of use so it is clearly a dummy argument. It is only third person agreement affixes that are ever lexicalised. The lexical entry of expressions with lexicalised agreement consist of a verb root plus the features of the lexicalised agreement index.

Verbs with dummy transitive subjects such as that in (1) have been described as 'experiencer object' verbs and are found throughout the Iwaidjan language family, which includes Mawng (Evans 2004). However, in Mawng

any type of core argument index can be lexicalised. An example of a Mawng verb with lexicalised object agreement is shown in (2).

- (2) *Ngarrung-marrajpu-ng kapa ara Anjumu*
 1PLEX/3LL-walk-PC there all.the.way.to place.name
 ‘We walked over to Anjumu.’ HH Text 1¹

1.2 How rare is it?

Lexicalised agreement has been reported in these language groups:

- non-Pama Nyungan languages of the Top End, Australia
 - Mawng, Iwaidja (Iwaidjan)
 - Gaagudju (family level isolate)
 - Tiwi (family level isolate)
- Kiowa-Tanoan language family
 - Kiowa
 - Southern Tiwa
- Athabaskan languages
 - Navajo
 - Tolowa
- Ket (Yeniseic)

Lexicalised agreement has been treated as unique in descriptions of individual languages but in fact, it is not restricted to a single language, language family or linguistic area. There has been little typological work on the phenomenon which only affects a small proportion of the verbal lexicon in any one language.² Earlier discussions of lexicalised agreement are found within the description of a single language and erroneously describe it as unique to that language. This paper is the first step towards a typology of lexicalised agreement. The next section discusses lexicalised agreement in Mawng, the language for which the best description of lexicalised agreement is available (Singer 2006).

Lexicalised agreement is defined as the lexicalisation of *otherwise productive* agreement morphology. Thus as a typological category it excludes lexicalised agreement morphology which is no longer used productively. This

is why absolutive Iwaidja pseudo-arguments are not analysed as lexicalised agreement, as discussed in Section 3. In Sections 4 and 5, lexicalised agreement in Gaagudju and Southern Tiwa is discussed. Finally in Section 6, I suggest what features of a language favour the development of extensive lexicalised agreement.

2 Mawng lexicalised agreement

2.1 Typological background

Mawng is a non-Pama Nyungan Australian language spoken as a first language by around three hundred people, in north-west Arnhem Land, Northern Territory. It is a member of the Iwaidjan language family, sister language to Iwaidja and is still being acquired by children. The location of Mawng in relation to other languages of north-west Arnhem land is shown in Figure 1. Mawng neighbours Iwaidjan languages to the west and Gunwinyguan languages to the south and east.

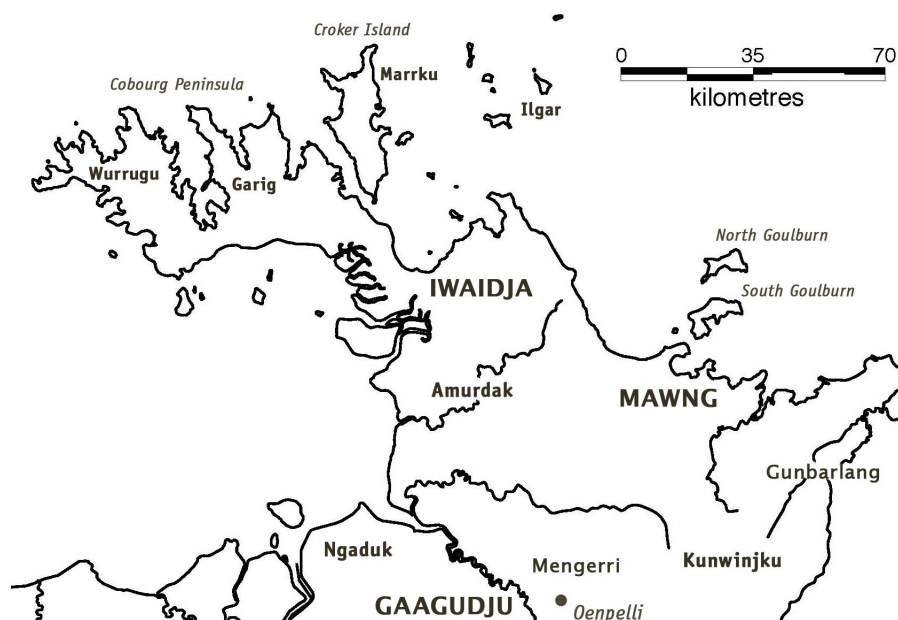


Figure 1. Location of Mawng and neighbouring languages

Table 1. Mawng genders

MA	Masculine	male humans, most animals
FE	Feminine	female humans, crabs, some birds, no body parts
LL	Land	land, landscape features, liquids, stories, customs, knowledge
VE	Vegetation	most plants, non-edible parts, plant products, wood, fire
ED	Edible	all edible plant products, buildings and household items

Mawng has five genders which have a strong semantic basis but gender assignment is by no means predictable. The five genders are listed in Table 1 with brief descriptions of their semantic foci.

Mawng gender is encoded in a number of different ways. Within the NP gender is encoded by articles, demonstratives, free pronouns, the prefixes of a small number of adjectival nominals and body part nouns. Gender is also encoded in verbal pronominal prefixes for third persons. Since these are the prefixes that may become lexicalised, the existence of five genders means that there is a wide range of possible lexicalisation patterns.

Examples of canonical verbs (i. e. verbs without lexicalised agreement) are shown below. The verb in (3) is intransitive while the verb in (4) is transitive although they share the same verb root.

- (3) *K-i-langali-∅*
PR-3MA-stand-NP
'He stands.'
- (4) *K-ini-langali-∅*
PR-3MA/3MA-stand-NP
'He stood it (MA) up.'

Gender agreement is restricted for transitive subject prefixes — there is only a two-way distinction between Masculine and Non-masculine. Non-masculine forms are glossed 'GEN'. Note that only a small proportion of Mawng verb roots are labile like the verb root *-langali* used in (3) and (4).

In addition to simple verbs like those in (3) and (4), Mawng has a type of complex verb found in a number of northern Australian languages that will be referred to as a 'coverb construction' following other authors (e. g. Schultze-Berndt 2000). For example, in (5) the coverb *pirl* forms a coverb construction with the inflecting verb *-wu* 'hit'.

- (5) *Ati-wu-k pirl ta kantijawa.*
 3MA/3ED-HIT-PP cut.through ED bread
 ‘He cut through the bread.’ (Hewett et al. 1990)

The coverb cannot occur on its own without an inflecting verb. In Mawng 700 verbs have been recorded and 300 coverbs out of a total dictionary of 3,500. In tables in this paper, coverb constructions will be listed by giving the inflecting verb first and then the coverb, although this is the opposite order to that in which they appear in speech.

Word order in Mawng is relatively free. Overt NPs are not usually required unless the referent is new and not recoverable from the context. Noun incorporation is not found in Iwaidjan languages although it is found in nearby Gunwinyguan languages such as Bininj Gun-wok and Nunggubuyu (Evans 2003; Heath 1984).

2.2 Classes of verbs with lexicalised agreement

Approximately 200 Mawng verbs with lexicalised agreement have been recorded — roughly one quarter of all verbs. Many verbs with lexicalised agreement fall into verb classes delineated by semantic and syntactic properties. The verb *-marrajpu* (LL O) ‘walk’ illustrated in (2) earlier is part of a class of motion verbs with lexicalised object agreement. These have either Masculine or Land gender lexicalised object agreement. Some examples are shown in Table 2. However, the most common motion verbs are canonical intransitive verbs as we expect. These include the two ‘go’ verbs *-a* ‘go1’ and *-arra* ‘go2’ and *-ngurri* ‘run, move fast’.

The verb *-ngartpanpu* (MA A) ‘have headache’ illustrated in (1) earlier, is part of the group of ‘experiencer object verbs’. Their animate identifiable argument is encoded as object while their subject slot is lexicalised. The ex-

Table 2. Motion verbs with lexicalised object agreement

Predicate	Lexicalised agreement	Gloss
<i>-ma rlirr</i>	MA O	move along
<i>-ma karliwirrk</i>	MA O	flee
<i>-marrajpu</i>	LL O	walk
<i>-marranyi</i>	LL O	wave
<i>-ukpu</i>	MA O	turn around

periercer object verbs include predicates of emotion, cognition, bodily sensation and illness. Examples of this class of verbs are listed in Table 3. Experiencer object verbs are also found in other Iwaidjan languages (Evans 2004). The experiencer object verbs encode their animate identifiable argument as object while their subject is a dummy argument. The canonical synonyms of experiencer object verbs are intransitive predicates in which the experiencer is encoded as an intransitive subject such as *-maju* ‘sick, suffer’.

Communication predicates are another class of lexicalised agreement verbs which all denote communication and have lexicalised object agreement. Some examples are shown in Table 4. The meaning of all communication predicates involves communication. The speaker is encoded as subject while the hearer is encoded using an oblique pronoun that follows the verb. These verbs also subcategorise for a third argument which is the message communicated. The message argument is optionally expressed as a Land gender nom-

Table 3. Experiencer object predicates

Predicate	Lexicalised agreement	Gloss
<i>-arajpu</i>	GEN A	feel pain, ache
<i>-ngartpanpu</i>	MA A	have headache
<i>-martalkanyi</i>	MA A	sneeze
<i>-wulkparrki</i>	MA A	stomach rumble
<i>-ma arrarr</i>	MA A	have diarrhoea
<i>-aka arlarl</i>	GEN A	confused, oblivious, go astray (spiritually)
<i>-wu warryak</i>	MA A	have groin itch
<i>-ma wiliny</i>	MA A	have cramp, pins and needles, numbness
<i>-wu ngurrij</i>	GEN A	feel sleepy

Table 4. Communication predicates

Predicate	Lexicalised agreement	Gloss
<i>-lakula</i>	MA O	explain (to OBL)
<i>-ma lagarr</i>	MA O	tell lies (to OBL), bullshit, mislead OBL
<i>-ni monmon</i>	MA O	tell important information, explain (to OBL)
<i>-unma</i>	LL O	tell story (to OBL), explain (to OBL)
<i>-lakpurnangani</i>	LL O	translate (for OBL)
<i>-ma mayali</i>	VE O	explain (to OBL)

inal or a *ta* nominalised clause and may appear to be the object argument of those communication predicates with lexicalised *Land gender* object agreement. For example, in (6) the communication verb *-unma* ‘tell story’ has a message argument that takes the form of a Land gender nominalised clause (shown in square brackets).

- (6) *Nganti yirrik anny-unma-ngung nuwu [ta ani-wu-ning*
 somebody 3MA/3LL-tell.story-PC 2SG.OBL LL 3MA/3LL-HIT-PC
ngurlk ja kiyap].
 make.water.murky MA fish
 ‘Somebody was telling you that the water was stirred up by fish.’ Child
 078

The message argument of *-unma* ‘tell (story)’ is a good candidate for object because *-unma* has lexicalised Land gender object agreement but the same types of message argument can also occur with communication predicates like *-lakula* ‘explain’ which have lexicalised Masculine object agreement as shown in (7).

- (7) *Tukapa ngiwa-lakula-Ø [ta k-awunp-u-n mawawiny la*
 DEM.G.LL 1SG/3MA.F-explain-NP LL PR-3PL/3PL-give-NP child-in-law CONJ
wemin].
 3PL.CONTR
 ‘I’ll explain how they arrange marriages (lit: give each other children-in-laws)’ Mawawiny 001

The verb *-unma* (LL O) ‘tell story’ is used to mean ‘count, appraise’ with canonical agreement. The development of the lexicalised agreement verb *-unma* (LL O) ‘tell story’ from the canonical verb *-unma* ‘count’ follows a similar pattern to the development of the Germanic verbs *recount* and *erzählen*. Analogy between the canonical verb *-unma* ‘count, appraise’ and the lexicalised agreement verb *-unma* (LL O) ‘tell story’ suggests that the message argument is the object of the latter. However, many canonical verbs such as the intransitive verb *-mi* ‘say, tell’ and the transitive verb *-lalkpa* ‘let know, remind’ subcategorise for a message argument and these verbs never encode the message argument as a core argument. Tests for object status are few in Mawng and those that exist can only be used with animate referents. A uniform analysis of communication predicates and their message arguments requires that *-unma* (LL O) ‘tell story’ be analysed as a verb with lexicalised object agreement.³

Table 5. Other complement-taking predicates with lexicalised agreement

Predicate	Lexicalised agreement	Gloss
-wu larr	MA O	finish (event)
-ma rtalk	MA O	start (event)
-wurru	LL O	think, know (about event)
-miyarma	LL O	want (event)
-arnama	LL O	not want (to do event)

Many complement-taking predicates have lexicalised object agreement. These include aspectual predicates such as *-wu larr* (MA O) ‘finish’ and cognition predicates such as *-wurru* ‘think, know’. Examples are shown in Table 5. Note that the cognition predicates in Table 5 encode a real object when they have an entity object but have lexicalised object agreement when they take an event complement. We will see in Section 5 that many complement-taking predicates in Southern Tiwa also have lexicalised object agreement. Gaagudju has many predicates with lexicalised agreement that express communication.

2.3 Two other unusual types of verbs in Mawng

In addition to the types of lexicalised agreement verbs described above, a small group of Mawng verbs have what will be called ‘replacive lexicalised agreement’. In this type of lexicalised agreement the features of a particular agreement slot of a verb have become fixed but this agreement slot does correspond to a real core argument. Only around two dozen Mawng verbs with replacive lexicalised agreement have been recorded. An example of such a verb is *-warlkanyi* (VE S) ‘fall’ illustrated in (8).

- (8) *Kayirrk la ma-warlkanyi-ny yirrk-ju yanatjanat*
 then 3VE-fall-PP COMPL-only 3DU.MA
 ‘Then they both fell down together.’ Kurrana 016

Note that the expected subject of *-warlkanyi* (VE S) ‘fall’ in (8) is in fact the real syntactic subject as shown by the fact that the third person dual Masculine pronoun *yanatjanat* is used to refer to the participant. There is a mismatch between the features encoded by the agreement affix and the features of the real argument. Replacive lexicalised agreement is a simple

morphology-syntax mismatch whereas the mismatch involved in ordinary lexicalised agreement is both semantic and syntactic. Unlike the ordinary type of lexicalised agreement described earlier, replacive lexicalised agreement has not been reported in other languages and so may be unique to Mawng.

Another unusual property of the Mawng verbal lexicon is the large number of verbs which have very narrow selectional restrictions. These verbs are referred to as having ‘restricted arguments’. There are around 50 of these verbs and they may appear to have lexicalised agreement because the agreement features of one argument do not vary but in fact the motivation for this is in their semantics rather than due to lexicalisation of agreement (cf. Plank 1984). For example, the transitive verb *-aka* ‘throw’ occurs with a range of agreement patterns but when it has the sense ‘build’ it only ever has Edible gender object agreement. All built structures are Edible gender so the apparently lexicalised agreement pattern is the result of the strong semantic basis of the gender system and the selectional restrictions of the verb sense. Restricted argument verbs tend to have an implicit default argument type so that their argument is only rarely expressed as an NP but it is a real argument of the verb, unlike the dummy argument of lexicalised agreement verbs.

3 Iwaidja: experiencer object verbs and pseudo-argument verbs

Iwaidja is the sister language to Mawng and the only other surviving member of the Iwaidjan language family (Evans 2000). Iwaidja has lost the five gender system found in Mawng, although it distinguishes between masculine and feminine transitive subjects. Iwaidja has experiencer object verbs like those in Mawng, which are genuine lexicalised agreement verbs. However, it also has a set of verbs that are described by Evans (2007) as having absolutive pseudo-argument affixes. These must or can optionally occur with a vestigial gender affix that is reconstructable as a Vegetation or Land gender agreement prefix. Many of the Iwaidja pseudo-argument verbs are cognate with Mawng verbs with lexicalised agreement. However, the fact that the vestigial Vegetation and Land gender affixes are not part of the productive agreement system of Iwaidja means that such verbs cannot be analysed as lexicalised agreement verbs like their Mawng cognates (see definition of lexicalised agreement discussed earlier). Some of these cognates are shown in Table 6 on the following page.

Around 60 pseudo-argument verbs have been recorded in Iwaidja compared to 230 Mawng verbs with lexicalised agreement or restricted argu-

Table 6. Iwaidja pseudo-argument verbs cognate with Mawng lexicalised agreement verbs (source: Evans and Birch (n. d.))

Iwaidja verb	Pseudo-argument	Mawng verb	Lexicalised agreement	Gloss in both languages
<i>-marraj pang</i>	'land gender' O	<i>-marraj pu</i>	LL O	walk
<i>-unma</i>	'land gender' O	<i>-unma</i>	LL O	tell story
<i>-wurrwun</i>	'land gender' O	<i>-wurru</i>	LL O	think (event)

ments. The total number of verbs recorded in each language is similar. The existence of pseudo-argument verbs in Iwaidja shows that Mawng lexicalised agreement is not a sign of the decay of the agreement system as in fact the lexicalised agreement patterns can be preserved even when the gender agreement system itself is lost as in Iwaidja. Verbal agreement often becomes lexicalised. However, what is particular to the typological category referred to as 'lexicalised agreement verbs' here is that the same morphology can be lexicalised in some verbs but productive in others. Iwaidja experiencer object verbs are a type of lexicalised agreement verb but Iwaidja verbs with absolute pseudo-arguments are not.

4 Gaagudju

Gaagudju is a language spoken near Mawng but not adjacent to Mawng. It is a family level isolate which has a four-way gender distinction in verbal agreement. Forty verbs with lexicalised agreement are listed by Harvey (2002). These fall into similar classes to those found in Mawng. They include communication predicates and verbs of involuntary motion. In some cases the lexicalised index is a dummy argument like the noun class IV object indexed by the verb *ma=waala* 'bear (young)'. The person born is encoded as an indirect object. In other cases, Harvey (2002) argues that the lexicalised index encodes a cognate object. These types of cognate objects are not morphologically related to the form of a noun but are similar to cognate object verbs in other languages in that the meaning of the verb already specifies the nature of the object, which is often left unspecified, but can be expressed with a nominal. An example of this type of verb is *barla-bu* 'sing' which always has class III object agreement. Other such verbs are listed in Table 7 on the next page.

Harvey suggests that the verbs in Table 7 with a class IV object have *dja-gaardu* 'language, noise, word' as object. However, there is little evidence

Table 7. Gaagudju lexicalised object verbs: cognate object type

Verb	Lexicalised agreement	Gloss
<i>barla-bu</i>	III O	sing
<i>bala-bu</i>	IV O	talk
<i>bolo-boyo-ma</i>	IV O	lie, trick
<i>gala-bi</i>	IV O	call out

that we are really dealing with a cognate object rather than a dummy object apart from one example in which the class III noun *mabalaabala* ‘corroboree, song’ occurs with the verb *barla-bu* (III O) ‘sing’. These verbs present challenges to analysis similar to the Mawng communication predicates discussed in Section 2.

Lexicalised agreement verbs are also found in Tiwi, another family-level isolate spoken west of Mawng (Lee 1987; Lee 1993). This suggests that lexicalised agreement verbs are an areal feature of part of coastal North-west Arnhem land and the land to the west. However, verbs with lexicalised agreement form a much smaller proportion of the total verbal lexicon in Gaagudju and Tiwi than they do in Mawng.

5 Southern Tiwa

Lexicalised agreement is not restricted to Mawng and neighbouring languages. Around 150 verbs with lexicalised agreement have been recorded in Southern Tiwa, a member of the Kiowa-Tanoan language family, spoken by Pueblo Indians living near Albuquerque, New Mexico (USA) (Frantz 1995; Frantz n.d.). Southern Tiwa verbal agreement distinguishes three genders and encodes two arguments on the verb. Despite the lack of genetic or areal relations between Southern Tiwa and Mawng, the types of events encoded by verbs with lexicalised agreement verbs in the two languages are remarkably similar. Table 8 on the following page gives an example of each of the main classes of verbs with lexicalised agreement in Southern Tiwa.

Note that Southern Tiwa verbal agreement prefixes have suppletion for person, number and gender and so are simply labelled ‘A, B’ and ‘C’. Some experiential predicates have a dummy subject and encode the experiencer as a dative such as *-bewate* ‘get happy’. However, there are no experiencer object verbs in Southern Tiwa as only absolutive prefixes can be lexicalised. There are verbs such as *-weuri* ‘go out’ which we would expect to be encoded by

Table 8. Representative Southern Tiwa verbs with lexicalised agreement

Verb	Lexicalised agreement	Gloss	Class
<i>-bewate</i>	C S	OBL get happy	nonsubject experiencer
<i>-weuri</i>	C O	go out	motion
<i>-hwawi</i>	C O	sigh	involuntary bodily process
<i>-p'euy'am</i>	A O	start	Complement-taking predicate

an intransitive verb but instead are encoded by a transitive verb with dummy object agreement. Compare this with *-ukpu* (MA O) 'bend over' in Mawng. Other verbs encode involuntary bodily functions such as *-hwawi* 'sigh' which has dummy object agreement. As in Mawng, a number of complement-taking predicates in Southern Tiwa have lexicalised object agreement such as *-t'am* 'help OBL' and *-p'euy'am* 'start'.

