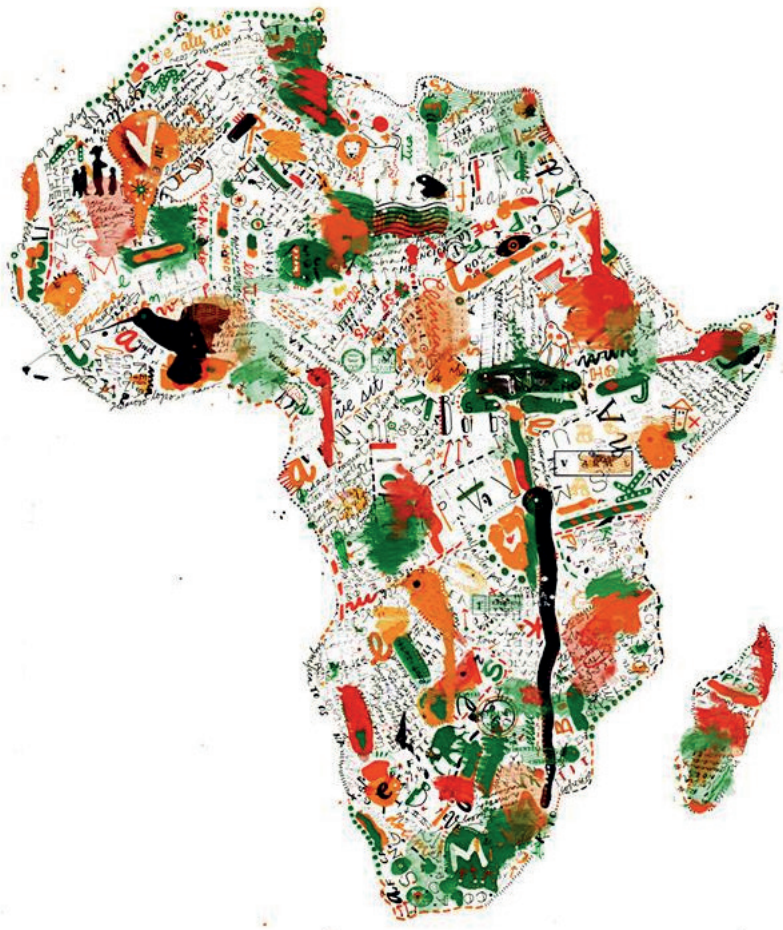
Stockholm University

Gender and its interaction with number and evaluative morphology

An intra- and intergenealogical typological survey of Africa

Francesca Di Garbo



Doctoral Thesis in Linguistics at Stockholm University, Sweden 2014

Gender and its interaction with number and evaluative morphology

An intra- and intergenealogical typological survey of Africa

Francesca Di Garbo

Gender and its interaction with number and evaluative morphology

An intra- and intergenealogical typological survey of Africa

Francesca Di Garbo

ISBN 978-91-7447-952-2

Printed in Sweden by US-AB, Stockholm 2014
Distributor: Department of Linguistics, Stockholm University
Cover design: Paolo Cerva
Map design: Ljuba Veselinova

Calipso: Tu che hai visto l'Oceano, i mostri e l'Eliso,
potrai ancora riconoscere le case, le tue case?
Odisseo: Tu stessa hai detto che porto l'isola in me.
Cesare Pavese, *I dialoghi con Leucò*

Abstract

This dissertation investigates the interactions between gender and number and gender and evaluative morphology in a sample of 100 African languages, and provides a method for assessing the role that these interactions play in the grammatical complexity of gender systems. Within the sampling space of one continental area, namely, Africa, in-depth studies of genealogically related languages are combined with large-scale comparison across unrelated genealogical units. The dissertation is organised around three main research foci.

First, the dissertation provides a systematic overview of patterns of interaction between gender and number along the following dimensions: exponence, syncretism, indexation, correlations in type of marking, and semantic interactions through gender assignment. Several of these patterns are identified, and the dissertation provides examples of their distribution across the languages of the sample. The study provides evidence for the possibility that, similar to the verbal domain, nominal features are also organised in a relevance hierarchy and that this can be studied by looking at nominal features as encoded both on nouns and on indexing targets. In addition, the study shows that, in languages with sex-based gender, gender and number may compete through indexation patterns and that animacy and/or lexical plurality play a crucial role in the expansion of the domains of use of special patterns of plural indexation. The study also shows that the development of pervasive indexation systems in the languages of Africa tends to always involve both gender and number. Finally, the study shows how gender assignment can be used as a means for encoding variation in the countability properties of nouns and noun phrases.

Second, the dissertation provides a systematic overview of patterns of interaction between gender and evaluative morphology in the languages of the sample. Two major types of interactions are found, and their occurrence depends on whether or not languages have dedicated diminutive and augmentative genders. The study shows that the distribution of the two types depends on three factors: (1) the type of gender system (sex-based vs. non-sex-based), (2) the number of gender distinctions and (3) the possibility of assigning the same noun to more than one gender.

Third, the dissertation investigates the role that interactions of gender and number and gender and evaluative morphology play in the absolute complexity of gender systems. The study proposes a metric for gender complexity and uses this metric to compute complexity scores for each of the gendered languages of the sample. The results suggest that the gender systems of the languages of the sample lean toward high complexity scores, that languages from the same genealogical units tend to have the same or similar complexity scores, and that the distribution of the outliers can often be understood as the result of language contact.

Finally, this dissertation provides an insight into the advantages of looking at nominal features through their morphosyntactic and semantic interactions rather than as isolated domains of grammar.

Contents

List of Figures ix

List of Tables xi

Abbreviations xiii

Acknowledgements xvii

1 Introduction 1

2 Background: gender, number, evaluative morphology and relevant interactions 3

2.1 Gender . . . 3
2.1.1 Gender assignment . . . 4
2.1.2 The morphosyntax of gender: indexation . . . 7
2.1.2.1 Beyond agreement: the notion of indexation and what we gain with it . . . 7
2.1.2.2 Patterns of gender indexation . . . 9
2.1.3 The function, distribution and diachrony of gender systems . . . 14
2.1.3.1 The function of gender . . . 14
2.1.3.2 The distribution and stability of gender systems . . . 15
2.1.3.3 The diachrony of gender . . . 15
2.2 Number . . . 18
2.2.1 Countability properties of nouns and noun phrases . . . 18
2.2.2 Types of number systems and number values . . . 20
2.2.2.1 General number . . . 20
2.2.2.2 Obligatory number marking and possible number values . 22
2.2.3 Number marking and the Animacy Hierarchy . . . 22
2.2.4 The morphosyntax of number: indexation . . . 23
2.2.5 Distribution and diachrony of number systems . . . 25
2.2.5.1 The distribution of number systems . . . 25
2.2.5.2 Sources of number . . . 25
2.3 Evaluative morphology . . . 26
2.3.1 The semantics, pragmatics and functions of evaluative markers . . 27
2.3.2 Types of marking in evaluative morphology . . . 29
2.3.2.1 Phonetic iconicity in evaluative morphology: universal or language-specific? . . . 29

2.3.2.2 Morphosyntactic encoding of evaluative markers 30
2.3.3 Distribution and diachrony of evaluative markers 31
2.3.3.1 The distribution of evaluative markers 31
2.3.3.2 The diachrony of evaluative markers 32
2.4 Gender and its interaction with number and evaluative morphology 32
2.5 The interaction between gender and number: state of the art 33
2.5.1 Exponence 34
2.5.2 Syncretism 35
2.5.3 Cumulative exponence, syncretism and relevance hierarchies 37
2.5.4 The interaction between gender and number in the languages of Africa 38
2.6 The interaction between gender and evaluative morphology: state of the art 39
2.7 Grammatical complexity of gender systems 40
2.8 Research questions 43
2.9 Summary of the chapter 44

3 Method: language sampling, data collection and organization 45
3.1 Sampling methodology 45
3.2 African languages: genealogical classification 47
3.3 African languages: large-scale language contact 52
3.4 Sampling procedure 53
3.5 Data collection and organization 58
3.6 Summary of the chapter 59

4 Gender, number and evaluative morphology in the languages of the sample: an overview 61
4.1 Introduction 61
4.2 Gender 61
4.2.1 Criteria of classification of gender systems: summary 69
4.2.2 Sex-based gender 71
4.2.3 Non-sex-based gender 74
4.2.3.1 Noun Classes in Bantu 76
4.2.3.2 Noun classes in Atlantic 78
4.3 Number 81
4.3.1 General number 87
4.3.2 Tripartite number systems 89
4.4 Evaluative morphology 91
4.5 Summary of the chapter 93

5 Gender and number 95
5.1 Introduction 95

5.2 Cumulation . . . 97
5.2.1 Cumulation between gender and number: results . . . 98
5.2.1.1 Cumulative encoding of gender and number with all the indexes . . . 100
5.2.1.2 Cumulative encoding of gender and number with some indexes . . . 101
5.2.1.3 Noncumulative encoding of gender and number on the indexing targets . . . 102
5.2.1.4 Cumulation between gender and number on nouns . . . 105
5.2.1.5 Summary . . . 109
5.3 Syncretism . . . 109
5.3.1 Syncretism of gender in the context of number: results . . . 110
5.3.2 *Semanticization* of gender distinctions under syncretism . . . 113
5.3.3 Summary . . . 116
5.4 Cumulation and syncretism: summary and discussion . . . 117
5.4.1 Cumulation, syncretism, and grammatical complexity . . . 117
5.4.2 Syncretism and relevance hierarchy for nominal features . . . 117
5.5 Split plural indexation systems . . . 118
5.5.1 Animacy-based uses of Dedicated Plural Indexation . . . 119
5.5.2 Dedicated Plural Indexation in Cushitic . . . 121
5.5.2.1 Introduction . . . 121
5.5.2.2 The distribution of DPI in the Cushitic sample . . . 122
5.5.2.3 Two peculiar languages within the Cushitic sample: Kambaata and Daasanach . . . 128
5.5.3 Discussion . . . 128
5.6 Presence of gender and type of number marking . . . 130
5.7 Gender shifts and semantic interactions between gender and number . . . 134
5.7.1 Gender shifts, number/countability in Bantu and North-Central Atlantic . . . 136
5.7.2 Gender shifts, number/countability in Berber, Semitic and Eastern Nilotic . . . 139
5.7.3 Gender shifts and semantic interactions between gender and number: summary . . . 141
5.8 Summary of the chapter . . . 142

6 Gender and evaluative morphology 145
6.1 Introduction . . . 145
6.2 Gender and evaluative morphology: overview of results . . . 145
6.3 Type 1: Diminutive and augmentative genders . . . 147
6.3.1 Distribution of diminutive and augmentative genders and general characteristics . . . 148
6.3.2 Multiple or additive class marking with diminutives and augmentatives . . . 152

6.3.3 The renewal of evaluative morphology in the southeastern Bantu languages . . . 154
6.3.4 Diminutive marking in Sɛlɛɛ . . . 158
6.3.5 Type 1: summary . . . 160
6.4 Type 2: Diminutives and augmentatives in languages with sex-based gender 161
6.4.1 Examples of Type 2 languages . . . 163
6.4.2 The distribution of Type 2 languages . . . 167
6.4.3 Type 2: summary . . . 169
6.5 Evaluative morphology and biological gender in languages without (sex-based) gender . . . 170
6.5.1 Bantu . . . 170
6.5.2 Western Nilotic . . . 172
6.5.3 Akan . . . 173
6.5.4 Summary . . . 174
6.6 Absence of interaction between gender and evaluative morphology . . . 174
6.7 Summary of the chapter . . . 175