Lexicalised agreement has also been reported in another Kiowa-Tanoan language — Kiowa (Watkins 1984). As in Southern Tiwa, only absolutive verbal prefixes are lexicalised in Kiowa. Lexicalised agreement has also been reported in the Athabaskan languages Navajo (Young, Morgan, and Midgette 1992) and Tolowa Athabaskan (Givón and Bommelyn 2000) as well as the Yeniseic language Ket (Vajda 2003) which Vajda (forthc.) argues is related to Athabaskan languages. These languages have much more complex verbal morphology than the Australian languages with lexicalised agreement and they also allow noun-incorporation. However, it is not clear how extensive lexicalised agreement is in any of these languages as it depends on the interest of the linguist describing the language whether it is reported at all. To give an example, there is no mention at all of lexicalised agreement in Capell and Hinch (1970) a grammar of Mawng that is otherwise thorough. It is likely that lexicalised agreement is also found elsewhere but no detailed survey of its distribution has been carried out as yet.

6 Conclusions: rare but not exotic

Lexicalised agreement has previously been treated as unique to particular languages but although rare it is clearly not unique to any one language. The distribution of lexicalised agreement suggests that verbal agreement for at least two arguments that distinguishes three or more genders favours its development. Lexicalised agreement has been found in a small set of languages and is

restricted to a small set of verbs within each language. Two phenomena mentioned earlier that are likely to be very rare or unique are Mawng replacive lexicalised agreement verbs and Iwaidjan pseudo-argument verbs.

Although lexicalised agreement verbs appear at first to be rather exotic, there are clearly links with expressions have developed through the freezing of a verb plus argument combination. These types of verb-argument idioms include lexicalised noun-incorporations in which an incorporated generic noun does not refer to an argument, and the lexicalisation of verb-argument relations in typical European noun-verb idioms (Nunberg, Sag, and Wasow 1994). In addition, there are similarities between lexicalised agreement verbs and cognate-object type verbs in Australian languages which also show a mismatch between their apparent transitivity and number of arguments (Austin 1982; Simpson 1991). There also are similarities between Mawng and Iwaidja experiencer object verbs with lexicalised agreement and experiencer object expressions in other languages. However, experiencer object expressions in other languages usually have overt nominal subjects rather than dummy subjects. There are a few examples in the literature of languages with experiencer object expressions that have a small number which take dummy subjects: in the Papuan language Kalam (Pawley et al. 2000) and the Kiranti language Limbu (van Driem 1987).

Abbreviations

1 = first person; 2 = second person; 3 = third person; A = transitive subject agreement (Mawng); fused number / gender agreement category (Southern Tiwa); C = Southern Tiwa agreement category; CONJ = conjunction (particle); COMPL = completive postverbal particle; CONTR = contrastive pronoun; DEM.G = given demonstrative pronoun; DU = dual; ED = Edible gender; EMPH2 = verbal emphatic particle; EX = exclusive (plural); F = future tense form of pronominal prefix; FE = Feminine; GEN = nonmasculine gender; III = Gaagudju gender; IV = Gaagudju gender; LL = Land gender; MA = Masculine; NP = nonpast TAM suffix; O = object agreement; OBL = oblique; PC = past continuous TAM suffix; PL = plural; PP = past punctual TAM suffix; PR = present tense prefix; S = intransitive subject agreement; SG = singular number; VE = Vegetation gender

Some notes on Mawng interlinear glosses: Capitals are used to gloss inflecting verbs that are part of a coverb construction. A slash (/) is used between the gloss of subject features and object features for transitive verb prefixes. Free pronouns are glossed simply with their person, number and gender features. Articles are glossed simply with their gender feature or number feature.

Acknowledgements

Mawng data is from my own fieldwork unless otherwise noted. Thanks to Waruwi Community Inc. for supporting the fieldwork and Nita Garidjalalug in particular. Financial support for fieldwork came from the University of Melbourne, Department of Linguistics, the Australian Research Council funded project 'Reciprocals across languages', Sonderforschungsbereich project SFB-632 D2 'Typology of information structure' and the Foundation for Endangered Languages. The contents of this paper contains the material presented at the *Rara & Rarissima* conference 2006. Many topics discussed are covered in more detail in Singer (2006).

Notes

1. All examples are from my fieldnotes or other linguist's unpublished notes unless the cross-reference takes the author-date form of other citations, in which case it is taken from published materials.
2. Some typological work on the phenomenon is currently being done by the Surrey Morphology Group, as part of their project 'Extended deponency', see Corbett et al. (2005).
3. Message arguments are a type of frame argument. Frame arguments are subcategorised for by verbs but bear no indication of their syntactic status through cross-referencing on the verb or case marking on the NP (see Singer 2006 for a full account).

References

- Austin, Peter
1982 Transitivity and cognate objects in Australian languages. In *Studies in Transitivity*, Paul J. Hopper and Sandra A. Thompson (eds.), 37–47. (= Syntax and Semantics; 15). New York: Academic Press.
- Capell, Arthur and Heather E. Hinch
1970 *Maung Grammar* The Hague: Mouton de Gruyter.
- Corbett, Greville G., Matthew Baerman, Dunstan Brown, and Andrew Hippisley
2005 Extended deponency project online database. Accessed 10 March 2006. <http://www.surrey.ac.uk/LIS/MB/WALS/WALS.htm>
- Evans, Nicholas
2000 Iwaidjan: a very un-Australian language family. *Linguistic Typology* 4: 91–142.
- Evans, Nicholas
2003 *Bininj Gun-wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.

- Evans, Nicholas
2004 Experiencer objects in Iwaidjan languages (Australia). *Non-nominative Subjects*, Peri Bhaskararao and Karumuri Venkata (eds.), 77–100. (= Typological studies in language; 60). Amsterdam: Benjamins.
- Evans, Nicholas
2007 Pseudo-argument affixes in Iwaidja and Ilgar: a case of deponent subject and object agreement. In *Deponency and Morphological Mismatches*, Matthew Baerman, Greville G. Corbett, Dunstan Brown and Andrew Hippisley (eds.), 271–296. Oxford: Oxford University Press.
- Evans, Nicholas, and Bruce Birch
n. d. Iwaidja dictionary file. Accessed December 2005.
- Frantz, Donald G.
1995 Southern Tiwa argument structure. In *Grammatical Relations: theoretical approaches to empirical questions*, Clifford S. Burgess, Katarzyna Dziwirek and Donna Gerds (eds.), 75–98. Stanford: CSLI publications.
- Frantz, Donald G.
n. d. Empty arguments in Southern Tiwa. Unpublished manuscript.
- Givón, Talmy and Bommelyn, Loren
2000 The evolution of de-transitive voice in Tolowa Athabaskan. *Studies in Language* 24: 41–76.
- Harvey, Mark
2002 *A Grammar of Gaagudju*. Berlin / New York: Mouton de Gruyter.
- Heath, Jeffrey
1984 *Functional Grammar of Nunggubuyu*. Canberra: Australian Institute of Aboriginal Studies.
- Hewett, Heather, Anne Dineen, David Stainsby, and Robyn Field
1990 *Maung dictionary*. Aboriginal Studies Electronic Data Archive (ASEDA), Australian Institute of Aboriginal and Torres Strait Islander Studies, Canberra.
- Lee, Jennifer
1987 *Tiwi Today: a study of language change in a contact situation*. (= Pacific Linguistics; C-96). Canberra: Pacific Linguistics.
- Lee, Jennifer C.
1993 *Ngawurranungurumagi Nginingawila Ngapangiraga: Tiwi-English dictionary*. Nguiu, Bathurst Island, Northern Territory: Nguiu Nginingawila literature production centre.
- Nunberg, Geoffrey, Ivan A. Sag, and Thomas Wasow
1994 Idioms. *Language* 70: 491–538.
- Pawley, Andrew, Simon P. Gi, Ian S. Majnep, and John Kias
2000 Hunger acts on me: the grammar and semantics of bodily and mental process expressions in Kalam. In *Grammatical Analysis: morphology, syntax and semantics. Studies in honour of Stanley Starosta*, ed. by Videa P. De Guzman and Byron W. Bender, 153–185. Honolulu: University of Hawai'i Press.
- Plank, Frans
1984 Verbs and objects in semantic agreement: minor differences between English and German that might suggest a major one. *Journal of Semantics* 3: 303–360.

- Schultze-Berndt, Eva
 2000 *Simple and complex verbs in Jaminjung. A study of event-categorisation in an Australian language*. Doctoral dissertation. Department of Linguistics, Katholieke Universiteit Nijmegen.
- Simpson, Jane
 1991 *Warlpiri Morpho-syntax. A lexicalist approach*. (= Studies in natural language and linguistic theory; 23). Dordrecht: Kluwer.
- Singer, Ruth
 2006 *Agreement in Mawng: productive and lexicalised uses of verbal gender agreement in an Australian language*. Doctoral dissertation. School of Languages and Linguistics, University of Melbourne.
<http://eprints.infodiv.unimelb.edu.au/archive/00003242/>
- Vajda, Edward J.
 2003 Ket verb structure in typological perspective. *Sprachtypologie und Universalienforschung (Language Typology and Universals)* 26: 55–92.
- Vajda, Edward J.
 forthc. A Siberian link with the Na-Dene. *Anthropological Papers of the University of Alaska, New Series*. (accessed draft 4 August 2009 at <http://www.uaf.edu/anlc/docs/vajda-2008.pdf>)
- Van Driem, George
 1987 *A Grammar of Limbu*. Berlin: Mouton de Gruyter.
- Watkins, Laurel J.
 1984 *A Grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Young, Robert W., William Morgan, and Sally Midgette
 1992 *Analytical Lexicon of Navajo*. Albuquerque: University of New Mexico Press.

Syllabic Obstruents in Ahtna Athabaskan

Siri G. Tuttle

1 Introduction

The goal of this paper is to demonstrate some unusual patterns of syllabification in Ahtna Athabaskan. This Alaskan Athabaskan language can have up to five consonants in prefixal clusters in verbs, yet prefix phonology shows that syllabification does take place in these words. This pattern is uncommon within the language family. Ahtna shares the morphological environments of these clusters only with its close neighbor Dena'ina, which allows even longer clusters; other Athabaskan languages allow clustering in fewer morphological environments, and clusters are shorter (usually only two consonants in a row.)

Kari (1990) divides Ahtna verb phonology into three domains: the stem, the conjunct prefixes and the disjunct prefixes. Both canonical morpheme shapes and syllabification patterns differ over these three domains. This description suggests a phonological regularity to Ahtna verbs that depends on both phonological and lexical facts. The reason for this is that because of the templatic structure of the Ahtna verb, domains end up being defined by the presence of particular individual prefixes. Morpheme-specific rules defining phonological domains are suspect, as they should be.

As one of the three main domains in Ahtna phonology, Kari posits a “first conjunct syllable” in Ahtna (that closest to the stem) of the canonical form shown in (1):

$$(1) \quad C_0^5 V_0^2 C_0^2 \text{STEM}$$

The formulation in (1) could be interpreted as indicating that Ahtna may have a syllable with five onset consonants followed by a long vowel, followed by a coda of up to two consonants, just in the environment immediately followed by a verb stem. However, the formalization is equally consistent with a string of unsyllabified consonants. For the record, a search of Ahtna digital files produces only one morphological context that allows this sequence, and very few actual examples of it; Kari's formalization delimits the maximum cluster he can provide evidence of. Typical and frequent consonant clusters contain

two to three consonants. It is the manner of the consonants involved that makes things interesting.

What is unusual about the Ahtna clusters is that there is considerable evidence for syllable adjustments in verbal prefixes, showing that not just any consonant can follow just any consonant. I propose an analysis of some Ahtna prefix syllables based both on phonological evidence, mainly provided by Kari (1990), and acoustic evidence from fluent speaker texts.

I will also discuss the consequences of my analysis for Alaskan Athabaskan syllabification more generally; the picture that emerges is that of a dominant morphological-phonological alignment constraint interacting with a rare, but very functional, hierarchy of syllable peaks.

2 Ahtna Athabaskan: geography and internal structure

Ahtna Athabaskan is spoken in south central Alaska, along the Copper River, the Matanuska River, and in the area of Denali. Ahtna is considered to form a subgrouping of Alaskan Athabaskan with the more southerly and coastal Dena'ina (Figure 1 on the next page).

There are four dialects of Ahtna: Central (Gulkana, Tazlina, Copper Center, all on the Copper River); Lower (Chitina); Western (Denali area, Cantwell); and Mentasta (near the headwaters of the Copper River). All dialects share the unusual syllabification pattern I will demonstrate in this paper, although there are differences in syllable-driven adjustments in the prefix phonologies of Mentasta and the other Ahtna dialects, which I will not deal with here.

3 Ahtna and Athabaskan syllables

The Athabaskan language family is the largest in North America, with forty or so languages documented, depending on how one counts dialects. The languages are dispersed geographically into three main areas. Northern Athabaskan includes Alaskan languages and those spoken to the east in the Yukon, British Columbia, the Northwest Territories and Alberta. Pacific Coast Athabaskan includes Hupa (Northern California), Tolowa (Southwest Oregon) and numerous extinct languages in western Oregon and Northern California. The other famous branch of Athabaskan is the Apachean branch: Navajo, Western Apache and Eastern Apache languages are spoken in Arizona and New Mexico.

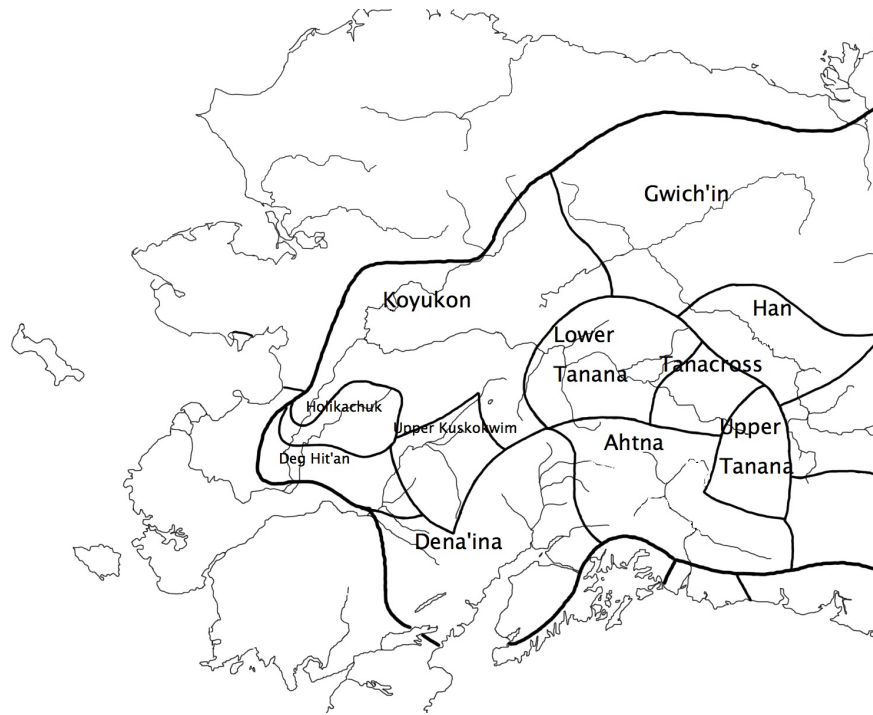


Figure 1. Athabaskan languages of Interior Alaska (after Krauss 1982; line map by Olga Lovick)

Some notes on the inventory in Table 1 on the following page are in order. The solitary labial stop is characteristic of Athabaskan consonantal systems, being a reflex of a proto-Athabaskan sonorant rather than a stop. The regular distinctions in the stop and affricate categories do not include voicing, but aspiration and glottalization; fricative releases of “plain” affricates are voiceless or partly voiced intervocalically, while those of aspirated affricates are voiceless with following aspiration. The alveo-palatal column includes sounds that are in free variation for place, as indicated in Table 1, section b. The “oral continuants” label hints at the phonological need to group together pairs of phones that alternate, despite their differing articulatory descriptions; [ç] is [j] devoiced, in the same word-final environment that renders [l] as [ɬ]. It does not, however, adequately demonstrate the contrast between the solid status of [ɬ] as a phoneme, and the marginal status of [ç]. This issue represents a family-wide phenomenon, and is worthy of another paper.

Table 1. Ahtna consonants (orthography and IPA)

1a: Ahtna Consonant Writing System

		Labial	Alve- olar	Later- al	Alveo- palatal	Pala- tal	Uvu- lar	Glott- al
Stops and Affricates	Plain	<i>b</i>	<i>d</i>	<i>dl</i>	<i>dz</i>	<i>g</i>	<i>gg</i>	'
	Aspirated		<i>t</i>	<i>tl</i>	<i>ts</i>	<i>c</i>	<i>k</i>	
	Glottalized		<i>t'</i>	<i>tl'</i>	<i>ts'</i>	<i>c'</i>	<i>k'</i>	
Oral Conti- nuants	Voiceless	<i>hw</i>		<i>ɬ</i>	<i>s</i>	<i>yh</i>	<i>x</i>	<i>h</i>
	Voiced			<i>l</i>	<i>z</i>	<i>y</i>	<i>gh</i>	
Nasals	Voiced	<i>m</i>	<i>n</i>				<i>ng</i>	

1b: Ahtna Consonant Inventory in IPA

		Labial	Alve- olar	Later- al	Alveo- palatal	Pala- tal	Uvu- lar	Glott- al
Stops and Affricates	Plain	<i>p</i>	<i>t</i>	<i>tɬ</i>	<i>ts ~ tf</i>	<i>c</i>	<i>q</i>	ʔ
	Aspirated		<i>t^h</i>	<i>tɬ^h</i>	<i>ts^h ~ tɬ^h</i>	<i>c^h</i>	<i>q^h</i>	
	Glottalized		<i>t'</i>	<i>tɬ'</i>	<i>ts' ~ tɬ'</i>	<i>c'</i>	<i>q'</i>	
Oral Conti- nuants	Voiceless	<i>ɱ</i>		<i>ɬ</i>	<i>s ~ ʃ</i>	<i>ç</i>	<i>χ</i>	<i>h</i>
	Voiced			<i>l</i>	<i>z ~ ʒ</i>	<i>j</i>	<i>ʁ</i>	
Nasals	Voiced	<i>m</i>	<i>n</i>				<i>ŋ</i>	

McDonough (2003: 4) discusses some of these issues with respect to the consonant inventory of Navajo.

3.1 Ordinary syllables in Athabaskan languages

Syllables in Athabaskan *stems* throughout the language family may include onsets, vocalic nuclei and single or occasionally restricted double codas. Onset clusters are rarely found: canonically, syllables in stems tend to look like CV(C)C or CVV(C), where VV is a metrically strong vowel, conventionally called “full”. In contrast, syllables in Athabaskan *affixes* throughout the family often include only an onset and a metrically light nucleus: CV. Some prefixes also contain full vowels. Table 2 on the next page shows some canonical Athabaskan verb words, with one light prefix and one heavy stem each, representing languages of each of the geographically separated areas of the family.

Table 2. Canonical Athabaskan verb words

	Orthography	IPA	Gloss
(a) Navajo (Southwest)	<i>dits'id</i>	[tə.ts'ət]	'It's tough'
(b) Galice (Pacific Coast)	<i>datq̣q̣</i>	[tə.t ^h ã:]	'It's thick'
(c) Ahtna (Alaska)	<i>dech'et</i>	[tə.tʃ'ət]	'It's tough'

3.2 Assumptions about syllables

I assume some things about syllables that I hope are uncontroversial, and will argue for other positions that facilitate description of the Ahtna data. I assume that syllables in Ahtna usually require onsets, and that codas are limited. Both of these assumptions are congruent with assumptions behind, for example, Optimality Theory, in which they are enshrined as constraints. I will argue, however, that the sonority hierarchy is less relevant to Ahtna syllable structure than other forms of rhythmic contrast.

I will count consonants as syllabic when their solitary appearance varies with Cə or əC in the same morpheme, in the same environment. I will also count as syllabic those consonants that speakers can separate as rhythmic beats. Usually, these two sets of syllables are congruent.

3.3 Assumptions about prefixal schwa in Ahtna

Any analysis of syllable structure in an Athabaskan language requires stated assumptions about prefixal schwa. Accounts vary by language and by linguist as to whether such schwas should be thought of as epenthetic (Kari 1990; McDonough 2003, for example) or underlying (Hargus and Tuttle 1997, for example), despite basic similarities in syllable patterns throughout the family. I will assume that many if not all Ahtna prefix vowels are underlying, and will show that this is the simplest assumption to adopt, given the variability in syllable expression to be seen below. However, this assumption holds only for the language under discussion. I assume for the Athabaskan family in general that prefixes may vary in representation both across and within languages, some bearing underlying vowels and some not, and that not all variations can be predicted based on comparative evidence.

Table 3. Syllabic nasals in Athabaskan languages

Language	Orthography	IPA	Gloss
(a) Jicarilla Apache (Southwest)	<i>ńyot</i>	[ń.joɬ]	‘It’s windy’
(b) Galice (Pacific Coast)	<i>nkei</i>	[n.k ^h ɛj]	‘Your foot’
(c) Ahtna (Alaska)	<i>nlaen</i>	[n.læ̃n]	‘(S)he is’

4 Unusual syllables in Ahtna

Ahtna prefix syllables are unusual both within the Athabaskan family and within Alaskan Athabaskan, but they are not outside of an overall family-wide pattern. First, omissible vowels clearly occur only in affixes, as in all Athabaskan languages with the exception of Dena’ina (Tuttle and Lovick 2007). Second, the Ahtna pattern is just a couple of steps farther along a scale of syllabic exoticism.