7 Gender and grammatical complexity 177
7.1 Introduction . . . 177
7.2 Background . . . 177
7.3 How to integrate interactions of gender into a metric for gender complexity 178
7.4 How to elaborate maximally local complexity measures . . . 180
7.5 A metric for grammatical gender complexity . . . 181
7.6 Method . . . 184
7.7 Results and discussion . . . 186
7.7.1 Genealogical and areal biases in the distribution of GCSs . . . 190
7.7.2 Same complexity score does not mean same type of gender system 193
7.7.3 Implicational relationships between the features in the metric . . . 193
7.7.4 Some features may be stronger predictors of gender complexity than others . . . 194
7.8 Summary of the chapter . . . 198

8 Summary of the main results, prospects for future research and concluding remarks 199
8.1 Summary and assessment of results . . . 199
8.1.1 Gender and number . . . 200
8.1.2 Gender and evaluative morphology . . . 201
8.1.3 Interactions of gender and grammatical complexity . . . 202
8.2 Assessment of the sampling methodology . . . 202
8.3 Prospects for future research . . . 204
8.3.1 Nominal relevance hierarchy . . . 204
8.3.2 Split plural indexation systems and Dedicated Plural Indexation . 205
8.3.3 Gender assignment and its manipulation . . . 205
8.3.4 Absolute complexity of gender systems . . . 206

8.4 Concluding remarks 207

A The language sample **209**
A.1 Genealogical units and internal composition of the subsamples 209
A.2 The language sample: alphabetical index 215

B Database coding sheet **219**
B.1 Language data 219
B.2 Gender systems 219
B.3 Number systems 221
B.4 Evaluative morphology 221
B.5 Gender complexity 223

C Gender- and number-indexing targets in the languages of the sample **225**

D Examples of lexical plurals in four Cushitic languages **227**

E Singular and plural suffixes in Mabaan, Dinka and Luwo **229**

F Complexity scores for the individual features in the metric **231**

Sammanfattning på svenska **235**

Bibliography **243**

List of Figures

2.1 NP-internal and NP-external indexation 9
2.2 The gender system of Romanian 12
2.3 The grammaticalization of gender from demonstratives 16
2.4 Universal semantic and pragmatics properties of diminutives 27
2.5 Possible patterns of syncretism 36

3.1 Greenberg's (1963a) genealogical classification of the African languages . . 48
3.2 African language families and isolates 51
3.3 Linguistic macro-areas within Africa 54

4.1 Sex-based and non-sex-based gender systems in the languages of the sample 72
4.2 General and obligatory number in the languages of the sample 86
4.3 Types of evaluative morphology systems in the languages of the sample . 94

5.1 Cumulation and syncretism 96
5.2 The noun class system of Wamey 104
5.3 Types of gender syncretism across number 110
5.4 Triggers of DPI in the Cushitic sample 126
5.5 DPI in the Cushitic sample 127
5.6 Gender- and number-indexing targets in the languages of the sample . . . 133

7.1 Distribution of the GCSs 188
7.2 Geographical distribution of the GCSs 189
7.3 Correlation coefficients between the features of the metric 195
7.4 GCSs (Average) stratified according to feature values 197

List of Tables

3.1 Genealogical relationships within Khoisan 49
3.2 Genealogical units in the sample 57

4.1 Sex-based and non-sex-based gender systems in the language sample . . . 63
4.2 Number of genders in the language sample 65
4.3 Systems of gender assignment in the languages of the sample 67
4.4 Number of gender-indexing targets in the language sample 68
4.5 Overt coding of gender on nouns in the languages of the sample 70
4.6 Noun classes and indexation patterns in Kirundi 77
4.7 Noun classes and indexation patterns in Bandial 79
4.8 Nominal number systems in the languages of the sample 82
4.9 Number values in the languages of the sample 83
4.10 Number of number-indexing targets in the languages of the sample 85
4.11 Singular marking in Kambaata, Borana-Arsi-Guji Oromo and Tsamai . . 88
4.12 Plural marking in Kambaata, Borana-Arsi-Guji Oromo and Tsamai 88
4.13 Evaluative morphology systems in the languages of the sample 92

5.1 Cumulation between gender and number on the indexing targets 99
5.2 The semantics of the noun class system of Ju|'hoan 101
5.3 Independent Personal Pronouns in Kabyle 102
5.4 The Definite Article in Beja 103
5.5 Cumulation between gender and number on nouns 106
5.6 Gender syncretism and its relationship with cumulation in the language sample 111
5.7 Gender syncretism in Turkana restrictive markers 112
5.8 Gender and number distinctions in Gola 114
5.9 Gender and number distinctions in Nuclear Wolof 114
5.10 Morphologically plural nouns taking plural indexation in Baiso 124
5.11 Lexical plurals in Baiso 125
5.12 Number-indexing targets in Creissels et al.'s (2008) type (a) languages . . 131
5.13 Distribution of the use of gender shifts for the encoding of variation in the countability properties of nouns 135
5.14 Noun classes and singular-plural pairs of Swahili 137
5.15 Noun classes and singular-plural pairs of Bandial 137

6.1 Interactions between grammatical gender and evaluative morphology . . . 146
6.2 Distribution of diminutive and augmentative genders in Type 1 languages 149

6.3 The diminutive suffix in some southeastern Bantu languages 155
6.4 Number of gender distinctions and type of attested gender shifts in Type 2 languages 162
6.5 The meanings and function of *-hadi* in Sotho 171

7.1 Cues for assessing grammatical complexity of gender systems 183
7.2 Gender complexity metric 185
7.3 GCSs of the languages of the sample 187

A.1 The language sample 210
A.2 The language sample: alphabetical index 215

C.1 Number of gender- and number-indexing targets in the languages of the sample 225

D.1 Lexical plurals in Iraqw 227
D.2 Lexical plurals in Borana-Arsi-Guji Oromo 228
D.3 Lexical plurals in Rendille 228
D.4 Lexical plurals in Baiso 228

E.1 Singular markers in Mabaan, Dinka and Luwo 229
E.2 Plural markers in Mabaan, Dinka and Luwo 230

F.1 Complexity scores 231

Abbreviations

The glossing of the examples conforms to the Leipzig Glossing Rules: `http://www.eva.mpg.de/lingua/resources/glossing-rules.php`. In case of conflict between the Leipzig Glossing Rules and the abbreviations used in my sources, nonstandard abbreviations were substituted with the corresponding standard abbreviations in the Leipzig Glossing Rules. The glossing of the examples was adapted accordingly. The orthography used in the examples is the same as the one used in the individual sources. Tones are thus only marked if marked in my sources.

2	second person
3	third person
ABL	ablative
ABS	absolutive
ACC	accusative
ADJ	adjective
AUG	augmentative
C	common gender
CL	noun class
COP	copula
CRD	coordination
DAT	dative
DECL	declarative
DEF	definite
DEM	demonstrative
DET	determiner
DIM	diminutive

Abbreviations

DS	different subject
DUP	reduplication
F	feminine
GEN	genitive
GENERAL	general number
ICP	instrumental-comitative-perlative
INDF	indefinite
M	masculine
N	neuter
NEG	negation, negative
NOM	nominative
OBJ	object
PASS	passive
PAUC	paucal
PCO	perfective converb
PERM	permissive
PFV	perfective
PL	plural
POSS	possessive
PRF	perfect
PRS	present
PST	past
RECP	reciprocal
RED	reduplication
REL	relative

RP	recent past
SBJ	subject
SG	singular
SUBORD	subordination marker
UNM	unmarked
VOC	vocative

Acknowledgements

The path towards the completion of this dissertation is populated by people, places, random extraordinary events and ordinary daily routines that have inspired me in various ways and contributed to my growth as a linguist and as a human being. It is to this beautiful crowd of individuals and life experiences that I would like to say "Thank you!" today.

First of all, my deepest gratitude goes to the Department of Linguistics at Stockholm University for welcoming me as a PhD student and for giving me the opportunity to become part of the most friendly and open-minded environment that a researcher could wish to live in.

My supervisors, Maria Koptjevskaja Tamm and Ljuba Veselinova, provided me with rigorous guidance and friendly support from the first day to the last. I thank them for letting me work independently throughout my doctoral studies, while, at the same time, always being there to express the most thorough and thought-provoking criticism as well as the most sincere enthusiasm towards the progress of my work. Our relationship during these four years has been based on a perfect combination of mutual trust, respect, empathy and a good dose of humour, all of which have contributed to create an ideal atmosphere for me to work through during the ups and downs of my life as a PhD student. Thank you also, Masha and Ljuba, for being such a great source of inspiration on how to combine professional fulfillment with the challenges of being a woman and a mother.

A preliminary draft of this dissertation was read and examined by Östen Dahl in the spring of 2014. His sharp comments, suggestions and insights have been precious for improving and finalising my work. Matti Miestamo, Micke Parkvall and Bernhard Wälchli contributed invaluably to my chapter on complexity by helping me designing the study and reading the first drafts of the chapter. Raphaël Domange, Thomas Hörberg, Robert Östling, and Bernhard Wälchli patiently helped me with the statistical analysis of my data on complexity.

I am very grateful to Eva Lindström for translating the summary of the dissertation into Swedish and to Lamont Antieau for his meticulous proofreading. Sofia Gustafson Capková patiently helped me find my way through the six-month procedure that precedes the doctoral defence. Ljuba Veselinova managed to find some extra time to design the beautiful maps that accompany the thesis text. Paolo Cerva designed the perfect cover for this dissertation.

Many people helped me with data collection and answered endless questions on their languages or theoretical domains of expertise. First of all, my deepest gratitude goes to the members of the French research institute LLACAN (*Langage, Langues et Cultures d'Afrique Noire*) that hosted me for two weeks in the spring of 2012. I am particularly

grateful to Dmtry Idiatov, Maria Khachaturyan (Mande), Stefano Manfredi (Standard Arabic), Amina Mettouchi (Berber), Konstantin Pozdniakov (Atlantic), Nicolas Quint (Kordofonian), Guillaume Segerer (Atlantic), Yvonne Treis (Cushitic), Mark Van de Velde (Bantu) and Martine Vanhove (Beja) for giving me the opportunity to discuss my work with them and for sharing with me their knowledge, resources and fieldwork data. I am also very grateful to Clement Appah (Akan), Nana Aba Amfo (Akan), Tom Güldemann (Khoisan), Bernd Heine (Khoisan), Serge Sagna (Bandial), Graziano Savà (Cushitic and Omotic), Anne Storch (Western Nilotic), and Suzanne van de Meer (Cushitic) for their generosity in sharing data on their languages of expertise. Desalegn Hagos Asfawwesen answered a long list of questions on Amharic, his mother tongue, and Yvonne Agbetsoamedo helped me with Ewe, her mother tongue, and Sɛlɛɛ. Thank you also, Yvonne, for the great honour of working together on temperature terms and evaluative morphology in Sɛlɛɛ. The pioneering work on evaluative morphology conducted by Nicola Grandi and Livia Körtvélyessy greatly inspired my own research. I thank them for the always interesting discussions we had as well as for inviting Yvonne (Agbetsoamedo) and me to participate in their project on the *Handbook of Evaluative Morphology*. I am also very grateful to Jenny Audring for her inspirational work on gender as a complex feature of grammar and for the numerous email exchanges we had. Thanks to Harald Hammarström, Suzanne van de Meer and Hedvig Skirgård for incorporating and adapting some of the features of my database into their questionnaire for the *Nijmegen Typological Survey*.