4.1 Syllabic hierarchy: Athabaskan syllabic sonorants

Nasals, nearly always /n/, can occur preconsonantly in nearly every Athabaskan language. Examples are shown in Table 3. The environment shown in the three examples is the same, word-initial, immediately preceding a stem syllable. This is the most common position for a syllabic nasal, although they may occur elsewhere. Throughout the history of Athabaskan description, recognition of syllabic nasals has been general (Haile 1941: 45; McDonough 2003: 95); this is likely because Apachean languages, which were described earlier than some others, have tone-bearing nasals as in (3a); a tone-bearing unit is a very plausible syllable. Thus, an onset sequence [nC] has not been proposed for Ahtna or Galice, and unusual versions of the sonority hierarchy have not been suggested.

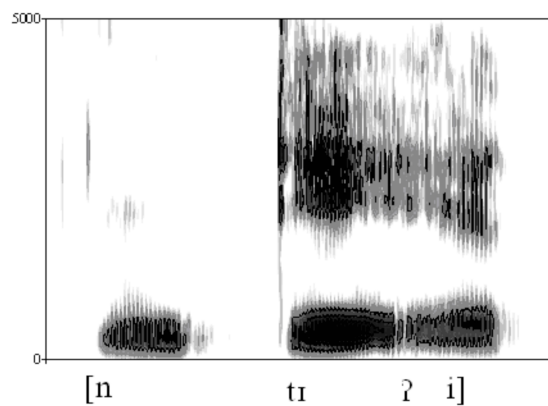
Less common among the languages, and also less plausible as a syllable, is a preconsonantal fricative or approximant. Examples are shown in Table 4. Most Alaskan Athabaskan languages do allow some preconsonantal fricatives, but they vary in how many of the fricatives participate. Ahtna has fewer constraints than most, but more than its neighbor Dena’ina. Ahtna does not sanction a voiced uvular fricative preconsonantly, but vocalizes this phone to a low back vowel. [ç] and [j] do not occur, nor is [χ] distinguishable from [h], in this context. [ʌ] is a nearly silent sound, for which the main perceptual cue is visual (lip rounding) but elders do not allow it to be elided.

Table 4. Preconsonantal fricatives and approximants in Ahtna

Orthography	IPA	Gloss
(a) <i>tsiin</i>	[t̚.ts ^h i:n]	‘(S)he made it’ (with complement)
(b) <i>ltsiin</i>	[l̚.ts ^h i:n]	‘It is made’
(c) <i>st’aaghe</i>	[s̚.t’a:ʔə]	‘Beneath me’
(d) <i>zdaa</i>	[z̚.ta:]	‘(S)he is sitting’
(e) <i>hdalts’i</i>	[h̚.tal.ts’i]	‘They are sitting, staying’
(f) <i>hwni’snikaen</i>	[ʌ̚.niʔ.s.ni.q ^h æ̃n]	‘We got stuck in a boat’

Rhythmically, speakers of Ahtna can pronounce these preconsonantal fricatives and approximants with some duration, and can even pause before beginning the next syllable. More common, however, is a weak-strong rhythmic pattern like a musical pickup note (the syllabic continuant) followed by a strong musical downbeat (the stem, or other strong syllable).

In most of the examples in Table 4, the IPA column represents the case where the syllable of interest can really consist only of consonantal material. The Ahtna writing system reflects this situation when it occurs regularly; that is, Kari’s transcription of Ahtna recognizes consonant-only pronunciation where it exists, and denotes the *impossibility* of such a pronunciation with a notated vowel. Figure 2 shows a nasal-initial word, *ndi’i* [n̚.ti.ʔi] ‘what’. There is no vowel preceding or following the nasal.

Figure 2. nC-initial word *ndi’i* ‘what’

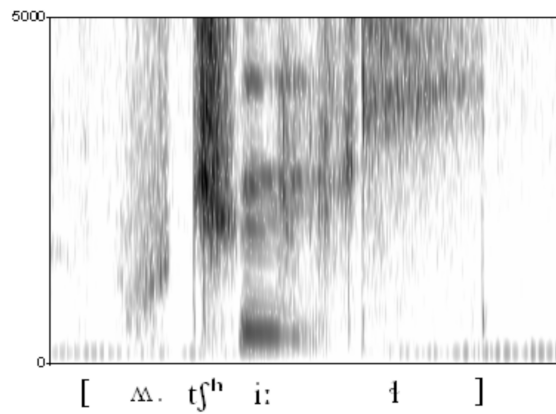


Figure 3. Word-initial, preconsonantal [ɬ]; *hwtsiit* 'bridge'

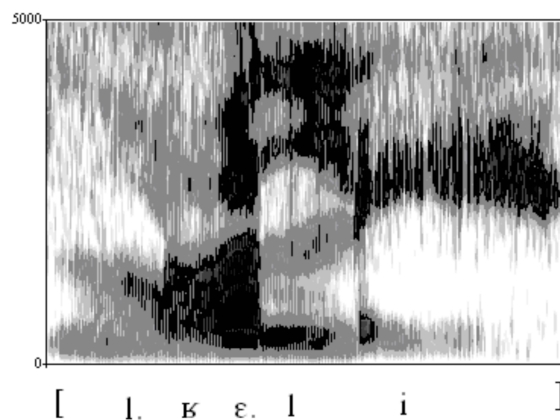


Figure 4. Word-initial, preconsonantal [l]; *lgheli* 'drum'

Figure 3 shows a fricative in word-initial, preconsonantal position, in the word *hwtsiit* [ɬ.tʰi:t] 'bridge'. Figure 4 shows word-initial, preconsonantal [l] in the word *lgheli* [l.ʁ.ɛ.li] 'drum'.

In many productions of word-initial, preconsonantal continuants, however, there may occur an extremely brief vowel. It tends to precede the continuant. Figure 5 on the facing page shows an example like this, with a very tiny (about 14 ms) and quiet vowel preceding a preconsonantal nasal. For Ahtna speakers, the production with the vowel and without the vowel are equiva-

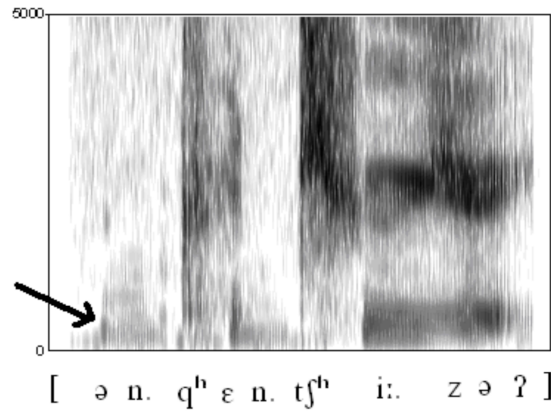


Figure 5. Preconsonantal nasal with brief vowel

lent. In voiceless environments within words, these very short vowels are often voiceless. Figure 5 shows the word spelled *nkentsiize*’ [n.qʰεn.tʃʰi:.zəʔ] ‘your moccasins’.

4.2 Syllabic aspirated stops in clusters

The hierarchy of syllabic consonants does not stop with fricatives. In Ahtna and in neighboring Dena’ina, but not in other Alaskan Athabaskan or other Athabaskan languages, aspirated stops can also form some sort of syllable, under particular conditions. Word-initially and in certain cluster conditions, aspirated [tʰ] seems to form a prosodic unit. Examples are given in Table 5 on the next page; these examples were published in Kari (1990); IPA is based on this author’s re-elicitations. I provide syllabification based on my analysis; naturally, the notion of a syllabic [tʰ] is not expected to be plausible without further argument.

There are two pieces to the argument for syllabic [tʰ] in the contexts seen in Table 5. The first piece comes from discussions with speakers, who with some variation count this consonant as a beat in the rhythm of the word. In slow syllabification, the breaks are as I have written them in Table 5. In different productions, a very short vowel may be found following the aspirated stop, or there may be no vowel. The number of syllables is the same, with some variation in speaker interpretation. The situation seems similar to that

Table 5. Aspirated stops in clusters

Ahtna Orthography	IPA	Gloss
(a) <i>uk'iile' tkut'ae</i>	[u.q'i:..ləʔ t ^h .q ^h u.t'æ]	'He is missed' (after death)
(b) <i>tket'aen</i>	[t ^h .q ^h ɛ.tæn]	'They are doing'
(c) <i>tkut'ae xu</i>	[t ^h .q ^h u.t'æ χu]	'The event' ('Where it is happening')
(d) <i>tkulaak</i>	[t ^h .q ^h u.la:q]	'He fixed it (a trail)'
(e) <i>tmedyaak</i>	[t ^h .nɛt.ja:q]	'It happened'
(f) <i>melzael</i>	[t ^h .nɛl.zæɫ]	'Music is playing'.
(g) <i>tes tnat'aan</i>	[t ^h ɛs t ^h .na.t'a:n]	'There is a series of hills'

described by Hoard (1978) for Bella Coola, another language with unusual syllable structure.

The second piece comes from phonological alternations in Ahtna prefixes. There is a set of systematic allophonic variations of prefix consonants in preconsonantal position; this includes the strikingly un-lenition-like shift of unaspirated [t] to aspirated [t^h] preconsonantly in the verb prefixes. The environment for these alternations refers to the *following* segment and not the preceding segment, so I believe that avoidance of onset clusters is part of the motivation for the alternations. In the next section I will show examples of the alternations and discuss their implications for prefix syllabification.

There are multiple possibilities for the prosodic role of the [t^h] in these word-initial syllables, if the reader will indulge my hypothesis that they are syllables. One is that they serve as onsets, and another that they serve as peaks. A third is that they constitute peaks with onsets — decent syllables on their own. This idea suggests that possible syllable peaks in Ahtna include (perhaps in scalar order, perhaps not) vowels, nasals, and fricatives — including aspiration. The possibility will be discussed further below.

4.3 Evidence for exhaustive syllabification in Ahtna

One logical possibility for consonantal Ahtna prefixes is that they are not syllabified. However, there is evidence for exhaustive syllabification of Ahtna verbal prefixes, in the form of a set of alternations that generally move consonants upward in sonority or continuance, making them better peaks, when they precede other consonants. These involve changes of [q^h] to [h], [c'] to [iʔ], and [t] to [t^h]. I consider these alternations to be changes that create bet-

ter syllable peaks. In the process of creating peaks, they also allow for full parsing of crucial consonant features, so that prefixes can be identified by hearers. The alternations thus serve both form and function.

4.4 Evidence for peak adjustments

As noted above, there are several lenition-type rules in Ahtna phonology that exchange laryngeally marked segments with unmarked segments specifically in preconsonantal position — suggestive of onset-rhyme alternation. Data in this section comes from Kari (1990), with underlying and additional forms added. The alternations shown in Table 6 exchange consonants with more closure for consonants with less closure: stops and affricates for fricatives, ejectives for vowel plus glottal stop. Therefore they seem appropriately termed lenitions, and since they occur in preconsonantal environments without reference to preceding environment, they would seem to be lenitions that provide consonantal syllables with better peaks.

Table 6. Syllable adjustment in Ahtna prefixes (Kari 1990)

Prefix gloss	“Onset” form	“Rhyme” form	“Onset” preceding vowel	“Rhyme” preceding consonant	What alternates
First person plural subject	ts’-	s’-	<i>ts’eyaan</i> ‘we are eating it’ [ts’ɛ.ja:m]	<i>stayiil</i> ‘We will eat it’ [ʔs.tʰa.ji:tʃ]	Ejective ~ glottal stop plus fricative
Indefinite object	c’-	i’-	<i>c’eyaan</i> ‘(s)he is eating sth.’ [c’ɛ.ja:m]	<i>i’tayiil</i> (CLW) ‘(s)he will eat sth.’ [iʔ.tʰa.ji:tʃ] <i>ctayiil</i> (Mentasta) [cʰ.tʰa.ji:tʃ]	Ejective ~ vowel and glottal stop
Third person human plural subject	k-	h-	<i>kiiyaan</i> ‘they are eating it’ [qʰi:ja:m]	<i>htayiil</i> ‘They will eat’ [h.tʰa.ji:tʃ]	Aspirated stop ~ voiceless fricative
Areal object or subject	k- (with following rounded vowel or rounding)	hw-	<i>koyaan</i> ‘(s)he is grazing, eating an area’ [qʰo.ja:m]	<i>hwtayiil</i> ‘(s)he will graze’ [ɰ.tʰa.ji:tʃ]	Aspirated stop ~ voiceless fricative

Less easy to class as lenitions are the prefixal alternations in voice onset time. These are alternations between aspirated *t* and unaspirated, voiceless *d* in the *D*-classifier, *t*- and *d*-qualifiers, and the disjunct *d*-# “proverb” prefix.

Kari (1990: 650) notes: “With stem-initial *b*, *n* and velar stops, *D* [the *D*-classifier] appears as a syllable coda.” The writing system recognizes this *D*-classifier allomorph as a *t*. Examples are given in Table 7, without syllabification (I will argue against the “coda” designation below). Again, I assume that vowels present in the output are underlying:

Table 7. Examples of *D*-classifier (spelled *t*) in postvocalic position

(a) /k(o)+e+ <i>D</i> +ba’/	<i>kutba’</i>	[q ^h ut ^h ba:]	‘It became twilight’
(b) /gh+e+ <i>D</i> +naa/	<i>atnaa</i>	[at ^h na:]	‘He is working’
(c) /d+e+s+ <i>D</i> +get/	<i>destget</i>	[tɛst ^h qɛt]	‘It is smoky’
(d) /da#e+ <i>D</i> +k’aali/	<i>datk’aali</i>	[tat ^h q’a.li]	‘snuff’

The right-hand environment for this alternation comprises all the non-coronal obstruents, plus the only nasal that occurs in stem-initial position. Preceding alveolar stops and affricates, the Ahtna *D*-classifier simply deletes. In addition to the *D*-classifier, the *d*-qualifiers (including object gender) and the disjunct “proverb” prefix have the underlying phonological shape *d* ([t]).

Other Ahtna prefixes exist in the conjunct sector of the verb that have the underlying shape *t* (aspirated alveolar stop, [t^h]). The most frequently encountered is the inceptive prefix, which is found in the first qualifier position (leftmost among the conjunct prefixes).

Kari (1990: 130) shows that underlying *d* prefixes surface as *t* before an aspirated or voiceless consonant. In addition, he notes (1990: 655) that “the conjunct prefixes *t*- and *d*- do not contrast before *n*”. In this environment, Kari spells both prefixes using the letter *t*. Examples are shown in Table 8. Kari’s examples suggest the possibility that syllabification processes are driving a laryngeal alternation between [t] and [t^h], but the direction of the alternation is puzzling. Why should a more marked value

Table 8. Neutralization of *t*-inceptive, *d*-qualifier before *n*.

(a) /y+ t +n+e+z+’iin/	<i>itnez’iin</i>	[i.t ^h .nɛz.ʔi:n]	‘He started to steal it’
(b) /na+ d +n+gh+e+l+nɛn/	<i>natnalnen</i>	[na.t ^h .nal.nɛn]	‘It flew down quickly’

(aspirated) be substituted for a less marked value (unaspirated) in *coda* position?

The observations behind Kari's analysis are easy to substantiate acoustically. Acoustic analysis of tokens of these relatively frequent prefixes in a Central Ahtna recorded text (Tuttle 2005) shows that Kari's notation using the aspirated symbol reflects a longer voice onset time for preconsonantal prefixal [t]. 193 examples of prefixes with underlying [t] and [t^h] were measured for voice onset time and closure. The examples came from two texts, narrated by a single storyteller at about the same time in her life (Kari 1978, 1982), Martha Jackson of Copper Center. The aspirated allophone of [t] in these prefixes is not nearly as long as that found in normal onset [t^h], but neither is it as short as that associated with normal onset [t]. Figures for this small acoustic study are given in Tables 9 and 10.

As the tables show, voice onset time is longer for underlying /t/ when it precedes a consonant than when it precedes a vowel; this is the reverse of the pattern seen for [t^h]. So the preconsonantal variant is not canonical for either /t/ or [t^h]; voice onset time is shorter for the aspirated, longer for the unaspirated members of the set. The neutralization is to neither underlying form.

Figure 6 on the next page shows a preconsonantal /t^h/ taken from audio text. The voice onset time of the initial alveolar stop in this word is about 18 milliseconds, contrasting with that of the following [q^h], which is over 40 milliseconds. The word is *tke'tiix* 'they do thus'.

Table 9. /t/ before consonants and vowels in Central Ahtna recorded text

	Before Consonants (8)	Before Vowels (60)	<i>p</i> -value (Scheffé's <i>S</i>)
Mean Closure	38.0	42.3	.51
Mean Voice Onset	18.3	8.0	.12

Table 10. [t^h] before consonants and vowels in Central Ahtna recorded text

	Before Consonants (24)	Before Vowels (39)	<i>p</i> -value (Scheffé's <i>S</i>)
Mean Closure	43.2	40.7	.77
Mean Voice Onset	11.5	23.4	.0017

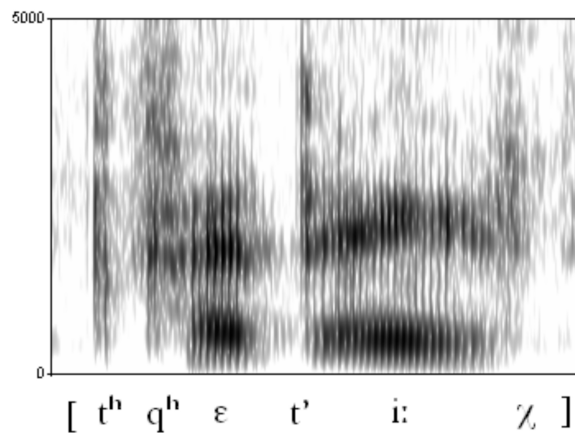


Figure 6. Preconsonantal word-initial aspirated stop

In this section I have demonstrated several unusual syllable shapes in Ahtna words, along with evidence that the prefixes expressed in these syllables are syllabified. In the following section I will suggest a way to make sense of the consonant clusters shown.

5 Discussion

Blevins (2004: 214) notes that clusters tend to be the result of

[...] uncommon convergence of significant rules of consonant or vowel loss resulting in long V and C clusters respectively; prosodic systems in which stress-timing, and not syllable-timing, prevails; and unambiguous rules of syllabification.

Ahtna's vowel deletion patterns are certainly related to durational effects of stress; in fact, the language has so regularized vowel quantity that nearly all long vowels are found in stems, which are stressed. All the consonant-only syllables seen in the examples above are found in prefixes, and in particular, in syllables that are unstressable because they contain, at best, a short vowel (Kari 1990; Tuttle 2003, 2005). The variability and apparent scalability of vowel presence in metrically light syllables suggest that vowels are assumed to exist in syllables, but that syllables with certain kind of consonantal noise as their peaks can work perfectly well in the same metrical context.

Blevins does not suggest that an unusual sonority hierarchy is at work in languages that build clusters of consonants or vowels, and it seems that the Ahtna facts would not make it necessary to make such a claim. What seems to be at work is a powerful constraint that aligns all prefixes to syllable edges. This must be accompanied, in Ahtna, by a constraint that maximizes the durational contrasts between stressed and unstressed syllables — which boils down, mostly, to stems versus prefixes. Without this constraint, the effect would more likely produce something like the Navajo pattern, where CV syllables are the norm in the conjunct prefixes.

The alignment of prefixes to syllable edges facilitates the perception, and thus the survivability, of prefixal consonants. The syllable edge aligned to /t/-prefixes in Ahtna is the left edge — a much better edge for parsing of the consonant's features than a right edge. Then, the extra release of added aspiration in the preconsonantal form of the prefix becomes a mechanism for optimization of parsing. As Blevins notes regarding similar facts in Georgian: “Presumably, the audible release of Georgian consonants [...] provide[s] salient cues to segment identification and syllabification” (2004: 214).

Blevins' appeal to perception brings to mind Ohala's (1990: 325) proposal that a collection of acoustic parameters should replace the single-valued “sonority” in phonetics: amplitude, spectral shape, periodicity and fundamental frequency may be manipulated. “Peak”, a notion from the sonority paradigm, gives way to “modulation”, emphasizing transitions rather than states. Since transitions are everything in perception of consonant contrast, the idea of modulation and transition as highly valued in syllable structure is a very practical one. As Ohala puts it, “The survivability of a given segmental sequence is proportional to the strength of [...] modulations” (Ohala 1990: 334). Functionally, the syllabification of obstruents is all about survivability of morphological content where light vowels are disfavored.

I propose, then, that the Ahtna syllabic obstruent is the result of a morphological-phonological alignment constraint dominating normal patterns of syllabification, which must be relativized for stems and affixes. While morphology-phonology interaction is a very common pattern in Athabaskan languages (see Hargus and Tuttle 1997) this particular type of interaction between syllables and prefixes is exemplified best in languages of central to western Alaska, in which stress is related to vowel length (the “stress-timing” noted by Blevins) and in which lexical tone is either in stiff competition with other prosodic systems (Lower Tanana, Tuttle 1998) or does not exist at all (Ahtna, Dena'ina.)