This dissertation could have never been accomplished without the support and generosity of the people mentioned above. Shortcomings and mistakes in the analysis and interpretation of the data are, of course, entirely mine.

I am immensely grateful to all my colleagues within the General Linguistics session at the Department of Linguistics at Stockholm University, for an exceptionally inspiring working environment, and for the always cheerful and intellectually stimulating chats during our *måndagsfika*, when I would try to practise my Swedish while listening to the most interesting stories about Swedish society, history and culture. I would not have made it through these four years without the solidarity of the other doctoral students in linguistics within and outside the department: Andrea (Kiso), Benjamin (Brosig), Calle (Börstell), Clelia (La Monica), Desalegn (Hagos Asfawwesen), Emil (Perder), Franco (Pauletto), Geraldine (Quartararo), Ghazaleh (Vafaeian), Hatice (Zora), Kerstin (Lindmark), Lena (Renner), Natalia (Perkova), Pernilla (Hallonsten Halling), Robert (Östling), Susanne (Vejdemo), Thomas (Hörberg), Yvonne (Agbetsoamedo) and all the others. Thank you guys for our countless lunch breaks in the gloomy darkeness and the bright sunshine of the Swedish winters and summers, for the many adventures we experienced together, the joyful moments and the stressful ones. Robert, Yvonne, Kerstin, and, although very briefly, Miina (Norvik) and Gintaré (Grigonyté), you have been the perfect office mates during these years. Thank you for tolerating my mess and for brightening up all my days at work. Finally, I would like to thank Nada Djokic, Linda Habermann, Cilla Nilsson and Rickard Franzén for their patience, efficiency and kindness and, most importantly, because, no matter what they were doing and how busy they might have been, they always welcomed me with a smile.

The story that brought me to Stockholm began between the spring and summer of 2009, in Albuquerque (New Mexico) where I studied for five months under the supervision of Bill Croft. Thank you, Bill, for giving me the first hint that interactions of gender with other nominal features could be an interesting phenomenon to look at, as well as for letting me work on your collection of reference grammars during the sunny mornings of the high desert summer.

Luisa Brucale, already more than ten years ago you made me fall in love with linguistics. Giorgio Iemmolo, from the old good days in the courtyard of the Department of Classical Philology at the University of Palermo, you have been the best companion of linguistic adventures. Shelece Easterday, our deeply felt friendship goes beyond the ocean and the many time zones that separate us.

It is not easy to move to a new country as an adult and build a life there. I was lucky, because in Stockholm I met some of the most precious people of my life. Adeile, Fabio, Franco, Geraldine, you have been and are my beloved "Little Italy" in Stockholm. Eti and Chris, my weeks are brighter since you came back to Stockholm and brought Elsie to this world. Ljuba and Vera-Katlheen, the sun always shines in your company. Andrea, you are the first friend that I made in Sweden, and have become one of the closest. Stockholm is much more empty since you and Gladstone left. Marjatta, I arrived at your place on a cold Sunday afternoon, you welcomed me with *kanelbullar* and a cup of strong Swedish coffee, and I immediately felt at home. Your beautiful house soon became the most secure place that I know of within Sweden. A couple of days later, on the day of my 27th birthday, Yvonne, you also arrived in Stockholm. It didn't take long to realise that you are my third sister, born on the other side of the world and from different parents, but from whom I will never be apart again. Benjamin, the world would be much happier if everyone could have a friend like you, your generosity and altruism trascend obviousness, and this is what makes them so precious.

Much of who I am has come to be thanks to the friends I grew up with. Together we explored our hometown, our island and many places outside of it. Together we also learned to be apart, to make the best out of the always little time available, and to find our way through one of the most brutal political and economical crises that Italy has ever experienced. Anna, Carla, Dario, Diana, Francesca, Graziana, Germana Co. and Germana Civ., Giorgio, Lollo, Marco, Nadia, Paolo, La Seppia (Peppe Filippi), Rosi and Sara, thank you for all the years together, for your visits to Stockholm and for making Berlin, Milan, Paris, Reggio Emilia, Tunisi or Zurich "another" Palermo to visit and in which to find myself.

My sisters, Chiara and Agnese, and my brother-in-law, Cristiano, are fundamental pillars of my daily existence. I am very grateful to them for being there when I cannot be there, and for visiting me here every time they could. Ever since I was born, my parents, Vera and Enzo, have never left me alone. Their support has been unconditional without ever being oppressive. They taught me to always look at what's around me with humility and to never stop learning. Raphaël gives a sense of direction to all my days. He is home and future. From my grandmother, Lilla, I learned how to cook, sew, knit and water plants, but, most importantly, she taught me how to stand my ground and to never give up. Unfortunately, she passed away in December 2013 and could not see

the end of this work. I dedicate this dissertation to her and the memory of her beautiful rebellious eyes that always saw beyond appearance.

1 Introduction

Gender is one of the most extensively researched domains of grammar. In the typological literature on gender systems, a great number of studies have focussed on identifying and classifying the types of gender systems attested in the world's languages, mapping their distribution across genealogical units and geographical areas, and, to a lesser extent, attempting to model their origins and functions in grammar and discourse. This dissertation focusses on a less-explored aspect of the nature of gender systems: their interactions with other domains of grammar.

Gender is often claimed to be closely linked to other grammatical domains such as argument marking, definiteness, number or evaluative morphology. These connections can be both morphosyntactic and semantic in nature. For instance, gender distinctions can be realised together with case, definiteness or number distinctions; in Italian, e.g., gender values are always realized together with number values. Or, individual genders in a language can be associated with the encoding of diminutive and augmentative meanings; e.g., in Serbian-Croatian, the Neuter Gender is associated with diminutive meanings (Grandi 2001).[1] Accounting for these interactions can be very relevant to understanding how the gender system of a language functions synchronically, or how it developed diachronically. Even though phenomena of this type are likely to be at least mentioned in reference grammars, a systematic survey of the kinds of interactions attested between gender and other domains is still missing in the typological literature.