There is a favorable byproduct of this study, which is that it seems clear that deletion of metrically light vowels in prefixes, rather than epenthesis to repair bad syllables, seems to be the pattern in Ahtna. Both Kari (1990) and Causley (1994) have proposed analyses for Ahtna that require epenthesis, and in both cases, considerable diacritic labeling is required in order to get vowels in the proper positions. The phonetic analysis of text shows that tiny, vestigial-looking vowels occur in free variation with vowel-less sequences. Given this variation and the very specific consonant manners involved in these syllables, it would make no sense to suggest the insertion of vowels, and every kind of sense to suggest that the vestigial vowels have eluded complete deletion. This consequence has another favorable consequence, which is that questions about epenthesis versus deletion in other Athabaskan languages might possibly yield to a close acoustic analysis of natural speech.

References

- Blevins, Juliette
2004 *Evolutionary Phonology*. Cambridge: Cambridge University Press.
- Causley, Tricia
1994 *The minimal word in Ahtna (Athapaskan)*. MA Thesis, University of Toronto.
- Haile, Father Berard
1941 *Learning Navajo*. St. Michael's, Arizona: St. Michael's Press.
- Hargus, Sharon and Siri G. Tuttle
1997 Augmentation as Affixation in Athabaskan Languages. *Phonology* 14 (2): 177–220.
- Hoard, James
1978 Syllabification in Northwest Indian Languages, with remarks on the nature of syllabic stops and affricates. In *Syllables and Segments*, Alan Bell and Joan Bybee Hooper (eds.), 59–72. Amsterdam: North Holland.
- Kari, James
1978 Kudadzaey Ts'akae. (Spider Story). Manuscript. Fairbanks: Alaska Native Language Center.
- Kari, James
1982 Bac'its'aadi. (Salmon Boy Story). Manuscript in author's possession.
- Kari, James
1990 *Ahtna Athabaskan Dictionary*. Fairbanks: Alaska Native Language Center.
- Krauss, Michael E.
1982 *Native Peoples and Languages of Alaska*. Map. Fairbanks: Alaska Native Language Center. [originally published: 1975].
- McDonough, Joyce
2003 *The Navajo Sound System*. Dordrecht: Kluwer.

- Ohala, John
1990 Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *CLS 26: Papers from the 26th Regional Meeting of the Chicago Linguistic Society, vol. 2: The parasession on the syllable in phonetics and phonology*. Michael Zielkowski, Manuela Noske and Karen Deaton (eds.), 313–339. Chicago: Chicago Linguistic Society.
- Tuttle, Siri and Olga Lovick
2007 Intonational marking of discourse units in two Dena'ina texts. *Nouveaux cahiers de linguistique française* 28: 305–316.
- Tuttle, Siri
1998 *Metrical Structures in Tanana Athabaskan*. Unpublished doctoral dissertation, University of Washington.
- Tuttle, Siri
2003 Realizations of stress and intonation in an Ahtna text. Presented at the Society for the Study of the Indigenous Languages of the Americas, Atlanta.
- Tuttle, Siri
2005 Phonology of Prefix Consonant Clusters in Ahtna Athabaskan. In *ANLC Working Papers Vol. 5*. Suzanne Gessner (ed.), 128–145. Fairbanks: Alaska Native Language Center.

The puzzle of two terms for red in Hungarian

Mari Uuskiila & Urmas Sutrop

1 Introduction

The theory of the evolution of basic colour terms, created by Berlin and Kay (1969), states that in a language with a fully developed colour term system (black, white, red, green, yellow, blue, brown, grey, purple, pink, and orange) the maximum number of basic colour terms is 11. Berlin and Kay suggested that some languages may possess more than 11 basic colour terms, e. g. Russian and Hungarian. They argued that Russian has two basic blues – *sinij* and *goluboj* (cf. Davies and Corbett 1994; Paramai 2005) – and Hungarian has two basic terms for red — *piros* and *vörös*. Following this tradition, Plank sampled Hungarian as a rare language into his *Raritätenkabinett*.¹

Note, however, that Berlin and Kay interviewed only a single Hungarian subject, who lived in the San Francisco Bay region, and they consulted only one, non-authoritative and out-of-date Hungarian dictionary printed in the USA. As a result they found not only two basic terms for red in Hungarian, but also that the basic term for white is not *fehér* but *fejér*, which is actually an archaism. Instead of the basic *narancssárga* ‘orange’ they proposed *narancs* ‘orange’ meaning both colour and a fruit. That could be the result of English influence on the Hungarian subject who lived in the English-speaking environment, as in English the colour and fruit are designated with the same term (cf. Andrews 1994 for the influence of American English on the Russian émigré colour terminology).

The aim of this paper is to show that there are exactly 11 basic colour terms in Hungarian. One term *piros* for red is basic, whereas the second term *vörös* is not. We will report briefly on the results of our empirical study and discuss the puzzle of two terms for red in Hungarian as well as in the neighbouring Czech language. The definition of the basic colour term used in this paper is worded in the tradition in which Berlin and Kay used to define basic colour terms:

A basic colour term is a psychologically salient, in most cases morphologically simple and native word, which belongs to the same word class and has the same grammatical potential as the prototypical well-established basic

colour terms. It is a term that generally denotes an object, a quality, or a phenomenon at the basic level and which is applicable in all relevant situations (Sutrop 2001: 275).

2 Hungarian basic colour terms

2.1 Methodology and experiment

The data was collected by using the field method of Davies and Corbett (1994, 1995) that was developed further by Sutrop (2001, 2002). All subjects completed three tasks:

1. The list task where the subjects were asked to name as many colours as they knew.
2. The City University colour vision test for assessing the subject's ability to see colour (Fletcher 1980).
3. The colour-naming task. The task involved showing the subjects 65 (or 80 in a modified colour naming task with special emphasis on the yellow-red region) different colour squares at a time, in a random sequence. The subjects were asked to name the colour of the tiles.

The colour-naming task leads us to a technical definition of the basic colour term, which is different from our initial definition “the basic colour term is applicable in all relevant situations” (see above). Technically, a colour term is a basic colour term first if it is salient in a colour-listing task and second if the term crosses several technical hurdles in a colour-naming task. It is characteristic that, in the first task, the subject lists colour names without any reference to the real colour world and, in the second task, the sample colours are shown to the subject context-free on a colour chart or in random order against a light grey background.

Our data was collected from different regions in Hungary — in Budapest, Debrecen, Győr, Pécs, Dejtár, Ipolyvece, Balassagyarmat and Budaörs from 22 to 31 October 2002, and from 19 October to 5 November 2003. The interviewer (Uusküla) spoke Hungarian with the subjects.

There were 125 subjects in total, 66 women and 59 men, whose ages ranged from 9 to 82 years with the mean of 36 years. The subjects were all native speakers of Hungarian, with different dialect backgrounds. Some of them could not name their specific dialect — they just knew they were speaking “a little dialectally”.

The subjects did not know, until the immediate beginning of the test, that the questions would refer to colours and colour terms. All subjects did the colour-name list task first and then the colour-naming task. In the preliminary study in 2002, 40 subjects passed the standard colour-naming task with 65 colour tiles. In 2003, in the second series of interviews, 15 colour tiles from the yellow-red and pink-purple region were added. Yellow tiles were added because our preliminary results showed that sometimes the basic term for yellow was replaced with 'lemon-yellow' in Hungarian. Tiles from the red region were added to give a chance to the so-called second basic term for red, *vörös*. An additional reason for that was our wish to affirm or contradict the highly cited statement that Hungarians use the colour term *vörös* to indicate the dark red colour. The colour tiles added from the (Color Aid®) red spectrum were Life Red (normal red), ROR T1 (slightly lighter than a normal red), RV S1 (darker purplish red), R S1 (darker red), and R T1 (lighter red). The other 10 tiles were from the yellow, purple or pink region and are not considered in this article. 85 subjects performed the modified colour-naming task with 80 colour tiles.

In this paper, composite colour names are written, for the purpose of easier reading and transparency, with a hyphen “-” unlike the Hungarian literary norm, e. g. we write *narancs-sárga* instead of *narancssárga* ‘orange’.

2.2 Results

Here we present our results very briefly, concentrating on the terms for red. The full results will be published elsewhere (preliminary results have been published in Bogatkin-Uusküla and Sutrop 2005a, 2005b; Uusküla and Sutrop 2007; Uusküla 2008b). In 2002 the standard stimuli of 65 colour tiles and in 2003 the 65 standard stimuli and 15 extra colour tiles from the yellow-red region of the spectrum were used. The field work carried out in 2002 yielded a total of 3,432 colour names, the number of different colour names was 595. In 2003, the 85 subjects named 8,754 colour names in total, of which 979 were different. In all, the 125 Hungarian subjects named 1,149 different colour names.

When we sum the results of two years, we can see that in the list task 123 subjects of 125 named the colour term *piros*, while only 43 out of the 125 named *vörös*. The results of the list task are presented in Table 1, where 20 most frequently named terms are shown together with their naming frequency, mean position and cognitive salience index.

Table 1. The most salient colour names in the list task, where Fr = frequency, R = rank, Mp = mean position, S = salience

Term	Gloss	Fr	R	Mp	R	S	R
<i>piros</i>	'red'	123	1.0	3.480	1.0	0.283	1
<i>kék</i>	'blue'	115	2.5	4.357	2.0	0.211	2
<i>zöld</i>	'green'	114	4.0	5.018	4.0	0.182	3
<i>sárga</i>	'yellow'	104	8.0	4.962	3.0	0.168	4
<i>fehér</i>	'white'	115	2.5	5.678	5.0	0.162	5
<i>fekete</i>	'black'	113	5.0	6.327	6.0	0.143	6
<i>lila</i>	'purple'	110	6.0	8.655	7.0	0.102	7
<i>barna</i>	'brown'	106	7.0	9.085	8.0	0.093	8
<i>rózsaszín</i>	'pink'	94	10.0	10.011	9.0	0.075	9
<i>narancs-sárga</i>	'orange'	79	11.0	10.354	10.0	0.061	10
<i>szürke</i>	'grey'	98	9.0	13.204	15.0	0.059	11
<i>bordó</i>	'claret, bordeaux'	69	12.0	12.754	14.0	0.043	12
<i>citrom-sárga</i>	'lemon yellow'	49	13.0	12.612	13.0	0.031	13
<i>világos-kék</i>	'light blue'	45	14.0	15.000	21.0	0.024	14
<i>vörös</i>	'red'	43	16.0	14.419	20.0	0.024	15
<i>drapp</i>	'beige'	40	17.0	16.000	26.5	0.020	16
<i>sötét-kék</i>	'dark-blue'	44	15.0	17.636	31.0	0.020	16
<i>okker-sárga</i>	'ochre yellow'	37	18.0	15.459	24.0	0.019	18
<i>bézs</i>	'beige'	32	19.0	15.063	22.0	0.017	19
<i>világos-zöld</i>	'light green'	30	20.5	16.400	28.0	0.015	20

The results of the colour-naming task for the red region carried out in 2003 (with 80 colour tiles) are shown in Table 2 on the facing page. In this task 85 subjects named the colour tiles with the word *piros* on 210 occasions and with the colour name *vörös* only on 19 occasions. In most cases the term *piros* was applied to the colour tiles RO red-orange (51 times), ROR red-orange-red (44 times), ORO orange-red-orange (22 times) and R red (16 times). The colour word *vörös* was mostly used with the colour tiles RO (5 times) and ROR (3 times) — these tiles are not dark red in colour. From this result it is impossible to conclude that *vörös* refers to dark red, because it is used randomly for nearly all red tiles.

In addition to examining the usage of two reds, the whole Hungarian basic colour term inventory was investigated. The collected data enables us to conclude that in Hungarian only *piros* is a basic colour term, and the colour term *vörös* is not. In Hungarian there are exactly 11 basic colour terms: *fehér*

Table 2. Tiles named by 85 subjects with colour names *piros* and *vörös*, where tile codes in italics refer to extra tiles

Tile code	<i>Piros</i>	<i>Vörös</i>
ORO (orangish red)	22	0
RO (normal red)	51	5
ROR (normal red)	44	3
<i>ROR Tl</i> (slightly lighter than normal red)	13	1
R (normal red)	16	3
<i>R Tl</i> (lighter red)	8	3
<i>R Sl</i> (darker red)	1	2
RVR (purplish red)	3	0
<i>Life Red</i> (normal red)	46	1
Rose Red (pinkish red)	6	1
Total	210	19

Table 3. Basic colour terms in Hungarian

Present study	Berlin and Kay (1969)	English equivalent
<i>fehér</i>	<i>fejér</i>	white
<i>fekete</i>	<i>fekete</i>	black
<i>piros</i>	<i>piros</i>	red ₁
—	<i>vörös</i>	red ₂
<i>zöld</i>	<i>zöld</i>	green
<i>sárga</i>	<i>sárga</i>	yellow
<i>kék</i>	<i>kék</i>	blue
<i>barna</i>	<i>barna</i>	brown
<i>lila</i>	<i>lila</i>	purple
<i>rózsaszín</i>	<i>rózsaszín</i>	pink
<i>narancssárga</i>	<i>narancs</i>	orange
<i>szürke</i>	<i>szürke</i>	grey

‘white’, *fekete* ‘black’, *piros* ‘red’, *zöld* ‘green’, *sárga* ‘yellow’, *kék* ‘blue’, *barna* ‘brown’, *lila* ‘purple’, *rózsaszín* ‘pink’, *narancssárga* ‘orange’, and *szürke* ‘grey’. Thus, Hungarian is in line with the basic colour term theory.

We present our results in Table 3, with a comparison to the Hungarian basic colour terms as listed by Berlin and Kay (1969: 95).

The differences between the results of our study and the basic colour terms in Hungarian as offered by Berlin and Kay are briefly as follows. The second basic colour term for red offered by Berlin and Kay – *vörös* – is not a basic

colour term of Hungarian. First, its naming frequency was very low in the list task, and second, there was no consensus of opinion whatsoever as to which colour *vörös* might refer to in the colour-naming task. This result coincides with the opinion of MacLaury and his colleagues that the Hungarian *piros* is a basic word and *vörös* is not (MacLaury et al. 1997). Our understanding of the essence of the difference between *piros* and *vörös*, however, differs from MacLaury's (see below). Second, in modern Hungarian the name for 'white' (*fehér*) differs from that offered by Berlin and Kay (1969) (*fejér*). The difference may evidently be caused by the vernacular of Berlin and Kay's emigrant subject and the specifics of the dictionary they had at their disposal. Third, Berlin and Kay have a different name for 'orange'. The Hungarian *narancs* means also the fruit orange, while orange colour is mainly called *narancssárga*. Evidently, this difference is also due to the emigrant's language usage. Among the Russian emigrants in the USA, for example, some assimilation in the use of Russian colour terms with those of English has been observed (Andrews 1994). In English, indeed, *orange* means both the fruit and the colour.

What is interesting about the Hungarian colour names is that the basic term for yellow – *sárga* – is linguistically unstable and is in many circumstances replaced by another colour term *citromsárga*, 'lemon yellow'. This colour name forms, by analogy, a symmetric pair with the colour term *narancssárga* 'orange', which glosses as 'orange yellow'.

3 Discussions of the two terms for red in Hungarian

The Hungarian words denoting red – *piros* and *vörös* – have constituted a constant topic of discussion since the late 19th century. Today, this controversy does not only seem a good subject for linguists, but is also being studied by naturalists (chemists, physicists), psychologists and others. The discussion of two reds in Hungarian was actually initiated by the physician Csapódi in 1899, who suggested that in medicine (as well as in other sciences) there should be a clear distinction between what should be denoted by the colour word *piros* and what with the colour word *vörös*. He proposed a simple rule founded on experiments with colour-blind people. He suggests that while colour-blind people mix up one type of red colour with yellow and the other type of red with blue, the first red should always be referred to with the name *vörös*, while *piros* should be earmarked for the second (Csapódi 1899: 201).

The discussion is carried on by Gárdonyi, who is concerned with the rule in a new spelling dictionary which allows every language user to choose how to name the first colour of the Hungarian tricolour – *piros* or *vörös* ‘red’ – to be followed by *fehér* ‘white’ and *zöld* ‘green’ (Gárdonyi 1920: 84). Gárdonyi asks whether those two reds really mean the same as they are supposedly so freely replaceable. He offers many examples indicating that if *vörös* was the red colour with a yellowish overtone, then it could not occur in such compound words as *vörösrépa* ‘beet’ and *vöröskáposzta* ‘red cabbage’, because neither of them is yellowish in colour, being rather bluish red. Gárdonyi concludes that both *piros* and *vörös* are absolutely one and the same colour and the difference between them is cognitive — i. e. the colour name *piros* is mainly associated with good emotions and nice feelings, whereas *vörös* is linked with anger, bad emotions and is used in cautionary idioms, e. g. *elpirosodik* ‘blush’, *piros hajnal* ‘red dawn’, *piros a taréja kakasnak* ‘red is the rooster’s ridge’, *piros az én rózsám szája* ‘red is my darling’s mouth’; while *vörös kaka* ‘fire, literally red rooster’, *vörös terror* ‘red terror’, *vörös fene* ‘red devil’, *vörös számum* ‘red sand-storm’ and *elvörösödik* ‘turn purple (with anger)’ (Gárdonyi 1920: 85–86).

This discussion is taken further by Kenedy in 1921 who agrees with Gárdonyi in most of his statements. He adds, though, that the distinction between *piros* and *vörös* is not only cognitive, but the real hue of the two reds is also different (Kenedy 1921: 33). There follows a long pause, which is finally broken in 1947 by the chemist Selényi who tries with physical experiments to ascertain the colour of *piros paprika* ‘red paprika’. He suggests that the colour of paprika should be described by neither of the two terms for red but by the colour term *veres* (which is considered to be the phonetic variant of *vörös* in some dialects) (Selényi 1947: 12). He also finds that the distinction between *piros* and *vörös* is based on cognitive knowledge.

The issue of two reds in Hungarian became important in the study by Berlin and Kay (1969), which we have already handled in this article. Their study of basic colour terms has inspired many scholars to thoroughly examine several different aspects set up there. The two reds in Hungarian have also been a subject of discussion for general linguists and Slavists.² In some cases the research of the two names for red is limited to merely stating that *vörös* means ‘dark red’ (e. g. Crystal 1987: 106).

In 1997 an article titled “Hungarian *piros* and *vörös*: Color from points of view” by MacLaury and his colleagues Almási and Kövecses was published in the journal *Semiotica*. As a basis of appraisal the study by Berlin

and Kay is used, an extension of which is the vantage theory by MacLaury himself (1997). MacLaury and his colleagues involved nine Hungarian native speakers in Budapest. They used Munsell colour chips. According to their results only *piros* is a basic colour term, while *vörös* is not. Falling back on the vantage theory, the two Hungarian reds are studied according to the criteria of lightness and darkness. The authors bring out several interesting points: seven of their subjects, for example, relate *piros* and *vörös* to each other; *piros* is focused at middle brightness; *piros* is always named and mapped on more chips than *vörös*; mapping of *piros* includes the mapping of *vörös* etc. Their conclusion was that *piros* is not light red, while *vörös* should be still treated as dark red (MacLaury et al. 1997: 75). The latter statement is in contradiction with our empirical results (see above).

The two reds in Hungarian are also discussed by Wierzbicka, who defines colour terms through semantic components. She supplements the general definition of red with some semantic components, adding 'light' to the definition of *piros* and 'dark' to that of *vörös* (Wierzbicka 1996: 317). Often the two reds in Hungarian are also discussed by those linguists who study the two Russian blues, *sinij* and *goluboj* (e. g. Moss 1989a, 1989b; Paramei 2005).

Since the beginning of the present century the two reds in Hungarian have been studied by Hungarians themselves. In 2001 an article titled "Piros, vörös — red, rot, rouge" was written by Kiss and Forbes. In this article (available in Hungarian) they describe an experiment conducted with 98 subjects (all Hungarian native speakers). Relying on the results, the authors claim that *vörös* is a basic colour term in Hungarian, as is *piros* (Kiss and Forbes 2001:198). The most interesting part of this article is where the subjects have been asked to name collocations with *piros* and *vörös* (Kiss and Forbes 2001: 195–198). Collocations of these colour words are also discussed in Kiss (2004), where he assembles frequency lists on the basis of the Hungarian corpus (*Magyar Nemzeti Szövégtár*) to confirm the hypothesis of Berlin and Kay that Hungarian has two basic colour terms to indicate red. Besides that Kiss also presents lists of nouns co-occurring with either *piros* or *vörös* (Kiss 2004).

In 2003, De Bie-Kerékjártó studied the use of the colour word *vörös* in the Hungarian biggest daily newspapers *Magyar Nemzet* and *Népszabadság* (De Bie-Kerékjártó 2003: 69). She dealt with contexts where *vörös* is replaceable with *piros*. She compiled 12 context categories where *piros* is not able to substitute for the colour word *vörös* and thus concludes that both reds in Hungarian are basic colour terms (De Bie-Kerékjártó 2003: 70–78).

4 Etymology and motivation of *piros* and *vörös*

In Hungarian many adjectives are derived from nouns by adding an ending *-s* (also *-as*, *-es*, *-os*, *-ös*), the meaning of which is diverse: it may express that something is covered or decorated with something, contains something or is similar to something, or that somebody is dressed in something; in addition, it may express quantity, measure, age, time, distinctive feature or something with less intensiveness (e. g. *só* ‘salt’ > *sós* ‘salty’; *erő* ‘force’ > *erős* ‘strong, powerful’; *kalap* ‘hat’ > *kalapos* ‘with a hat’).

Both words for red in Hungarian are derived in the same way. They are denominal adjectives with an ending *-s*. The Hungarian *vörös* probably has a Finno-Ugric origin. The old Finno-Ugric root **wire* ‘blood’ can be found in the Finnic and Saami languages, in Mordvin, Mari, Komi and all the Ugric languages (Khanty, Mansi, and Hungarian). In a few languages a term for red is derived from this root, for example *verev* and *verrev* ‘red’ in South-Estonian dialects, *wiyr* ‘red’ in Mansi dialects, *wərtz* ‘red’ in Khanty (UEW: I, 576, Futaky 1981: 49–50, Sutrop 2002: 167). The root of the colour term *vörös* was probably *vér* ‘blood’ and its older form *véres* meant ‘bloody, covered with blood’. The primary meaning of this word was ‘the colour of blood’, i. e. ‘red’ (TESz: 1178). Over time the *é* of the first syllable was shortened and developed into *veres*, which is a phonetic variant of *vörös* even today. In our empirical tests *veres* was named only once, by a 50-year-old male subject from Carpathian Ruthenia (nowadays a part of Western Ukraine).

The colour adjective *piros* derives from the nominal stem *pír* ‘blush, rudiness’ and has suffix *-s*. It has no counterparts in other Uralic languages. It is argued that the stem *pir* may be identical with an onomatopoeic (sound-symbolic) descriptive stem (*pir-* ~ *per-* ~ *por-* ~ *pör-*) (TESz: 206–208, 275). It is possible that it was an old *nomen verbum*; its sound shape refers to the sound of baking over an open fire. Synaesthetically the meaning of sound may change into one of colour. Initially *piros* indicated the colour of grilled meat or roasted bread, which is either brown or at least dark in colour. Only later did it come to indicate the colour of blood (TESz: 208). Every modern Hungarian would probably say “*a vér piros*” ‘the blood is *piros* / red’ (De Bie-Kerékjártó 2003: 70). There are many nouns and verbs that belong to the same word family with the colour name *piros*, e. g. *piroslik* ‘to blush’, *pirosság* ‘blush’, *pirul* ‘to blush’, *pirkad* ‘to dawn’, *pirkadat* ‘dawn’, *pirók* ‘finch’, *perzsel* ‘scorch’, *pirít* ‘to roast, to fry’, *pörköl* ‘to roast’ etc.

Several of the Hungarian basic colour terms have a Turkic or Altaic origin (e. g. *kék* ‘blue; initially also indicated green’, *sárga* ‘yellow’), while several others are loanwords from Indo-European languages (e. g. *lila* ‘purple’, *barna* ‘brown’)

5 Compounds and collocations, semantics of the two reds

To have a better overview of the two terms for red in Hungarian it is first necessary to take a look at the compound words which are formed using *piros* and *vörös*, and second, at the similarities and differences in the behaviour of the terms *piros* and *vörös* in a context.

For compounds, the results gathered in 2002 and 2003 are used as material for analysis. We will view the first parts of the compounds (the second part is always *piros* or *vörös*) on the basis of a classification which was used for German colour words by Oksaar (1961). This system is used because it is easily adaptable to Hungarian and also because no other such classification has been offered for Hungarian so far. The words are grouped according to their semantics.

First, all those first parts of compounds which can equally well co-occur with both *piros* and *vörös* are presented; then those first parts co-occurring only with *piros*; and finally those with *vörös* are discussed. The written words should be read according to the scheme: first part (which is given) + *piros/vörös*. For example, the written *barnás* should be read either *barnáspiros* or *barnásvörös*. As regards collocations (where more than two words are written out) *piros* or *vörös* is affiliated to the second component, e. g. written *mély téglavörös* should be read *mély téglavörös*.

5.1 Common first parts in compounds and collocations

1. **other colour name + *piros/vörös***: *barnás* ‘brownish’, *lilás* ‘purplish’, *narancssárgás* ‘orangish’, *sárgás* ‘yellowish’;
2. **adjective + *piros/vörös***: *fakó* ‘pale’, *tompá* ‘blunt’, *halvány* ‘pale’, *közép* ‘middle’, *sötét* ‘dark’, *világos* ‘light’;
3. **pigment (artificial colour) + *piros/vörös***: *bíbor* ‘scarlet, purple’, *cinóber* ‘cinnabar’, *kármin* ‘carmine’;
4. **stones, minerals, precious stones + *piros/vörös***: *tégla* ‘brick’;
5. **oxides, metals + *piros/vörös***: *rozsa* ‘rust’;

6. **fruits, berries, vegetables + *piros* / *vörös***: *narancs* ‘orange’;
7. **liquids, water, beverages + *piros* / *vörös***: *vér* ‘blood’, *alvadt-vér* ‘clotted blood’;
8. **fire, ice, snow, sky, celestial bodies + *piros* / *vörös***: *tűz* ‘fire’

5.2 First parts which only co-occur with *piros*

1. **other colour name + *piros***: *feheres* ‘whitish’, *feketés* ‘blackish’, *narancs-sárgás-rózsaszínes* ‘orangish-pinkish’, *rózsaszín* ‘pink’, *zöldes* ‘greenish’, *sötétebb-feheres* ‘darker-whitish’, *sötét-lilás* ‘dark- purplish’;
2. **adjective + *piros***: *élénk* ‘vivid’, *érdekes* ‘interesting’, *erős* ‘strong’, *igazi* ‘real’, *meleg* ‘warm’, *mély* ‘deep’, *közönséges* ‘common’, *neon* ‘neon’, *rikító* ‘glaring’, *sima* ‘smooth’;
3. **pigment (artificial colour) + *piros***: *karmazsin* ‘carmine’;
4. **fruits, berries, vegetables + *piros***: *meggy* ‘cherry’, *paprika* ‘paprika’;
5. **liquids, water, beverages + *piros***: *bordó* ‘claret’³;
6. **plants, plant parts, flowers + *piros***: *ciklámen* ‘cyclamen’, *mályva* ‘mallow’;
7. **bones + *piros***: *foszfor* ‘phosphor’

5.3 First parts which only co-occur with *vörös*

1. **other colour name + *vörös***: *rőt* ‘reddish, russet’, *mély tégl*a ‘deep brick colour’, *sötét tégl*a ‘dark brick colour’;
2. **adjective + *vörös***: *matt* ‘flat’;
3. **pigment (artificial colour) + *vörös***: *terrakotta* ‘terracotta’;
4. **stones, minerals, precious stones + *vörös***: *cserép* ‘glazed tile’;
5. **oxides, metals + *vörös***: *bronz* ‘bronze’;
6. **fruits, berries, vegetables + *vörös***: *granát* ‘pomegranate’;
7. **liquids, water, beverages + *vörös***: *bor* ‘wine’;
8. **fire, ice, snow, sky, celestial bodies + *vörös***: *láng* ‘flame’;
9. **plants, plant parts, flowers + *vörös***: *rózs*a ‘rose’;
10. **wood + *vörös***: *mahagon* ‘mahogany’;
11. **animals, birds, fishes + *vörös***: *róka* ‘fox’;
12. **interjection + *vörös***⁴: *jaj* ‘oh’⁵.

5.4 Collocations

Next we will give an overview of collocations with *piros* and *vörös* according to a (2001) study by Kiss and Forbes (based on experiments as well as extensive corpus research). They have analysed which nouns are used with *piros* and/or *vörös*. They single out three groups: 1) nouns that are in any case used (mainly) with the colour word *piros*, e. g. *piros alma* ‘red apple’; 2) nouns that are in any case used (mainly) with the colour word *vörös*, e. g. *vörös csillag* ‘red star’; 3) nouns which can be used either with *piros* or *vörös*, e. g. *piros könyv*, *vörös könyv* ‘red book’. In the last group the meaning of the expression may sometimes change depending on the word used, on all other occasions the meaning remains neutral (Kiss and Forbes 2001; Kiss 2004). The results of Kiss and Forbes are presented in Table 4. This represents collocations (associations) that native speakers of Hungarian will have with either one or the other of the two terms for red (Kiss and Forbes 2001: 197).

In further research it could be useful to appeal to the semantic field theory of Lyons (1995: 62). The scope of collocations, i. e. the number of contexts where *piros* and *vörös* can occur, should be found out empirically. The collocations that these two colour words share should be discussed in paradigmatic terms, which means that the words in question are freely exchangeable within one syntagma. The collocations that are used with only one of

Table 4. Collocations with *piros* and/or *vörös*

Syntagmatic collocations		Paradigmatic collocations
Only with <i>piros</i>	Only with <i>vörös</i>	With <i>piros</i> or <i>vörös</i>
<i>alma</i> ‘apple’	<i>haj</i> ‘hair’	<i>vér</i> ‘blood’
<i>cseresznye</i> ‘cherry’	<i>róka</i> ‘fox’	<i>rózsa</i> ‘rose’
<i>lámpa</i> ‘lamp’	<i>csillag</i> ‘star’	<i>toll</i> ‘pen’
<i>Mikulás</i> ‘Santa Claus’	<i>meggy</i> ‘cherry’	<i>ceruza</i> ‘lead pencil’
<i>pont</i> ‘dot, full stop’		<i>pulóver</i> ‘pullover’
		<i>könyv</i> ‘book’
		<i>rúzs</i> ‘lipstick’
		<i>száj</i> ‘mouth’
		<i>póló</i> ‘polo’
		<i>kabát</i> ‘coat’
		<i>táska</i> ‘bag’
		<i>zászló</i> ⁶ ‘flag’

the two reds should be seen as syntagmatic relationships between the colour word and its modifying noun (cf. Lyons 1977: 230–269).

It has also been discussed exactly which colour or colours *piros* and/or *vörös* indicate. Many scholars are of the opinion that *vörös* stands for a physically darker colour than *piros* (Kálmán 1990; Sipőcz 1994; Wierzbicka 1996; MacLaury et. al 1997; Kiss and Forbes 2001; De Bie-Keréjártó 2003), others suggest that *vörös* is used for a yellowish red, while *piros* stands for a bluish red (see below; Csapódi 1899). According to the third opinion the use of *piros* and *vörös* is, rather, cognitive and traditional, whereas the hue is the same for both *piros* and *vörös* (Gárdonyi 1920; Selényi 1947; Kicsi 1988a, 1988b, 1991). Many articles and monographs stay neutral on this point (Mátray 1910; Bartha 1937). Many of these opinions are only speculations because there are very few studies about Hungarian colour terms based on empirical-cognitive field work experiments. Many of the subjects questioned by Uusküla confirmed that *piros* and *vörös* are of the same hue (thus being physically synonymous), only their meaning may be different depending on usage context and tradition.

6 Two reds in the neighbouring Czech language

We find that the use of two terms for red in Hungarian (syntagmatic and paradigmatic collocations) could be an areal phenomenon, because a similar case of two reds – *červený* and *rudý* – is found in the Czech language as well. And, similarly to the case of *vörös* in Hungarian, most dictionaries of Czech add the meaning *tmavě, sytě červený* ‘dark, saturated red’ to the term *rudý*. Etymologically, *červený* is connected to the blood colour of a caterpillar, while *rudý* has its etymological origins in a word indicating (iron) ore, (cf. Machek 1957: 70–71, 426–427). Uusküla carried out an empirical field work study with 52 Czech native speakers in Prague and Brno in 2007 (Uusküla 2008a). The results enable to conclude that Czech has only one basic colour term for red — *červený*.

The relationship of Hungarian *vörös* and Czech *rudý* is not so easy to prove as it may seem, especially if we follow the syntagmatic and paradigmatic collocations or context. The beetroot in Hungarian is *vörösrépa*, in Czech, however, it is *červená řepa*, not **rudá řepa*; red cabbage is, respectively, *vöröskáposzta* and *červené zelí*. In Hungarian the only correct expression for red wine is *vörösbor*, while in Czech both *červené víno* and *rudé víno* are possible. In spite of the different use of *vörös* and *rudý* there

are expressions where these colour terms are used as equivalents. These are phrases connected with socialism and, even more so, with communism. For instance, the Red Army is *vörös hadsereg* and *rudá armada*, a Red soldier is *vöröskatona* in Hungarian; a red star⁷ is *vörös csillag* and *rudá hvězda*; the flag of the former Soviet Union and the Red Army is *vörös zászló* and *rudý praporek* lit. ‘red flag’. The colour words *vörös* and *rudý* also correspond to each other in some other expressions, such as *vörös bolygó* and *rudá planeta* meaning ‘red planet, i. e. the planet Mars’. Extensive field work in Czech and other neighbouring Slavic languages is necessary to establish their inventory of basic colour terms and to examine whether our hypothesis is correct.

7 Conclusion

There is only one basic colour term – *piros* – denoting ‘red’ in Hungarian. The other term for red – *vörös* – is not basic. It follows that Hungarian possesses exactly 11 basic colour terms as predicted by the theory of Berlin and Kay (1969). The choice between the two terms for red is conditioned on the cognitive, emotional and/or collocational levels rather than the basic level.

Such a phenomenon with a choice between two reds on similar principles may be areal, because a similar use of two terms for red can be found (at least) in the Czech language.

The status of the basic colour term *piros* remains slightly contradictory for its use is restricted by *vörös* in some contexts/collocations. According to the original definition of a basic colour term its application “must not be restricted to a narrow class of objects” (Berlin and Kay 1969: 6).

We recommend that the Hungarian example about the two basic reds should be removed from the *Raritätenkabinett*. Of course we realize that the example of the two basic reds for Hungarian has already acquired a life of its own and thus we are hardly able to stop the *prima facie* intriguing example of the so-called two basic reds and twelve basic colour terms in Hungarian from being repeated in one linguistics textbook after another.

Acknowledgments

The study has been supported by the Estonian Science Foundation grant nos. 6744 and 8168.

Notes

1. See <http://typo.uni-konstanz.de/rara/intro/>
2. Unfortunately typesetters have made several mistakes in spelling, like *vörös* (Forbes 1979: 296) or *voris* (Spence 1989: 473).
3. Here it means wine.
4. Els Oksaar (1961) does not have this category.
5. The term *jajvörös* could perhaps be translated into English as 'very red'. Someone who is hurt may make a *jaj*-sound, an equivalent to English 'ouch!' If someone uses a colour expression *jajvörös* he must perceive it as a very strong stimulus, something like 'oh, my God! It's so red'.
6. However, syntagmatically the first colour of the Hungarian tricolour is always called *piros*, while only the flag of the former Soviet Union can be described by using the colour word *vörös* (*zászló*).
7. The emblem of some communist countries.

References

- Andrews, David R.
1994 The Russian *color* categories *sinij* and *goluboj*: An experimental analysis of their interpretation in the standard and émigré languages. *Journal of Slavic Linguistics* 2: 9–28.
- Bartha, Katalin
1937 *Szókinccstanulmány a magyar nyelv színelnevezéseiről* [Research of the colour vocabulary in Hungarian language]. (Dolgozatok a Debreceni Tiszta István Tudományegyetem magyar nyelvészeti Szemináriumából. 2. szám. Szer. Bálint Csűri). Debrecen: Városi Nyomda.
- Berlin, Brent, and Paul Kay
1969 *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.
- Bogatkin-Uusküla, Mari, and Urmas Sutrop
2005a *Tänapäeva ungari keele põhivärvinimed* [Basic colour terms of modern Hungarian]. *Keel ja Kirjandus* (Tallinn) 7: 558–570.
- Bogatkin-Uusküla, Mari, and Urmas Sutrop
2005b *Kas ungari keeles on kaks punase värvi põhinime piros ja vörös?* [Are there two basic colour terms for red in Hungarian?]. *Emakeele Seltsi Aastaraamat* (Tallinn) 50: 93–110.
- Csapódi, István
1899 *Vörös és piros*. *Magyar Nyelvőr* 28.5: 201–204.
- Corbett, Greville G., and Gerry Morgan
1987 Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics* 24: 31–64.
- Crystal, David
1987 *The Cambridge encyclopedia of language*. Cambridge / New York: Cambridge University Press.

- Davies, Ian R. L., C. MacDermid, G. Corbett, H. McGurk, D. Jerrett, T. Jerrett, and P. Sowden
1992 Color terms in Setswana: a linguistic and perceptual approach. *Linguistics* 30: 1065–1103.
- Davies, Ian R. L., and Greville G. Corbett
1994 The basic color terms of Russian. *Linguistics* 32: 65–89.
- Davies, Ian R. L., and Greville G. Corbett
1995 A practical field method for identifying basic colour terms. *Languages of the World* 9: 25–36.
- De Bie-Kerékjártó, Ágnes
2003 A vörös színnév használata a magyarban [The usage of colour word *vörös* in Hungarian]. In *Ünnepi kötet Honti László tiszteletére*. Marianne Bakró-Nagy, Károly Rédei (eds.), 67–79. Budapest.
- Fletcher, Robert
1980 *The City University Colour Vision Test*. 2nd ed. London: Keeler.
- Forbes, Isabel
1979 The terms *brun* and *marron* in modern standard French. *Linguistics* 15: 295–305.
- Futaky, István
1981 Zur Herausbildung des Farbfeldes im Finnougrischen. *Lakó emlékkönyv*, 44–58. Budapest.
- Gárdonyi, József
1920 Piros vagy vörös. *Magyar Nyelv* 16.4–6: 84–87.
- Heider, Eleanor Rosch
1971 “Focal” color areas and the development of color names. *Developmental Psychology* 4.3: 447–455.
- Heider, Eleanor Rosch
1972 Universals in color naming and memory. *Journal of Experimental Psychology* 93.1: 10–20.
- Kálmán, Béla
1990 Semantische Parallelen in den finnisch-ugrischen Sprachen. *Nyelvtudományi Közlemények* 91: 107–111.
- Kenedy, Géza
1921 Piros vagy vörös. *Magyar Nyelv* 17.1–3: 33–34.
- Kicsi, Sándor András
1988a Az alapszínnevek lexikalizálásáról [About the lexicalization of basic colour terms]. *Magyar Nyelvőr* 112.4: 456–467.
- Kicsi, Sándor András
1988b Az alapszínnevek lexikalizálása [Lexicalization of the basic colour terms]. *Természet Világa* 1: 43–44.
- Kicsi, Sándor András
1991 A magyar nyelv alapszínnevei [Hungarian basic colour terms]. *Nyelvünk és Kulturánk* 82: 50–53.
- Kiefer, Ferenc
2005 Van-e magyar jelentéstan? [Is there such thing as Hungarian semantics?]. *Magyar Nyelv* 2.2: 129–140.

- Kiss, Gábor
2004 *A piros, vörös és más színnevek használata a magyar nemzeti szövegtár alapján* [The usage of *piros*, *vörös* and other colour words in the basis of Hungarian Corpus]. *Variabilitás és nyelvhasználat*. Geccső Tamás (ed.), 160–165. (= Segédkönyvek a Nyelvészet Tanulmányozásához; 34). Segéd: TINTA Könyvkiadó.
- Kiss, Gábor, and Isabel Forbes
2001 *Piros, vörös – red, rot, rouge*. In *Kontrasztív szemantikai kutatások*. Geccső Tamás (ed.), 190–199. (= Segédkönyvek a Nyelvészet Tanulmányozásához; 11). Segéd: TINTA Könyvkiadó.
- Koski, Mauno
1983 *Värien nimitykset suomessa ja lähisukukielissä* [Colour terms in Finnish and related languages]. Savonlinna: Suomalaisen Kirjallisuuden Seura.
- Lyons, John
1977 *Semantics*. Vol. 1. Cambridge: Cambridge University Press.
- Lyons, John
1995 *Linguistic semantics: An introduction*. Cambridge: Cambridge University Press.
- Machek, Václav
1957 *Etymologický slovník jazyka českého a slovenského* [Etymological dictionary of the Czech and Slovak languages]. Prague: Nakladatelství Československé Akademie Věd.
- MacLaury, Robert E.
1997 *Color and cognition in Mesoamerica: Constructing categories as advantages*. Austin: University of Texas Press.
- MacLaury, Robert E., Judit Almási, and Zoltán Kövecses
1997 Hungarian *piros* and *vörös*: Color from points of view. *Semiotica* 114.1–2: 67–81.
- Mátray, Ferenc
1910 *A magyar színnevezésekről*. [On Hungarian colour names]. Kalocsa: Jurcsó Antal könyvnyomda.
- Moss, Anthony E.
1989a Basic colour terms: problems and hypotheses. *Lingua* 78: 313–320.
- Moss, Anthony E.
1989b Does Russian have a basic term for purple? *Linguistics* 27: 145–155.
- Oksaar, Els
1961 Über die Farbenbezeichnungen im Deutsch der Gegenwart. *Muttersprache* 71: 207–220.
- Paramei, Galina V.
2005 Singing the Russian blues: An argument for culturally basic color terms. *Cross-Cultural Research* 39.1: 10–34.
- Selényi, Pál
1947 *Piros és veres*. *Magyar nyelvőr* 72.2: 12–14.
- Sipőcz, Katalin
1994 *A vogul nyelv színnevei* [Mansi colour terms]. (= Studia Uralo-Altaica. Suppl.; 3). Szeged.