This dissertation investigates the morphosyntactic and semantic interactions of gender systems with number and evaluative morphology. This is done by combining crosslinguistic research and in-depth studies of genealogically related languages within one continental area, Africa. Diachronic and synchronic aspects of the interactions among the above-mentioned grammatical domains are considered. The notion of interaction is used in this dissertation in a rather loose way. In fact, in African languages in particular, gender and number, on the one hand, and gender and evaluative morphology, on the other, can be so closely intertwined with each other that they do not only interact but exhibit indications of actually being merged.

The dissertation is built on three main research foci:

(1) Interactions between gender and number

[1] In this dissertation, I use capital letters to refer to language-specific categories (e.g., the Neuter Gender in Serbian-Croatian) and lowercase letters to refer either to a specific marker within a language (e.g, the neuter marker in Serbian-Croatian) or to grammatical domains as objects of crosslinguistic comparison (e.g., the neuter genders in the languages of the sample). See Haspelmath (2007) for a discussion on this topic.

  - Do gender and number have the same encoding? How are gender distinctions mapped across number values?
  - Do gender and number compete through indexation patterns?
  - Do gender and number have the same relevance to nouns?
  - Do gender and number interact semantically?

(2) Interactions between gender and evaluative morphology:

  - Can *size* be a criterion for gender assignment?
  - How do diminutive and augmentative genders interact with the other gender distinctions of a language?

(3) Interactions of gender and grammatical complexity:

  - Is it possible to measure the grammatical complexity of gender systems?
  - Can interactions of gender with other domains of grammar be accounted for by such a complexity metric?
  - What is the role of these interactions in the overall complexity of a gender system?

The dissertation is organised as follows. In chapter 2, I introduce the three grammatical domains under investigation by looking at their functional and semantic properties, their structural encodings and typological distributions. The chapter also provides an overview of previous literature on the interaction between gender and number and gender and evaluative morphology as well as on grammatical complexity. The research questions addressed in this study, and already partially introduced here, are outlined at the end of chapter 2. In chapter 3, I describe the sampling methodology established for this study as well as the protocol for data collection and organization. Chapter 4 provides an overview of the gender, number and evaluative morphology systems attested in the languages of the sample. The results of the investigation are presented and discussed in chapter 5 (gender and number) and chapter 6 (gender and evaluative morphology). In chapter 7, I propose a metric for measuring the grammatical complexity of the gender systems attested in the language sample and show how morphosyntactic and semantic interactions with number and evaluative morphology can be accounted for by this metric. A summary of the results, prospects for future research and concluding remarks are presented in chapter 8.

2 Background: gender, number, evaluative morphology and relevant interactions

In this chapter, I define the grammatical domains investigated in the dissertation – gender, number and evaluative morphology. I also provide an overview of previous literature on the interactions between gender and number and gender and evaluative morphology, on the one hand, and the grammatical complexity of gender systems, on the other. The chapter thus consists of two parts. In the first part, gender, number and evaluative morphology are defined along four dimensions:

(1) Semantics and functions

(2) Morphosyntax

(3) Typological distributions

(4) Diachrony

The three domains investigated in the dissertation are each allotted a separate section, starting with gender (§2.1), continuing with number (§2.2) and concluding with evaluative morphology (§2.3).

Previous studies on the patterns of interaction between gender and number and gender and evaluative morphology as well as on the grammatical complexity of gender systems are discussed in the second part of the chapter (§§2.5, 2.6 and 2.7, respectively). In §2.8, an outline of the research questions addressed in the dissertation is provided. A summary of the chapter is presented in §2.9.

2.1 Gender

Gender is a type of noun classification strategy. The label *noun classification strategy* is conventionally used to refer to heterogeneous sets of grammatical constructions that are specialised in the categorisation of nouns (Aikhenvald 2003). Different types of noun classification strategies can be determined according to the following criteria: meaning, number of distinctions within the system, locus of marking, historical development and degree of grammaticalization. Following Corbett (1991, 2013a,b), I define gender as the particular type of noun classification strategy which must be reflected beyond the nouns themselves, that is, through indexation patterns or, following Corbett's terminology, "agreement" (see §2.1.2.1 for a critical reappraisal of the notion of agreement).

The notion of gender adopted in this dissertation conforms to that of the mainstream typological literature, in which gender and noun classes are viewed as the same grammatical phenomenon and the term gender is used as a hyperonym of the two. When describing language-specific types of gender systems, I use only the label *gender* for systems that are built on a limited number of distinctions (from two to three, e.g., animate vs. inanimate, or masculine vs. femine vs. neuter). Conversely, for richer systems (that make four or more distinctions) I use the term *noun class* to refer to individual singular and plural markers and the term *gender* to refer to the system as a whole and, particularly, to the pairings between singular and plural class markers (see also Aikhenvald 2003; Corbett 1991, 2013b).

In §2.1.1, I discuss gender assignment rules, that is, the dynamics according to which, in a given language, nouns are distributed across different genders. The morphosyntax of gender and patterns of gender indexation are discussed in §2.1.2. Finally, the function of gender, the distribution of gender systems across the world's languages and aspects of their historical development are discussed in §2.1.3.