- Spence, Nicol C. W.
 1989 The linguistic field of colour terms in French. *Zeitschrift für Romanische Philologie* 105: 472–497.
- Sutrop, Urmas
 2000 The basic colour terms of Estonian. *Trames* 4.1: 143–168.
- Sutrop, Urmas
 2001 The list task and a cognitive salience index. *Field Methods* 13.3: 263–276.
- Sutrop, Urmas
 2002 *The Vocabulary of Sense Perception in Estonian: Structure and History.* (= Opuscula Fenno-Ugrica Gottingensia; 8). Frankfurt am Main etc.: Lang.
- Uusküla, Mari
 2008a The basic colour terms of Czech. *Trames* 13.1: 3–28.
- Uusküla, Mari
 2008b *Basic Colour Terms in Finno-Ugric and Slavonic Languages: Myths and Facts.* (= Dissertationes Linguisticae Universitatis Tartuensis; 9). Tartu: Tartu University Press.
- Uusküla, Mari and Urmas Sutrop
 2007 Preliminary study of basic colour terms in modern Hungarian. *Linguistica Uralica* 43.2: 102–123.
- TESZ
 1967 *A magyar nyelv történeti-etimológiai szótára.* Három kötetben: I A–Gy, II H–Ó, Ö–Zs, IV Mutató [Hungarian etymological dictionary]. Főszerk. Loránd Benkő. Lajos Kiss, László Papp (eds.). Budapest: Akadémiai Kiadó.
- UEW = Rédei, Károly
 1998 *Uralisches etymologisches Wörterbuch.* Bd. 1. Uralische und finnisch-ugrische Schicht. Wiesbaden: Harrassowitz.
- Wierzbicka, Anna
 1990 The meaning of color terms: semantics, culture, and cognition. *Cognitive Linguistics* 1.1: 99–150.
- Wierzbicka, Anna
 1996 *Semantics: Primes and Universals.* Oxford/New York: Oxford University Press.

Possessive voice in Wolof: a rare type of valency operator

Sylvie Voisin-Nouguier

1 Introduction

The purpose of this paper is to describe an uncommon valency operator in Wolof, the verbal morpheme *-le* which encodes a possessive relation between the subject and the object of a derived verb. After a brief presentation of background information to facilitate the discussion, internal possessive constructions in Wolof will be presented. First, the particularities of the “*le*-construction” will be compared to different strategies of external possession construction observed in different languages following the typology proposed by Payne and Barshi (1999). Second, the “*le*-construction” will be compared to the Japanese possessive passive, as well as to double derivation constructions, such applicative-passives in other languages. Finally, even if the possessive construction *-le* in Wolof, and external possession constructions or derivations including passive share some characteristics, I will conclude that this valency operator in Wolof is specific and will propose a hypothesis of its emergence.

2 Wolof: background information

Wolof¹ is a West Atlantic language spoken in Senegal and also in Gambia and Mauritania. Like many Niger-Congo languages, Wolof has a nominal class system. It is reduced in comparison with other Atlantic languages. It is composed of 8 consonants for singular and 2 consonants for plural:

- singular: *b-*, *k-*, *l-*, *w-*, *m-*, *g-*, *s-* and *j-*
- plural: *y-* and *ñ-*

In this system, the consonants are used as support for determinant markers and can't be analyzed as affixed class markers (**ñ-góor*; *góor ñ-i* ‘the men’; *góor ñ-ale* ‘these men’).

Wolof has also a complex verbal inflexion system, including focus marking (Subject Emphatic, Verbal Emphatic and Object Emphatic).

The distinction between subjects and objects (without any distinction between transitive and intransitive subjects) involves contrasts in both con-

stituent order (relatively rigid SVOX constituent order) and indexation of arguments in the verb form, but no case contrast. There is no class concordance between the nuclear function and the corresponding clitics. The subject clitics are combined with tense-aspect-mood morphemes and the object clitics have the following forms:

Table 1. Object clitics

1S	<i>ma</i>	1P	<i>ñu</i>
2S	<i>la</i>	2P	<i>leen</i> ²
3S	<i>ko</i>	3P	<i>leen</i>

The system of verb suffixes coding valency changes is complex. Verbal suffixes encoding valency changes are:

Table 2. System of valency changes

medio-passive	<i>-u</i>
causative	<i>-e, -al, -le, -lu, -loo</i>
applicative	<i>-e, -al</i>
co-participative	<i>-e, -oo, -ante, -andoo, -aale</i>
antipassive	<i>-e</i>
possessive	<i>-le</i>

This system of valency alternations calls for some remarks. First, notice that a passive derivation is missing in this inventory. Second, Wolof includes in its system a derivation *-le* that I call possessive. This suffix will be detailed in the following sections. Third, even if most of these derivations are common cross-linguistically, Wolof uses different markers for encoding the same valency modifications: the causative is expressed by five different markers, the applicative by two and the co-participative by five. In each voice alternations, the particular suffixes are specialized either for different meanings, or for different verbal classes.

For example, causative suffixes can be divided in two classes of derivation according to verbal classes: (i) *-e* and *-al* for intransitive verbs, (ii) *-loo, -lu* and *-le* for dynamic³ verbs. But, in these sub-classes, each morpheme has a specific meaning, so the values of causation are different and the syntactic structures of the causative proposition may also be different.

The suffix *-e* is a lexicalized causative suffix for some unergative⁴ verbs, like *gén* ‘go out (intr.)’ vs. *gén-e* ‘go out (tr.)’ and some inaccusative verbs, like *aay* ‘be forbidden’ / *aay-e* ‘forbid’. The derivation *-al* is a very productive causative derivation. It is compatible with all unaccusative verbs for product transitive causative counter-parts like *bax* ‘be boiled’ / *bax-al* ‘to boil’, *wex* ‘be white’ / *wex-al* ‘to whiten’. The suffix *-loo* is the most common causative derivation. It is used with unergative and transitive verbs with an indirect causative meaning. The suffix *-lu* is used only on transitive verbs. It introduces a new argument, the causer, in subject position, but this derivation also deletes the causee. In other words, the *-lu* suffix has a double effect on the valency of the derived verb: an augmentation in subject position (causer) and a reduction in object position (the former subject, the causee). Thus, the result of this derivation on transitive verbs is a transitive verb with new grammatical relations. In (1a), *ñaw* ‘to sew up’ has an agentive subject *keen* and a theme object *mbubb*. When it is derived with the suffix *-lu* (1b), the subject is the causer of the event and the theme is the object, but the former subject (the causee / agent) can’t appear (1c). In some sentences, the whole construction has a specific meaning, the causer is perceived as being the beneficiary of the event.

- (1) a. *Kenn ñaw-agul mbubb mi.*
 nobody sew.up-NEG3S booboo DEF
 ‘No one has sewn the booboo yet.’
 b. *Dama-y ñaw-lu roob.*
 EV1S-INACC sew.up-CAUS dress
 ‘I am making sew up the dress.’ (or: ‘I am getting the dress sewn’)
 c. **Dama-y ñaw-lu roob ci/ak Mamadou.*
 ‘I am making sew up the dress by Mamadou.’

The last suffix, the causative *-le* combines with unergative verbs and transitive verbs. It introduces a new argument in subject position. This derivation has a cross-linguistically unusual meaning, forming exclusively an associative causation. Traditionally, causative derivation is divided into two semantics groups: direct and indirect causation. Shibatani and Pardeshi (2001) show that these meanings constitute a semantic continuum, where the so-called associative meaning is an intermediary sense between direct causation and indirect causation. Associative causation is more frequently conveyed by a causative morpheme also coding direct or indirect causation. But, in some languages like Wolof, this meaning has its own dedicated marker.

- (2) a. *Tabax naa kër-am.*
 build P1S house-POSS3S
 'I built his house.'
- b. *Tabax-le naa ko kër-am.*
 build-CAUS P1S O3S house-POSS3S
 'I helped him build his house.'

Note that the causative derivation *-le* is different from the possessive *le*-derivation presented in this study. In possessive *le*-construction, the possessive relationships are not marked, they don't use a specific morpheme or a possessive marker. Whereas in causative *-le*, if there is a possessive relationship between two entities, possession is obligatorily marked by the internal possessive marker. Without this marker, the possession is not clear. For example, *Tabaxle naa ko kër gi* means 'I helped him build the house'; the house can be possessed by me, by him or by somebody else.

Thus, each causative suffix has a particular effect, and the specific meaning is linked to particular verbal classes. In Table 2, when several suffixes are listed to encode a same syntactic operation of valency change, note that some semantic differences are involved, like we come to demonstrate for the several causative suffixes. This also shows the complexity and the richness of the verbal derivation system of this language. In some cases a diachronic hypothesis can be made for the emergence of this plurality of forms, but synchronically these derivations all co-occur. Evidence for a diachronic hypothesis can be observed in co-occurrence patterns. Compound suffixes can be used on a verbal class, but each component derivations that make up the compound can't be used alone on this class of verb. For example, the derivation *-lu* is presented by Church (1981) as the fusion of an applicative derivation *-al* and the medio-passive derivation *-u*.

"Sur le plan formel, ce suffixe résulte vraisemblablement de la combinaison de *-al₃* bénéfactif et de *-u₃* réfléchi. Cependant la voyelle *a* n'est jamais présente : on dit *ràbblu* 'faire tisser pour soi', et non **ràbbalu*. Ce suffixe [*-lu*] diffère du *-al₃* en ce que le bénéficiaire est sous-entendu ; *dinaa* (**ma*) *beylu sama tool* 'je ferai cultiver mon champ (pour moi)'." (Church 1981: 287)

[On the formal aspect, this suffix probably results from the combination of the benefactive *-al₃* and the reflexive *-u₃*. However, the vowel *a* is never conspicuous: we say *ràbblu* 'to make weave for himself', and not **ràbbalu*. This suffix [*-lu*] dissent from *-al₃*, in that the beneficiary is implied; *dinaa* (**ma*) *beylu sama tool* 'I will make grow my field (for myself).']

The applicative derivation *-al* adds a beneficiary object (4). On this new derived verb, the adjunction of medio-passive *-u* involves the deletion of the agent participant (the subject) and the promotion of the beneficiary to subject function, leading at the same time to the interpretation of the subject as the initiator of the event (5).

- (3) *Sàmba tabax na kër gi.*
Samba build P3S house DEF
'Samba built the house.'
- (4) *Sàmba tabax-al na Waly kër gi.*
Sàmba build-APPL P3S Waly house DEF
'Sàmba built the house for Waly.'
- (5) *Waly tabax-al-u na kër gi. (tabax-al-u > tabax-lu)*
Waly build-APPL-MDP P3S house DEF
'Waly has got the house built.'

The meaning of this double derivation has been reinterpreted with a causative signification, described sometimes as a benefactive-causative. This composition can only be conceived of as the result of a diachronic process.

3 Possessive constructions

In Wolof, a possessor-possessum relationship can be expressed, as in other languages, by different strategies. Possession is built with the possessive lexical predicate *am* 'have' (6), *moom* 'to own' (7) and *yor* 'to have with oneself' (8).

- (6) *Moo am xale.*
ES3S have child
'She has a child.'
- (7) *Ku moom fas wii*
INTER own horse DEM
'Who owns this horse?'
- (8) *Su yor-oon caabi ji, jox ma ko.*
HYP have.with-PAS key DEF give O1S O3S
'If he had the key with him / on him, he would have given it to me.'

Moom and *yor* are used essentially for alienable possession, while *am* conveys both alienable and inalienable possession: *Am na kër yu bare* ‘He has several houses’; *Am na ñetti doom* ‘He has three children’.

Possessive predicates can combine with the suffix *-e*. The derivational *-e* morpheme present on these verbs is not productive; it operates only with possessive predicates. It doesn’t change the argument structure, but introduces a temporary possession.

- (9) *Mën-ul ñów, moo am-e xale bi.*
 can-NEG3S come ES3S have-e child DEF
 ‘She can’t come, she is the one who has the child.’

Possession can also be expressed by a genitive construction, as illustrated in (10) or by a possessive determiner (cf. Table 3).

- (10) *Woto-u Sàmba bi* (> *Wotoo Sàmba bi*)
 car-CONN Samba DEF
 ‘The car of Sàmba.’

Table 3. Possessive determiner

<i>sama woto</i>	‘my car’	<i>suñu woto</i>	‘our car’
<i>sa woto</i>	‘your car’	<i>seen woto</i>	‘your car’
<i>woto-am</i> > <i>wotoom</i>	‘his / her car’	<i>seen woto</i>	‘their car’

Against this background, we now turn to the possessive *-le* construction, the main topic of this paper. The derivation *-le* involved in this construction is included in the voice system and the features of the construction lead us to connect it to external possession constructions.

4 Possessive derivation

The verbal possessive marker *-le* is a valency-changing suffix. It increases the argument structure of the verb. Appearing on an intransitive verb, it derives a transitive verb with the following argument structure:

- an additional argument with the semantic role of possessor is introduced in the subject position;
- the object of the derived possessive verb cumulates the role of possessum, and the semantic role assigned to the subject in the non-derived construction.

- (11) a. *Woto bi gaaw na.*
 car DEF be.fast P3S
 ‘The car is fast.’
- b. *Sàmba gaaw-le na woto.*
 Sàmba be.fast-POSS P3S car
 ‘Sàmba has a fast car.’
- c. *Sàmba moo ko gaaw-le.*
 Samba ES3S O3S be.fast-POSS
 ‘Samba has a fast one.’

In example (11a), the subject is a patient/theme argument of which a state is predicated by the monovalent verb *gaaw* ‘to be fast’. In (11b), the same verb *gaaw* contains an additional morpheme, the suffix *-le*. The same participant *woto bi* occurs with the same semantic role of patient. However, a new argument *Sàmba* is introduced in the sentence. Thus, the derivation *-le* changes the grammatical relations. The patient-subject *woto* is demoted to object position and shares all the features of an object, e. g. it can be pronominalized with *ko* (11c). The object *woto* receives an additional meaning of possessum. *Sàmba* is the new argument introduced in the subject position, and its semantic role is possessor. It can’t be regarded as the agent, as it has no effect on the state of the patient *woto*.

The valency-increasing effect of the derivation *-le* differs from causative and applicative markers, which constitute the commonest types of valency-increasing operators. *Sàmba* in (11b) is not a causer, as is the new argument added by causative derivation. Even if the system of voice in Wolof also has a causative suffix *-le*, this causative derivation has an associative semantics, as we have seen in example (2a) on page 380. Moreover, *woto*, in example (11b), is neither a beneficiary, nor a comitative, nor an oblique argument promoted as the object of a clause, as it is the case in applicative constructions.

Returning to the features of the possessive construction with *-le*, we can also remark that the possessive relation between the two arguments of the derived verb does not involve any possessive morphology. Thus, the derivation *-le* builds a possessor-possessum relationship without specific possessive morphology, except the suffix *-le* itself.

This derivation is only possible with a limited class of intransitive verbs that can be characterized as unaccusatives, since a common feature of all those verbs compatible with the possessive *-le* is the non-agentivity of the subject. However, although most unaccusative verbs accept this marker, *-le*

seems to be more productive with verbs expressing quality, e. g. *rafet* ‘be beautiful’, *dee* ‘be dead, die’, *baax* ‘be good’ etc.

- (12) *Baax-le na ay tééré.*
 be.good-POSS P3S INDEF book
 ‘He has good books.’
- (13) *Góor gii, moo dee-le jabar.*
 man DEM ES3S die-POSS wife
 ‘This man’s wife is dead. (This man became a widower.)’
- (14) *Maa réér-le xar.*
 ES1S be.lost-POSS sheep
 ‘I lost my sheep. (I’m the one who lost a sheep.)’

5 External possession constructions and other possessive constructions

The features of *-le* constructions can be compared with so-called external possession, i. e.

“[...] constructions in which a semantic possessor-possessum relation is expressed by coding the possessor (PR) as a core grammatical relation of the verb and in a constituent separate from that which contains the possessum (PM).” (Payne and Barshi 1999: 3).

If we look again at the example (11), repeated in (15), the possessor *Sàmba* – the subject – and the possessum *woto* ‘car’ – the object – are two distinct constituents, treated as nuclear arguments of the verb *gaaw-le*.

- (15) *Sàmba gaaw-le na woto.*
Sàmba be.fast-POSS P3S car
 ‘Sàmba has a fast car.’

However, even though the construction of Wolof possessive verbs shares some features with the possessor-possessum relation described by Payne and Barshi in the external possession construction (EPC), it is nevertheless distinct from the different strategies of external possession found cross-linguistically. In their typology, Payne and Barshi (1999) describe four different strategies:

- incorporation;
- possessor raising;
- applicative voice;
- External possession construction without any morphological marker, including the so-called “double unaccusative” in Sinitic languages

5.1 Incorporation

In external possession constructions with incorporation, such as in Guaraní (example 16), the possessor is encoded as the subject of the clause and the possessum is incorporated in the verb.

(16) Guaraní (spoken in Paraguay, Velázquez-Castillo 1999: 78)

- a. *Che akã-jere.*
1INACT head-turn
'To me turns head' = 'I'm dizzy'
- b. *Hetymã-po'is*
3INACT=leg-thin
'(S)he had the legs thin (= thin legs)'

5.2 Possessor raising

The term "possessor raising" or "dative construction" is commonly used for dative clitic constructions in Romance languages, such as French (17), Spanish (18) and Romanian (19). In Romanian, we can see that "dative constructions" are not restricted to inalienable possession. In this construction, the possessor is an argument of the verb distinct from the possessum. In French (17), the possessor *vous* 'you' is the dative argument and the possessum the object argument of the verb *couper* 'cut'.

(17) French

Il vous coupe les cheveux.
'He cuts your hair.'

(18) Spanish

Me da vueltas la cabeza.
'To me turns head.' = 'I'm dizzy.'

(19) Romanian (Timoc-Bardy 1996: 242)

Îi pleac trenu-l
(he) him leave train-the
'His train is leaving.'

5.3 Applicative voice

External possession constructions can also be found in applicative clauses. In some languages, the applicative derivation does not require an internal possessive marker, as in Mohawk (20). In other languages, an internal possessive

marker must encode the possessive relationship between the two nuclear arguments, cf. example (21a).

- (20) Mohawk (Baker 1999: 293)

Wa-hi- 'sere-ht-óhare- 's-e'.

FACT-1SG:SUBJ/MASC:SG:OBJ-car-NOM-wash-BEN-PUNC

'I washed his car.' (better: 'I washed the car for him.')

- (21) Oluta Popoluca (Zavala 1999: 340)

- a. $\emptyset = ?o:k-u=k$ *tan=majaw.*

B3(ABS)=die-CMPL=ANIM A1(POSS)=wife

'My wife died.'

- b. *ta=küj-?o:k-ü-w=ak* *tan=majaw.*

B1(ABS)=APPL2-die-INV-CMPL=ANIM A1(POSS)=wife

'My wife died on me.' (or: 'I got affected by the fact that my wife died.').

In Wolof, the applicative derivation can be linked to possession relationship. With transitive verbs, the applicative marker *-al* adds a new object. When this object is a beneficiary, it can be the possessor of the "patient" object. Nevertheless, in this case, as in Oluta Popoluca, the possessor-possessum relationship must be expressed by an internal possessive marker, the possessive determiner *sama* 'my' (22).

- (22) *Mamadou seet-al na ma sama jabar.*

mamadou look.for-APPL P3S 1S POSS1S wife

'Mamadou looked for my wife for me.'

5.4 External possession constructions without morphological strategy and "double unaccusative"

External possession construction without morphological marking is the last strategy. It presents many similarities with the *-le* constructions. In some languages, monovalent verbs can have a transitive structure with a meaning of possession without resort to any particular morphology.

In Ilkeekonyokie (a dialect of Maasai), the subject encoded by the prefix *áa-* must be interpreted as the possessor of the object *entitó* 'girl' (23a). This possessive interpretation is induced only by the transitive use of monovalent verbs. Payne and Barshi (1999: 4) point out that, in this language, all monovalent verbs (derived or not) can appear in this type of construction.

- (23) Maasai, Ilkeekonyokie dialect (Payne and Barshi 1999: 4)

a. *áa-yshú en-titó.*

3>1-be.alive FEM.SG-girl.NOM

‘My girl is alive (with presumably positive effect on me).’

b. *k-áa-ból ol-páyyàn en-kútók.*

DSCN-3>1-open MASC.SG-man.NOM FEM.SG-mouth.ACC

‘The man will open my mouth.’ (lit.: ‘The man will open me the mouth.’)

These external possession constructions share some features with *-le* constructions. There is no internal possessive morphology and these constructions are restricted to monovalent verbs. But they differ by the absence of restriction to a specific verbal class, and by the absence of verbal derivation.

“Double unaccusative” constructions in Sinitic languages are also reminiscent of the Wolof constructions. Hilary Chapell (1999: 205) notes that ‘double unaccusative’ constructions are restricted to unaccusative verbs.⁵ She describes this construction as syntactically “aberrant”:

“[...] syntactic construction where the rules of grammar, narrowly understood, are violated: in the double unaccusative, intransitive process verbs take two arguments, one more argument than the verb valency should allow. [...] The two arguments of the intransitive verb designate possessor and possessum.” (1999: 195)

- (24) Taiwanese Southern Min (Chapell 1999: 204)

I pái tloh tò-kha.

3SG lame PRES left-leg

‘He has gone lame in the left leg.’

- (25) Cantonese Yue (Chapell 1999: 207)

Kui⁵ sei² joh² taai³ taai².

3SG die PERF wife

‘He was bereaved of his wife.’ (lit.: ‘He died wife.’)

Thus, as is the case with the *le*-construction in Wolof, the two arguments of the clause are in the syntactic position of subject and object. Subject and object arguments are respectively, the possessor and the possessum. But, contrary to the Sinitic and Maasai languages, in Wolof, the possessive constructions are not syntactically ‘aberrant’ because they are morphologically marked by *-le*. The external possession *le*-construction, in distinction to all strategies described, shows a specific morphology.

Now, the question is what kind of marker is this morpheme? Can we see in the *-le* suffix a complex verbal derivation which, in particular conditions, for example with unaccusative verbs, would take on a possessive interpretation, as found in some languages with passive voice?

5.5 External possession resulting from passive voice or applicative-passive derivation

In this section, two uses of passive derivation in different languages will be examined: a) a specific use of the *-(r)are* suffix in Japanese, sometimes called the possessive passive or adversity passive, and b) a double derivation with the passive, similar to the applicative-passive derivation in Tswana.

In the possessive passives in Japanese, the passive derivation does not have its canonical function and this specificity allows possessive interpretations.

5.5.1 *Non-canonical voice strategy: the possessive passive in Japanese*

When the passive suffix *-(r)are* is used with unaccusative verbs, a possessive relationship is established between the subject of the passive clause and the dative argument (prior subject) or affected relationship.

(26) Japanese (Gunji 1987: 63–64)

- a. *Tomio-wa Susan-ni keimusyo-ni ik-are-zunisun-da.*
 tomio-TOP susan-DAT jail-to go-PASV-do.without-PAST
 ‘Tomio was not adversely affected by Susan’s going to jail.’
- b. *Susan-wa Naomi-ni nak-are-ta.*
 susan-TOP naomi-DAT cry-PASV-PAST
 ‘Susan was adversely affected by Naomi’s crying.’

5.5.2 *Double derivation: applicative-passive constructions*

In Tswana, a double applicative-passive derivation on intransitive verbs produces an intransitive clause with an oblique argument introduced by *ke* ‘by’. With some verbs, a possessive relationship is established between the subject and the oblique.

(27) Tswana (Creissels, p. c.)

- Batho ba shelwa ke mantlo.*
 2people SC2 burn.APPL.PSV by 6house
 ‘People’s houses are burning.’

To summarize, on the one hand the *le*-construction in Wolof presents features of external possession construction. There is no internal possessive marker in the clause. The possessor and the possessum are in different nuclear argument positions, respectively subject and object. The best syntactic test revealing that the possessum is an object in Wolof is pronominalization (cf. (11)); recall that there is no passive derivation in this language. The restriction on intransitive verbs connects the *le*-constructions to the “double unaccusative” (Sinitic languages) and external possession construction without a morphological strategy (Maasai). But the verbal derivation *-le* in Wolof removes this construction from strategies of external possession construction, since the verbal derivation *-le* is not an applicative marker.⁶ On the other hand, the comparison with the possessive passive of Japanese and double applicative-passive derivation of Tswana shows that possessive relationship between different arguments of the clause can be produced by passive or compound derivation. But in this case, the linking between grammatical relations and the pair possessor-possessum is different from that occurring in *le*-construction. In Japanese, if the possessum is an accusative / object argument, the possessor is a subject / dative. In Tswana, the possessor has the subject function, but the possessum is demoted to oblique by passive derivation. Moreover, the absence of passive derivation in Wolof has been pointed out and the *-le* suffix can’t be related to the medio-passive derivation *-u ~ -ku*. Even though a synchronic passive derivation is absent in Wolof, a hypothesis of complex derivation will be developed in the next section.

6 Evolutional hypothesis of the *-le* morpheme

The comparison with possessive passive and applicative-passive constructions leads us to consider the present form *-le*, in Wolof, as the possible result of the grammaticalization of a double applicative-passive derivation. But, synchronically, the fact that Wolof does not have a passive derivation makes it impossible to analyze the valency change encoded by the possessive suffix as a combination of applicative and passive, as seems to be the case in Tswana. A plausible diachronic explanation however is that Wolof possessive *-le* results from the grammaticalization of a complex marker where the second marker is **-e*, at a stage of evolution when passive in Wolof was coded by a suffix **-e*.

6.1 Passive derivation in Atlantic languages

Doneux and Podzniakov (p. c.) have reconstructed a suffix **-i* in Proto-Atlantic as a passive derivation. In some Atlantic languages, for example in different dialects of Diola, the reflex is *-i* (see Sapir (1965) for Diola-Fogny and Bassène (2006) for Diola-Banjali).

(28) Diola-Banjali (Bassène 2006: 226)

- a. *Atejo na-jug-e figen si-bé sasú.*
 Atéjo s3s-see-TAM yesterday CL4-cow CL4.DEM4
 'Atéjo has seen the cows yesterday.'
- b. *si-bé sasú su-jug-i figen.*
 CL4-cow CL4.DEM4 CL4-see-PSF yesterday
 'The cows have been seen yesterday.'

In other languages like Singandum (a dialect of Sereer), Buy or Peul, the passive suffix is *-e*.⁷ All of these languages are classified in the North Branch of the Atlantic family, as is Wolof.

(29) Peul, dialect spoken in East-Niger (Labatut 1982)

- a. *ngelooa monn-at Iisa.*
 camel annoy-TAM Iissa
 'The camel is annoying Iisa.'
- b. *Iisa monn-ete.*
 Iisa annoy-TAM.PASSIVE
 'Iisa is being annoyed.'

So, the existence of a passive derivation *-e* in an earlier period in Wolof is conceivable. Some traces of this derivation are maintained in Wolof. For example we have already seen one derivation *-e*, which is specific to the possessive verbs and encodes temporary possession, in the same way that *ser* and *estar* do in Spanish. *Ser* is the copula used in possessive construction, but *estar* can be used to indicate temporary possession; note that *estar* is also the copula expressing resultative states.

In the system of voices, several derivations involve a reduction of valency or a remodelling of grammatical relations (cf. Table 4 on the next page).⁸ All of these voice markers include a suffix *-e*.⁹ Alone, this suffix has two functions: antipassive and reciprocal.

The antipassive (AP) effect occurs with transitive and ditransitive verbs with recipient objects. When these verbs are derived by *-e*, the object / recipi-

Table 4. Suffixes of reduction or remodelling of grammatical relations

co-participative	-e, -oo (<*u-e), -ante (<*ant-e), -andoo (<*ànd-u-e), -aale (<*aal-e)
antipassive	-e

ent is obligatorily deleted, but this deletion doesn't mean that no recipient is implied in the process; rather, similarly to antipassive in ergative languages, the antipassive derivation conveys a habitual or generic meaning.

- (30) a. *Xaj bi du màtt-e.*
 dog def. ENeg3S bite-AP
 'The dog doesn't bite.'
 b. **Xaj bi du matte xale yi.*
 The dog doesn't bite the children.
- (31) a. *Alal du jox-e màqaama.*
 fortune ENég3S give-AP prestige
 'Fortune doesn't give one prestige.'
 b. **Alal du joxe màqaama sàcc bi.*
 'Fortune doesn't give the prestige to the thief.'

With naturally reciprocal events as defined by Kemmer (1993)¹⁰, this derivation has a reciprocal function. The reduction of verbal valency can be observed either through the plurality of the subject (33) or the demotion of the object to an oblique function, see the clitic preposition *ag* 'with' (34).¹¹

- (32) *Lu mu daj, sànni la ko.*
 REL N3S meet throw O2S O3S
 'Anything that he meets, he throws it to you.'
- (33) *Ñu daj-e foofu ci pénc mi.*
 N1P meet-REC there LOC square DEF
 'We meet there at the square.'
- (34) *Ñu daj-e=eg ay waxambaane yu takku.*
 N3P meet-REC-with INDEF comrade CONN be.numerous
 'They met up with several comrades.'

The other co-participative derivations also contain *-e*. Combined with other morphemes, *-e* produces different meanings such as distributive actions (*-andoo*), as in (35).¹²

(35) *Ñoo dugg-andoo kàrce.*

ES1P enter-DISTR army

‘We went into the army at the same time.’

This suffix *-e* corresponds to the Plurality of Relations marker¹³ (PR) described by Lichtenberk (2000) for Oceanic languages. This term refers to a morphological marker which is used to encode reciprocal and certain other situations, but not reflexive situations. These others functions are: chaining, collective, converse, distributed, repetitive, depatientive, middle, kinship relations, and collective plurals. In Wolof, *-e* encodes reciprocal and depatientive (our antipassive), and with other morphemes:

-ante encodes reciprocal situations with verbs denoting non-naturally reciprocal events,

-andoo encodes simultaneous distributed situations,

-oo encodes non simultaneous collective situations, and

-aale encodes distributed situations.

6.2 Hypothesis of grammaticalization: *-le* < *-al-e*

The former section presented a proposal aimed at resolving the problem of the identity of the suffix *-e* present in the *-le* possessive marker, given the absence of passive derivation in Wolof. However, the identity of the *-l* or *-al* suffix has not yet been treated. In the presentation of causative markers, it was shown that some causative derivations are complex and that this complexity involves a suffix *-al*, systematically reduced to *-l* in the compound suffixes (cf. Church’s (1981) hypothesis for causative *-lu*). In the current system of valency changes, two suffixes both have the form *-al* and the effect of valency-increasing: a causative and an applicative suffix. Thus, one can ask does the suffix *-al* contained in the possessive *-le* derivation result from the causative or the applicative derivation?

To answer this question, several characteristics can be explored. Presenting the *le*-construction as the result of the grammaticalization of several voice markers entails, on the one hand, explaining the syntactic configuration of the *le*-construction through a double reorganization of grammatical relations, and on the other hand, providing sound justification for the possessive reinterpretation of the *le*-construction through voice blending.

For the syntactic configuration of *le*-constructions, the approach adopted here is that the suffix *-e* is responsible for the remodelling of grammatical relations and it can have a canonical or non-canonical use as in Japanese, and either the causative or the applicative derivation is responsible for the increasing valence of the intransitive verb.

On one hand, several arguments speak in favour of the applicative derivation. First, only the applicative derivation seems to be related to external possession construction constructions, and this derivation has effects similar to those of the non-grammaticalized double derivations in Tswana and other Bantu languages. Second, Payne and Barshi (1999: 17) note that even if a causative analysis might be suggested,

“the difficulty of a causative analysis [...] is that clear causative morphology is not known to surface in external possession constructions (and to our knowledge, a causative solution for EP has never been proposed in the literature for any language).”

Third, if transitivity of the possessive construction results from the following two operations on valency: augmentation via applicative or causative, and reduction (or remodelling of grammatical relations) via passive (or more specifically in Wolof, by plurality of relations marker), then the possessive meaning is more difficult to explain in the *le*-construction with causative than with the applicative derivation. Indeed, if something has some quality and, at the same time, someone receives some emotional or psychological effects of this quality, then one possible reinterpretation is that the human being is the possessor of the thing, and the meaning of benefit is more plausibly conveyed by the benefactive applicative than by a causative derivation.

Since syntactic reorganization with the applicative derivation is obligatorily linked to an verbal augmentation of valency via the addition of a new object, the plurality of relations derivation might be responsible for the promotion of this new object to subject position and consequently the demotion of the former subject. But, this demotion is not correlated with an oblique function, as in the Japanese possessive passive where the possessum has an object function.

- (36) a. *Woto bi gaaw na.*
The car is fast.
- b. **Woto bi gaaw-al na Sàmba.* Applicative derivation
The car is fast for Sàmba.

- c. *Sàmba gaaw-al-e na woto bi.* Applicative-PR derivation
Sàmba has a fast car.

This is the single plausible reorganization that could explain the subject function of the possessor with applicative derivation. The possessive interpretation is induced from the benefactive meaning of the applicative derivation *-al*. But, in Wolof applicative derivations are incompatible with unaccusative verbs. *Woto bi gaaw-al na Sàmba* might mean ‘The car has made Samba fast’ or something else, but never ‘The car is fast for/to Samba’.

Several arguments, on the other hand, speak in favour of a causative derivation playing a role in the system of voice in Wolof. First, there is already one causative suffix *-le*. This derivation has an associative meaning (cf. example (2)) and it is used only on dynamic verbs. Second, double applicative derivation and the plurality of relations marker with an antipassive function are attested in Wolof with dynamic verbs. We have seen that the use of *-e*, with an antipassive function (cf. (30) and (31)) is possible with verbs having a recipient object, so this derivation is possible with some transitive verbs and is fully productive with ditransitive verbs, in particular with ditransitive verbs derived by means of the applicative marker *-al*.

- (37) a. *Togg naa yàpp wi.*
cook P1S meat DEF
‘I have cooked the meat.’
b. *Togg-al naa la yàpp wi.*
cook-APPL P1S O2S meat DEF
‘I have cooked the meat for you.’
c. *Togg-al-e naa yàpp wi.* (> *Toggale naa yàpp wi*)
cook-APPL-PR P1S meat DEF
‘I have cooked the meat (for people).’

And, as we can see, in this double derivation, the suffix *-al* keeps its form; it is not reduced to *-l*. Third, the applicative derivation is never compatible with stative verbs; only the causative derivation *-al* is employed with this verbal class.

- (38) a. *Woto bi gaaw na.*
The car is fast.
b. *Sàmba gaaw-al na woto bi.* Causative derivation
Sàmba makes the car is fast.

- c. *Sàmba gaaw-al-e na woto bi.* Causative-PR derivation
 Sàmba has a fast car.

While, syntactically, a double derivation involving an applicative in the *le*-construction in Wolof is difficult to motivate, a causative derivation is plausible. Semantically, the plurality of relations marker can be responsible for abandoning the direct causative interpretation normally linked with *-al* causative. *Sàmba* is reinterpreted as the endpoint of the event and not as the initiator.

7 Conclusion

The *le*-construction shares several features with external possession construction. The possessor and the possessum are in different core arguments of the derived verb, respectively the subject and the object. The possessive relationship established between the subject and the object is expressed only by the verbal derivation *-le*.

We compared the *le*-construction to external possession construction strategies. The strategy closest to Wolof is the strategy without morphology (Maasai and Sinitic languages), essentially because the composition of suffix *-le* is synchronically not clear. It presents the same form as the associative causative *-le*, but apparently it shares no syntactic and semantic effects with this causative derivation. We also presented two kinds of voice derivation. The possessive passive in Japanese has the particularity to function with intransitive verbs. In this non-canonical use, the passive derivation builds a transitive clause where the two arguments maintain a possessive relationship. In Tswana, a double applicative-passive derivation creates the same effect. On this basis, a diachronic hypothesis has been proposed for the verbal derivation *-le*.

Despite the synchronic absence of passive derivation in Wolof, some remnants of an old passive derivation **-i* have been presented. The different functions of this trace lead us to identify it as a plurality of relations marker. Indeed, in its antipassive function, the derivation *-e* gives a habitual or general meaning of the process, and in this way it conveys the meaning that the process acts the same on every potential recipient. With its reciprocal function, the plurality of relations can be seen in that the relation between the A participant and the B participant is the same as the relation between the B participant and the A participant. But, with reciprocal function *-e* needs an added morpheme *-ant* to function with non-naturally reciprocal events. With other mor-

phemes, the plurality of relations marker encodes collective and distributed actions.

The second morpheme involved in the possessive, *-le*, has two possible sources: causative or applicative. With respect to strategies of external possession construction and double applicative-passive in Tswana, the applicative derivation should have been preferred. But, for the system of voice in Wolof, it appears more probable that the possessive construction is the counterpart of the associative causative derivation with unaccusative verbs.¹⁴

To summarize, the possessive marker in Wolof is the result of the grammaticalization of a causative marker – with an associative and direct meaning – and a plurality of relations marker. This diachronic hypothesis explains the syntactic organization of the possessive construction. The causative derivation adds a causer subject, and the theme argument, of which a state is predicated, is demoted to a syntactic object. On this basis, the plurality of relations marker has a remodelling effect on semantic role. The subject is reinterpreted as the endpoint of the event and not as the initiator. Consequently, the causative meaning is lost. The possessive meaning conveyed by the whole construction is perhaps due to the associative meaning sometimes linked with the causative derivation *-al*. Or, as in the external possession construction strategy without morphology, the possessive relationship can be simply induced by the non-canonical use of unaccusative verbs (i. e. they are used as transitive verbs). Indeed, with the plurality of relation effect, the derived verb is still non-dynamic, and yet has a subject, which presents many features of agent. In the same way, this diachronic hypothesis can explain why, with dynamic verbs, the causative meaning of causative *-le* ('help, assist') is fuzzy. The subject, as in possessive constructions, loses its role of initiator of the action; it is reinterpreted as an associative-agent participant to the process like in the following causative sentence: *Sàmba bey-le na ko tool yi* 'Sàmba helped him cultivate the fields'. The dynamicity of the verb also involves that the subject is not interpreted as an endpoint participant.

Thus, the particularity of Wolof is to have developed a specific marker for possessive constructions from a causative derivation, and the syntactic configuration of these constructions permits including them into external possession construction. Finally, the absence of such markers in others world languages can be explained first, by the absence of a causative marker in the expression of external possession construction and secondly, the voice system of Wolof. This complexity is not specific to the Atlantic family as a whole but is a language-specific property of Wolof.

Abbreviations

AP = antipassive; APPL = applicative; CAUS = causative; DEF = definite determiner; DEM = demonstrative; DISTR = distributive marker; ENEG = Emphatic negative; ES1S = Subject; Emphatic 1st person singular; ES3S = Subject Emphatic 3rd person singular; EV1S = Verbal Emphatic 1st person singular; HYP = hypothetic; INACC = imperfective (inaccompli); INDEF = indefinite determiner; INTER = interrogative; LOC = locative preposition; MDP = medio-passive; N3P = Narrative 3rd person plural; N3S = Narrative 3rd person singular; NEG = negative; O1S = object clitic 1st person singular; O2S = object clitic 2 person singular; O3P = object clitic 3rd person plural; O3S = object clitic 3rd person singular; P1S = perfect 1st person singular; P3S = perfect 3rd person singular; PAS = past tense; POSS = possessive voice; POSS3S = possessive determiner 3rd person singular; PR = plurality of relation marker; PRO = pronoun; REC = reciprocal; REL = relative marker.

Notes

1. The Wolof examples used here are extracted from various sources. The main sources are two dictionaries Wolof-French (Fal, Santos, and Doneux 1990; Diouf 2001) and two story books (Keteloot and Dieng 1989; Keteloot and Mbodj 1983).
2. The second and the third plural persons also have the same pronunciation [n¹u], but they are distinguished by two different orthographies, respectively *nu* vs. *ñu*.
3. By dynamic we understand all verbs with an agent subject, both intransitive and transitive.
4. The notions of unaccusative and unergative are taken from Relational Grammar developed by Perlmutter and Postal (1984) to divide intransitive verbs into two groups: those which have a patient subject are called unaccusative verbs; and those which have an agent subject: unergative verbs.
5. Except for *jau*² 'run, leave' in Cantonese.
6. Applicative markers, in Wolof, are *-al* and *-e*.
7. In Peul, the passive derivation is combined with tense-aspect-mood markers and presents *-a ~ -e* variations (Labatut 1982: 270).
8. In some cases voices which have a decreasing effect, such as middle or reciprocal, are described as voices where arguments receive a new syntactic function and this modification of valency involves the reorganization of grammatical relations. But, this kind of description focusses on syntactic effect and hides the semantic aspects of valency changes. In the reciprocal, for example, the reduction of syntactic valency is not correlated with the same reduction in semantic valency. The former object, at the same time that it is promoted to subject function, takes on a new semantic role of agent, but it also maintains its original semantic role of patient. And the former subject, which keeps its syntactic function (subject) and its semantic role of agent, receives a new semantic role of patient. Consequently, this kind of voice is best described in terms of a remodelling of grammatical relations, i.e. constructions where a verbal marker is implied in the redefinition of semantic roles of arguments.

9. For more details on co-participative voices in Wolof, see Creissels and Voisin (2008).
10. Naturally reciprocal events are events that are either necessarily (e.g. 'meet'), or else very frequently (e.g. 'fight', 'kiss') semantically reciprocal (Kemmer 1993: 102).
11. Wolof is an "and-language" in the terminology of Stassen (2000). In its vernacular varieties, the coordination of noun phrases in subject function is impossible (in urban variety, probably due to the influence of French, this kind of coordination is attested). Thus, if the speaker wants to present the participants in reciprocal actions as distinct entities, the best construction is to introduce the second participant into a prepositional phrase *ak ~ ag*.
12. The suffix *-andoo* results from the grammaticalization of *-ànd-u-e* 'go.with-MEDIO-PASSIVE-e'.
13. For Lichtenberk, the link between reciprocal function and these others situations is what he calls "plurality of relations":

"There is plurality of relations in an overall situation (event, state, etc.) if what can be considered to be basically one and the same relation holds more than once either between one or more participants and the event / state they are involved in, or between the relevant entities." (Lichtenberk 2000: 34)
14. As Payne and Barshi (1999: 9–10) note

"For the External Possessive Relation to assume a transitive subject relation is extremely rare cross-linguistically, and even where it does occur [...], it may mostly surface with rather stative transitive predicates."