2.1.1 Gender assignment

Nouns can be assigned to a given gender on the basis of a combination of semantic and formal (morphological and/or phonological) properties. All gender systems are in a sense mixed since "there is always a semantic core [...] but this is never the entire story" (Aikhenvald 2003: 22).[2]

The semantic underpinnings of the gender system of a language can be very blurred in actual use. Nonetheless, as shown by Corbett (1991) (see also Aikhenvald 2003; Dahl 2000a,b), it is always possible to recall the fundamental semantic notion(s) on which the categorisation is based, even if only for a restricted portion of the nominal lexicon of a language. Semantic gender assignment generally involves nouns denoting animate entities, with the cut-off points within the domain of animacy being highly language-specific: "between humans and animals, between higher and lower animals, or between animals and inanimates" (Dahl 2000a: 101). In many languages, for example, only nouns denoting sex-differentiable entities – that is, humans and animals – can be assigned to a gender on a semantic basis. The rest of the nominal lexicon is distributed in one gender or the other on the basis of semantically more arbitrary morphophonological patterns or mechanisms of semantic extension such as metaphorical or metonymic associations.

Animacy, sex, shape and size are the most common semantic notions upon which a gender system is based.[3] Their distribution across language-specific gender systems is not equal: sex-based systems are more frequent than other types of systems, which are generally built upon some notion of animacy (Corbett 2013c). According to the

[2] In Uduk (Koman), semantics seems to play no role in the allocation of nouns to one gender or the other (Don Killian, personal communication). The language is currently being described, but, if this preliminary analysis holds, the gender system of Uduk represents an interesting challenge for the current understanding of grammatical gender crosslinguistically.

[3] See Croft (1994) for a discussion of the significance of the notion of shape in the semantics of noun class systems.

data in WALS, in almost all geographic areas where gender is found, sex-based gender systems are also found: 84 out of the 112 gendered languages in Corbett's (2013c) sample have sex-based gender.[4] Moreover, on a general basis, animacy and sex are more frequent criteria of gender assignment than physical properties (shape and size). In many languages, "physical properties are only rarely employed to assign genders to animates"; their relevance as a criterion of gender assignment is restricted to inanimate entities (Aikhenvald 2003: 278). Finally, among the four most typical semantic underpinnings of gender, size seems to be the least likely to occur as an independent criterion for classification (see also §2.6).

In morphological gender assignment, individual word-formation strategies and/or inflectional classes tend to be associated with a particular gender. The nature of these associations is often problematic, and it is not obvious how or if one dimension (gender or morphological class) derives from the other. The regularity of these associations is largely language-specific, and, within languages, it can vary according to individual morphological patterns. In Italian, for example, the suffix for action nominalizations, *-zione*, is used to derive feminine nouns. Conversely, when the morphological criterion and the gender value that would be assigned to a noun according to its semantics are in conflict, morphology can be overridden by semantics. For example, the Italian noun *soprano* is morphologically equivalent to the noun *vaso* 'vase.' These nouns are grammatically masculine. However, since *soprano* typically denotes female singers, speakers tend to treat it as feminine.

The notion of phonological gender assignment is a tricky one. In most cases, it is in fact very difficult to tease it apart from that of morphological assignment. In order to distinguish between the two, Corbett (1991: 51) proposes the following rule: if the gender of a noun can be established by taking into consideration more than one form, we are dealing with morphological assignment; if the gender of a noun can be established based on one form only, we are dealing with an assignment rule that is also a phonological rule. The most frequently quoted language with phonological gender assignment is Qafar (Afro-Asiatic, Cushitic, East Cushitic) (Corbett 1991: 51-52). In Qafar, nouns whose citation form ends in an accented vowel are assigned to the Feminine Gender. The overlap between semantic and phonological properties of nouns is generally systematic. With a few animate nouns, phonology and semantics are in conflict; in such cases, semantics wins. Thus, the word for 'father' *abbà*, despite its phonology (it ends in an accented vowel), is assigned to the Masculine Gender.

Dahl (2000a) reassesses the dichotomy between semantic and formal dynamics of gender assignment by introducing a distinction between *lexical* and *referential* gender. In principle, both can be viewed as instances of semantic gender, even though in different ways and to different extents. Lexical gender refers to the denotation of nouns, that is, to semantic properties of nouns as lexical items; it may also be conditioned by formal properties that are characteristic of a noun as a lexeme belonging to a certain class with certain morphological properties. On the contrary, referential gender is concerned with

[4] The total number of languages in Corbett's sample is 257; out of this sample, 145 languages do not have gender.

semantic properties of the noun phrase (henceforth NP) referents in the extralinguistic world and is, as such, always semantic. The Russian word for 'judge' *sud'ja* is presented by Dahl (2000a) as an illustration of this polarity. Russian grammars treat the noun as masculine. This is in line with the general (semantic) pattern according to which underived nouns denoting professions are always masculine, with the exception being those nouns denoting professions that are stereotypically associated with women (*med-sestra*, '(medical) nurse,' from Dahl 2000a: 109). Conversely, *sud'ja* ends in *-a*, which is typically used as a feminine suffix. In actual spoken language, and when referring to a female judge, Russian speakers are conflicted between treating the noun as feminine *strogaja sud'ja* 'strict (fem.) judge'– thus based on referential semantics – or as masculine *strogij sud'ja* 'strict (masc.) judge' – thus based on the lexicalized, stereotypical association between the profession and the Masculine Gender. Given that the noun ends in *-a*, a prototypical feminine suffix, the use of masculine indexing forms associated with a female referent cannot be regarded as an instance of formal assignment. Following Dahl's analysis, this case can be rather viewed as a conflict between assignment criteria that are both semantic, even though in different ways, that is, lexically and referentially.