References

- Baker, Mark C.
 1999 External possession in Mohawk: Body parts, incorporation, and argument structure. In *External possession*, Doris L. Payne and I. Barshi (eds.), 293–323. Amsterdam / Philadelphia: Benjamins.
- Bassène, Alain-Christian
 2006 Description du diola-banjal (Sénégal). Thèse de 3e cycle, Université Lumière – Lyon II, Département Sciences du Langage.
- Chapell, Hilary
 1999 The Double Unaccusative Construction in Sinitic Languages. In *External possession*, Doris L. Payne and I. Barshi (eds.), 195–228. Amsterdam / Philadelphia: Benjamins.
- Church, Eric
 1981 *Le système verbal du wolof*. Dakar: Université Cheikh Anta Diop de Dakar.
- Creissels, Denis and Voisin(-Nouguier), Sylvie
 2008 The verbal suffixes of Wolof coding valency changes and the notion of co-participation. In *Reciprocals and Reflexives: cross-linguistic and theoretical explorations*, Ekkehard König and Volker Gast (eds.), 289–306. Berlin / New York: Mouton de Gruyter.
- Diouf, Jean-Léopold
 2001 *Dictionnaire wolof*. Tokyo: Ilcaa.

- Fal, Arame, Rosine Santos and Jean Léonce Doneux
 1990 *Dictionnaire français-wolof*. Paris: Karthala.
- Frajzyngier, Zygmunt and Traci S. Curl (eds.)
 2000 *Reciprocals: forms and functions*. Amsterdam / Philadelphia: Benjamins.
- Gunji, Takao
 1987 *Japanese phrase structure grammar: A unification-based approach*. Dordrecht: Reidel.
- Kesteloot, Lilyan and Bassirou Dieng
 1989 *Du tieddo au talibé*. Dakar: Agence de la coopération culturelle et technique, IFAN.
- Kesteloot, Lilyan and Cherif Mbodj
 1983 *Contes et Mythes Wolof*. Dakar: Les nouvelles éditions africaines.
- Kitagawa, Y. and S. Y. Kuroda
 1992 *Passive in Japanese*. Manuscript. Indiana University and University of California, San Diego.
- Labatut, Roger
 1982 *La phrase peule et ses transformations*. Thèse d'état, Paris III. Atelier de reproduction des thèses. Université de Lille III.
- Lichtenberk, František
 2000 Reciprocals without reflexives. In *Reciprocals: forms and functions*, Zygmunt Frajzyngier and Traci S. Curl (eds.), 31–62. Amsterdam / Philadelphia: Benjamins.
- Martin, Samuel E.
 1959 *Essential Japanese: An introduction to the Standard Colloquial Language*. Tokyo: Tuttle.
- Payne, Doris L. and I. Barshi
 1999 External Possession: What, Where, How, and Why. In *External possession*, Doris L. Payne, and I. Barshi (eds.), 3–29. Amsterdam / Philadelphia: Benjamins.
- Perlmutter, David and Paul Postal
 1984 The 1-advancement exclusiveness law. In *studies in relational grammar*, David Perlmutter and C. Rosen (eds.), 81–125. Chicago: University of Chicago Press.
- Sapir, J. David
 1965 *A grammar of Diola-Fogny: A language spoken in the Basse-Casamance region of Senegal*. London: Cambridge University Press.
- Shibatani, M. and P. Pardeshi
 2001 The Causative Continuum. In *The Grammar of Causation and Interpersonal Manipulation*, Masayoshi Shibatani (ed.), 85–126. Amsterdam / Philadelphia: Benjamins.
- Stassen, Leon
 2000 And-languages and with-languages. *Linguistic Typology* 4: 1–54.
- Timoc-Bardy, Romana
 1996 Appartenance implicite vs. appartenance explicite en roumain. *Faits de langues* 7: 241–250.

Velázquez-Castillo, Maura

- 1999 Body-part EP constructions: A cognitive / functional analysis. In *External possession*, Doris L. Payne and I. Barshi (eds.), 77–107. Amsterdam / Philadelphia: Benjamins.

Voisin (Nouguier), Sylvie

- 2002 Relations entre fonctions syntaxiques et fonctions sémantiques en wolof. Thèse de 3e cycle, Université Lumière Lyon II, Département de Sciences du Langage.

Zavala, Roberto

- 1999 External possession in Oluta Popoluca (Mixean). In *External possession*, Doris L. Payne and I. Barshi (eds.), 339–372. Amsterdam / Philadelphia: Benjamins.

Index

The index to this volume is split up into two separate indexes for reason of clarity and usability. The **Subject index** (starting below) lists all keywords and topics. Pages which contain definitions of keywords are marked by bold-face page numbers.

All languoids mentioned in the running text, in the examples and in the notes are indicated in the **Language index** (starting on page 405). Languoids that have different names or spelling variants of their names are always listed under the form which is actually used in the text by the respective authors. Cross-references were added if the different names bear no resemblance, e. g. *Hocąk* and *Winnebago*.

Subject index

- affixation, 35, 263
- Africa, 145
- agglutination, 189
- agreement, 7, 13, 17, 47–51, 53–66, 68, 76, 276, 306–311, 325–337
- aktionsart, 211, 221, 222, 224
- ambiguity, 11–13, 15, 16, 18–20, 102, 106, 248, 249, 253, 286, 287
- America
 - *North America*
 - *Mesoamerica*
 - *South America*
- analogy, 50, 51, 57, 59, 62–67, 89, 176, 256, 267, 286–288, 331, 364
- anaphoric marking, **4**
- animacy, 75, 79, 80, 84, 85, 87, 88, 111, 238, 248–253, 255–263, 308, 317, 318, 330, 331
- Asia, 75–77, 86, 91, 186, 325
- assimilation, 131, 148, 154, 164, 179, 264
- asymmetry, 28, 290, 291
- Australia, 229, 325–327, 335
- bilingualism, 196, 200, 201
- borrowing → *loanword*
- case marking, 262, 263
- case system, 1
- child language, 165, 167
- clitic, 52–57, 60–62, 65, 66, 134–141, 238, 378, 385
- code-switching, 196
- cognition, 23, 258, 261, 263, 289, 361, 365, 371, 372
- color terms, 236, 237, 240, 242–244, 359–372
- complexity, vii, 24, 25, 34, 43, 44, 130, 140, 179, 190, 200, 236, 272, 284, 288–290,

- 295, 298, 299, 305, 307,
309, 311, 377, 380, 388,
389, 392, 396
- consonant cluster, 23–44, 123, 124,
127, 176, 237, 239,
341–356
- context, 1, 8, 12–14, 17–19, 84, 85,
89, 98, 99, 101–103,
105–108, 110–113, 190,
192, 196, 197, 211, 220,
248, 252, 287, 297, 325,
329, 360, 371, 392
- context (phonology), 36, 38, 55, 59,
66, 99, 135, 136, 342, 346,
350, 354
- continuant, 178, 343, 347, 348
- coverb construction, 229, 328
- crazy rules, 145
- definiteness, 12, 13, 75–81, 83, 85,
89, 90, 114, 153, 249, 250,
252, 255, 313–315, 351
- diffusion, 7, 19, 55, 60, 271
- discourse, 12, 18, 19, 84, 96, 98, 101,
110–113, 238, 261, 263
- dissimilation, 175
- distribution (of characteristics), 48,
65, 117, 118, 189, 192,
229, 235, 244, 274, 275,
282, 336, 337, 378, 398
- distribution (phonology), 24, 26–28,
122, 152, 154, 159, 161,
164, 166, 168, 235
- ditropic clitics, 68
- Dorsey's Law-sequence, **124, 125**
- endangered languages, 120
- Europe, 48, 49, 248, 337
- extinct languages, 342
- fish, 15, 227, 241, 319, 331, 369
- focus, 7
- geminate, 33, 252
- gender, 57, 75, 76, 102, 309,
328–331, 333–336
- gender (of speaker), 76, 78, 82, 84,
87
- geographic distance, 167, 170
- geographic distribution, 48, 50, 55,
60, 65, 169, 178, 271, 335,
342, 371
- gesture, 229
- grammaticalization, vii, 62, 64, 65,
76, 86–90, 145, 257, 389,
391–395
- honorific, 63–65, 78, 268, 300
- iconicity, 231, 235, 244, 247,
258–263, 265, 299
- inflection, 2–6, 10–12, 18, 19, 53, 55,
57, 59, 64, 78–80, 86, 120,
131, 134, 160, 188, 203,
219, 230, 262, 263, 271,
274, 329
- information structure, 2, 12–15, 18,
19, 379, 383, 396
- language acquisition, 88, 164, 165,
178
- language change, 32, 34, 36, 38, 40,
43, 44, 49–60, 63, 66, 67,
78, 86–90, 123, 124, 128,
145, 155, 156, 159, 161,
164, 178, 193–195, 198,
200, 201, 203, 212, 243,
244, 251, 256–259, 263,
271, 272, 278, 281–289,
292–295, 321, 380, 381,
389, 390, 395
- language contact, 185–201, 243, 300
- language endangerment, 86
- lenition, 90, 350, 351
- lexicalization, 76, 122, 123, 130,
141, 221, 325–337, 379

- lexicon, 28, 53, 63, 117, 129, 220,
221, 235, 237, 243, 255,
341
- linguistic pressure, → *language
endangerment*
- loanword, 7, 28, 155, 179, 200, 213,
243, 245, 252, 256–258,
368
- markedness, 24, 61, 68, 78, 82, 164,
199, 200, 249, 250, 255,
259, 262, 352
- Mesoamerica, 236, 245, 321
- metathesis, 41, 151–152, 154–157,
164–166, 168, 175, 176,
178
- mood, 3–5, 8, 14, 95, 293, 300
- morphophoneme, 279
- names, 28
- naturalness, 68, 145, 391, 398
- negation, 115, 187, 188, 267–299
- New Guinea, 230
- North America, 1, 119, 321, 325,
335, 342
- noun phrase, 7, 82, 85, 96, 98–100,
102, 103, 105–109, 186,
187, 321, 328
- number, 4–7, 17, 19, 20, 49, 53, 58,
63, 65, 75, 76, 78, 79, 102,
155, 172, 179, 202, 245,
247–263, 309, 335, 377,
392, 398
- numeral classifier, 76
- obstruent, 28, 30, 31, 33, 34,
123–125, 127, 128, 237,
341–356
- Optimality Theory, 27, 28, 152, 178,
345
- orthography, 32, 189, 190, 344–347,
350, 361, 373, 397
- paradigm, 221
- part of speech, 2, 19
- politeness, 78, 92
- polysemy, 185, 190–193, 200
- portmanteau, 325
- postposition, 3, 14, 188, 189, 321
- postpositional phrase, 3
- preposition, 68, 188
- prepositional phrase, 187, 398
- productivity, 25, 44, 68, 86, 96, 123,
128, 145, 146, 160, 163,
168, 175, 176, 189, 212,
215, 237, 251, 259, 325,
326, 333, 334, 379, 382,
384, 394
- pronoun, 60
- Prosodic Morphology, 28
- prosodic trough, 167, 178
- rara, v–vii, 66, 185, 198, 200, 320,
321, 337, 359, 372
- rareness, 47, 48, 59, 60, 63, 65, 66,
86, 117, 119, 141, 145,
200, 236, 244, 298, 321,
326, 333, 336, 337, 372,
396
- Raritätenkabinett, 359, 372
- reanalysis, 88–90, 221, 256, 257,
286, 287, 321
- reduplication, 31, 37, 38, 42, 105,
125, 126, 155, 235,
237–241, 243, 244
- reference tracking, 2, 5, 7–9, 11–19,
21
- Relational Grammar, 320, 397
- resonant, 28, 128, 129
- sample, 47–49, 97, 117, 186, 275,
291, 359
- semantics, 76, 88, 209–231
- Similar Place Avoidance, 162, 163,
175
- social deixis, **78**

- sociolinguistic situation, 185, 186, 198–201, 209
- sonorant, 30, 32–34, 37, 38, 40–43, 123–125, 164, 343, 346
- Sonority Sequencing Principle, 24, 26–29, 43
- South America, 321
- speech errors, 164
- spoken language, 125, 188, 189, 194, 199, 201, 202, 225, 267, 269, 271, 282, 300, 350
- stranding, 321
- stress, 114, 117–141, 189, 202, 203, 235, 237, 238, 240, 243, 354, 355
- substitution, 292, 353, 366
- suppletion, 335
- syllable structure, 23–44, 59, 120–122, 132–135, 137, 153–158, 164, 168, 169, 179, 190, 212, 214, 238–243, 341–356
- symmetry, 290, 364
- tense, 95–113
- tense-aspect-mood, 96, 97, 99, 100, 135, 198, 274, 285, 286, 291, 294, 298, 378, 390
- tone, 79, 80, 83, 85, 90, 92, 173, 235
- transitivity, 10, 98, 213, 227, 313, 322, 323, 325, 326, 328–331, 333, 336, 337, 377–379, 381–383, 386–389, 391, 393–396, 398
- triggering, 6, 10, 20, 38, 82, 85, 114, 127–129, 169, 175, 288
- universal, 24, 43, 145, 162, 163, 167, 274, 290, 359
- valency, 10, 397
- verb, 60
- verb phrase, 99
- word, 29
- word order, 4, 7, 51–53, 100, 186–188, 190, 212, 325, 329, 377, 378
- written language, 188, 194, 201, 202, 267, 269, 282
- zero marking, 64
- zero negative, 267–299
- zero-anaphora, 1, 7–10, 12, 14, 16–18, 85

Language index

- Abui, 209–231
 Ahtna, 341–356
 Aleut, 1–19
 Algonquian lgs., 119, 305–321
 Alorese, 209
 Ancient Greek, 321
 Anejom, 48
 Apachean lgs., 342, 346
 Arabic, 162
 Arbore, 248
 Arowak, 61, 68
 Athabaskan lgs., 326, 336, 341–346, 355
 Atlantic lgs., 390
 Au, 48
 Australian lgs., 229, 325–328, 336, 337

 Badaga, 193, 269–271
 Bantu lgs., 63, 145, 152, 156, 166, 393
 Basaa, 157
 Basque, 162
 Bavarian, 63–66, 68
 Bayso, 248
 Bella Coola, 350
 Berom, 172, 176, 177, 179
 Bininj Gun-wok, 329
 Bobangi, 157, 161, 167
 Brahui, 271
 Buy, 390

 Cantonese, 397
 Cantonese Yue, 387
 Central Alaskan Yup'ik, 21
 Chamicuro, 115
 Chichewa, 163
 Chiwere, 119, 120, 123, 124
 Classical Mongolian, 162
 Cushitic lgs., 260, 263, 264

 Czech, 359, 371, 372

 Dakota, 120
 Dena'ina, 341, 342, 346, 349, 355
 Dhegiha, 120
 Diola, 390
 Diola-Banjai, 390
 Diola-Fogny, 390
 Dravidian lgs., 186, 193, 267–299
 Dutch, 47–68, 257, 258, 264

 Ekari, 48
 English, 48, 56, 61, 68, 104, 110, 164, 165, 186, 193, 200, 201, 227, 230, 235, 248, 249, 252, 257, 258, 264, 272, 323, 331, 359, 363, 364, 373
 Eskimo lgs., 12, 21
 Eskimo-Aleut lgs., 1
 Europe, 48, 49, 248, 337
 Ewe, 157

 Faroese, 48
 Fehan, 48
 Finnic lgs., 367
 Finnish, 236
 Finno-Ugric lgs., 367
 Flemish, 48, 50, 52–58, 68
 French, 48, 60, 62, 68, 164, 185–190, 195, 196, 198, 200–203, 259, 294, 301, 385, 397, 398
 French (Canadian), 188, 199, 202
 French (European), 188, 202
 French Creole lgs., 198–200
 Frisian, 47–49, 62
 Fula, 162

- Gaagudju, 326, 332, 334–335
 Galice, 345, 346
 Georgian, 23–44, 355
 German, 47–49, 61, 331
 Germanic lgs., 331
 Greek (Ancient), 321
 Gondi, 271, 275, 300
 Guaraní, 104, 385
 Gunwinyguan lgs., 327
 Guyana Creole, 199

 Haiti Creole, 199
 Haya, 154
 Hekou Hmong, 87–89, 91
 Hocak, 117–141, 321
 Hollandic, 56, 58
 Homeric Greek, 321
 Huayuan Qoxong, 90, 91
 Hungarian, 359–372
 Hupa, 342

 Icelandic, 48
 Ikalanga, 145, 146, 150, 157
 Ilkeekonyokie, 386
 Ilocano, 236
 Imonda, 260, 261, 263
 Indo-Aryan lgs., 300
 Indo-European lgs., 48, 248, 300, 321, 368
 Inuit, 192
 Iruḷa, 193
 Iwaidja, 326, 327, 333–334, 337
 Iwaidjan lgs., 325, 327, 329, 330, 333
 Izere, 147, 167, 170–178

 Japanese, 247–249, 265, 377, 388, 389, 393, 395
 Jicarilla Apache, 346

 Kaili Hmu-Qanao, 91
 Kaili Qanao, 90
 Kalam, 229–231, 337
 Kam-Tai lgs., 77

 Kannada, 193, 268, 271, 272, 275, 276, 278, 291, 300, 301
 Kartvelian lgs., 25, 31
 Kashaya, 166, 178
 Ket, 326, 336
 Khanty, 367
 Kiowa, 260, 261, 263, 326
 Kiowa-Tanoan lgs., 326
 Kobon, 229
 Koḍagu, 193
 Koiari, 48
 Kolami, 194
 Komi, 367
 Konda, 275, 300
 Koṇḍa, 194
 Kota, 193, 271, 280
 Koyo, 157–159, 161, 167, 178, 179
 Kuḍux, 300
 Kukuya, 157, 167–169
 Kuvi, 300

 Labu, 48
 Lahu, 236
 Latin, 60, 189, 202
 Laz, 25, 31
 Lezgian, 192, 193
 Limbu, 337
 Lolo, 77
 Lower Tanana, 355
 Luganda, 154

 Maasai, 386, 387, 389, 395
 Malagasy, 162
 Malay, 209
 Malayalam, 193, 271
 Malto, 300
 Mandarin, 77
 Mankon, 157
 Mansi, 367
 Margi, 192
 Mari, 367
 Martinique Creole, 199
 Mathimathi, 166, 167, 178

- Mauritius Creole, 199
 Mawng, 325–337
 Megrelian, 25, 31, 36, 38, 41
 Miao lgs., 78, 82, 86, 87, 89–92
 Miao-Yao lgs., 75
 Middle French, 189, 202
 Middle High German, 60
 Middle Indic, 256, 257
 Mohawk, 192, 203, 385, 386
 Mordvin, 367
 Movima, 95–113

 Nahuatl, 298
 Navajo, 326, 336, 342, 344, 345, 355
 Niger-Congo lgs., 178, 377
 Nungubuyu, 329
 Nuosu, 77

 Oceanic lgs., 48, 392
 Ojibwe, 305, 306, 313, 316–320,
 322–324
 Old English, 61
 Old French, 60, 190, 202
 Old Georgian, 31, 32, 35–38, 42
 Old Kannada, 268, 269, 277
 Old Tamil, 194, 203, 269, 270, 282,
 283, 286, 287, 289, 300
 Oluta Popoluca, 386
 Oromo, 260, 261
 Ottawa, 316, 318, 323

 Papuan lgs., 48, 209, 230
 Peul, 390, 397
 Picardian, 62, 64, 65, 68
 Pidgin English, 162
 Pondicherien Creole French,
 185–201
 Portuguese, 68, 257
 Proto-Athabaskan, 343
 Proto-Atlantic, 390
 Proto-Bantu, 146, 153–155,
 159–161, 163, 168, 176,
 179

 Proto-Dravidian, 268, 275, 280, 281,
 284, 287, 290
 Punu, 157

 Quechua, 162

 Réunion Creole, 199
 Rhaeto-Romance, 48
 Romance lgs., 385
 Romanian, 385
 Russian, 62–65, 359, 364, 366

 Saami lgs., 367
 Sanskrit, 257, 258
 Sereer, 390
 Singandum, 390
 Sinhala, 247–263
 Sinitic lgs., 384, 387, 395
 Siouan lgs., 119, 120, 123, 125, 127
 Slavic lgs., 372
 Slovene, 61, 68
 Southern Tiwa, 326, 332, 335–336
 Spanish, 67, 243, 245, 385, 390
 Svan, 25, 31, 36, 38
 Swahili, 163, 179

 Taiwanese Southern Min, 387
 Tamil, 185–197, 200, 201, 268,
 271–273, 275, 278, 279,
 282–285, 295, 300
 Teke lgs., 147, 159, 161, 167–169,
 175
 Teke-Gabon, 168–170
 Telugu, 271, 277, 278, 291
 Tibeto-Burman lgs., 77
 Tiene, 145–178
 Tiwi, 326, 335
 Toda, 193, 271, 273, 275, 300
 Tolowa, 326, 336, 342
 Totonac, 235–244
 Totonac-Tepehua lgs., 236, 244
 Tswana, 388, 389, 393, 395
 Tulu, 193, 275, 279

Turkic lgs., 63

Turkish, 192

Tuvan, 236

Udmurt, 192, 193

Ugric lgs., 367

Uralic lgs., 367

Vanimo, 48

Vietnamese, 77, 298

Weining Ahmao, 75–90

Welsh, 68, 259

West Germanic lgs., 49, 65–67

West Greenlandic, 21

Winnebago, → *Hocqk*

Wolof, 377–396

Yaka, 157, 167

Yi, 86

Yoruba, 236

Yue (Cantonese), 387

Yup'ik (Central Alaskan), 21