One last phenomenon to be mentioned in relationship with gender assignment is the existence of *non-rigid* gender assignment rules, whereby nouns are assigned to different genders depending on the context in which they occur. In his survey of gender systems of African languages, Heine (1982) introduces the distinction between *free* and *fixed* gender. Free gender systems are those that allow nouns "to change the gender affiliation" (Heine 1982: 198), while fixed gender systems are those in which nouns are rigidly and invariably assigned to a gender. Heine suggests that in free gender systems, ideally, any noun could be allocated to any gender. In reality, he notices that in the African languages with relatively free gender systems, semantic or morphological gender assignment preferences exist at least for some nouns. Corbett (2013a) refers to the possibility of multiple gender affiliation as *recategorization*, whereby speakers modify the construal of the NP referent by manipulating its gender assignment. The notion of *recategorization* was previously used by Corbett (2000, 2012) in relationship to nominal number to describe those cases in which speakers manipulate the countability properties of NPs by treating mass nouns as counts and counts as masses (see also Lyons 1968; Quirk et al. 1985). According to Corbett, typical instances of recategorization are the English utterances *I'd like three coffees, please*, and *There was dog all over the road* (Corbett 2000: 85). In the former case, the mass noun *coffee* is treated as countable and thus receives plural marking. In the latter case, the countable noun *dog* is construed as mass and thus cannot be pluralized.

Corbett (2013a) suggests that, when referring to gender systems, the most obvious type of recategorization is based on sex-differentiability. He mentions the case of Kupto, a West Chadic language spoken in Nigeria, in which animal nouns are usually treated as feminine. However, in storytelling, and if personified, certain animal nouns – e.g., the word for 'hyena,' 'elephant' or 'mouse' – can be treated as masculine. In languages with sex-based gender, inanimate nouns tend to be assigned to the feminine or the masculine gender depending on the way the NP referent is construed in the discourse (Corbett 2013a: 123). According to Corbett, recategorization through gender can also be based on

the notions of *size* and *value*. This happens in languages in which nouns can change their gender when speakers want to express diminutive and augmentative meanings – both in terms of size variation (*small* vs. *big*) and value (*good* vs. *bad*). To date, very little has been done to explore the crosslinguistic distribution of this phenomenon. Corbett (2013a) mentions a handful of languages where size and value-related recategorization is found – e.g., Lavukaleve (Lavukaleve), Maung (Iwaidjic), Savosavo (Savosavo), Walman (Nuclear Torricelli), Yawuru (Nyulnyulan). Similar phenomena are also investigated by Aikhenvald (2012) with a special focus on the languages of Papua New Guinea.

Recategorization phenomena related to gender are more commonly referred to in the literature as *gender shifts*. The latter label is adopted in this dissertation. I use the label *manipulable gender assignment* to refer to the possibility of assigning nouns to multiple genders ("free gender" according to Heine 1982) and the label *rigid gender assignment* ("fixed gender" according to Heine 1982) to refer to those languages where nouns are invariably assigned to one gender. Gender shifts and manipulation of gender assignment are explored in detail in §5.7, as well as in chapters 6 and 7.

2.1.2 The morphosyntax of gender: indexation

Hockett (1958: 231) defines gender as a grammatical category that is "reflected in the behavior of associated words." As a consequence, for the gender system of a language to be considered productive, gender needs to be cross-referenced by those elements in the utterance that entertain some kind of morphosyntactic and/or semantic relation with the noun itself or the referent of the NP (typically, demonstratives, determiners, pronouns, relative pronouns, adjectives, verbs, but also adpositions, complementizers and phrase markers). Cross-reference patterns are traditionally treated in the literature as *agreement phenomena*. However, following an already established tradition in typological literature (see, among others Croft 2001, 2003, 2013; Iemmolo 2011), in this dissertation the term *indexation* is used rather than *agreement*. The reasons behind this choice are briefly explained in §2.1.2.1.

2.1.2.1 Beyond agreement: the notion of indexation and what we gain with it

Typically, the term agreement refers to a particular type of asymmetric syntactic relationship between two entities within an utterance. This syntactic relation is such that one of the two entities takes an inflectional form that is determined by certain semantic or morphosyntactic properties of the other entity (typically gender, number, person). In his model of agreement, Corbett (1991, 2000, 2006) refers to the former entity as *target* and to the latter as *controller*. According to Corbett, the most typical or, to use his terminology, "canonical", instances of agreement need to comply with, among others, the following conditions (for a complete list of the properties of canonical agreement, see Corbett 2006: 9):

(a) The controller is *nominal* and *overtly expressed* in the discourse.

(b) The syntactic dependence between the controller and the target is *local* in the

> sense that, within a given utterance, the "structural distance" (Corbett 2006: 21) between controller and target is minimal.

Thus, the notion of agreement implies that "there is a phrase in the utterance that is 'agreed with' (the *controller*) that is necessarily present" (Croft 2003: 34). In fact, crosslinguistic evidence suggests that this is far from being the most common state of facts and that the whole notion of agreement can be very misleading. For instance, contrary to what is stated in (a), in many languages both pronouns and NP-internal targets do not necessarily presuppose the presence of a syntactic antecedent or a controller (for a useful discussion see Croft 2013). In addition, contrary to what is stated in (b), in the case of pronominal targets, even when an antecedent is overtly coded in the discourse, the anaphoric pronoun and the syntactic antecedent are usually not structurally close to each other (for a more detailed discussion see §2.1.2.2). Finally, to give an example that is even more closely related to the grammatical phenomena investigated in this thesis, in many languages in which number marking on nouns is not obligatory, information about the number construal of the NP referent can be marked throughout the discourse (e.g., on verbs or pronouns) despite there being no overt linguistic entity that can be interpreted as being agreed with.

In light of the examples discussed so far, I use the term *indexation* to refer to the grammatical strategies that languages use to signal (1) lexical and grammatical properties of nouns or (2) semantic properties of NP referents, *independently* of the presence of an overtly expressed syntactic antecedent/head noun (for a similar understanding of the notion of indexation, see also Croft 2001, 2003, 2013). More specifically, this dissertation investigates indexation patterns that are used to signal gender and number. Accordingly, I use the terms *indexing target* or, alternatively, *index* to refer to entities whose inflectional morphology is used to signal gender and number throughout the discourse (e.g., adjectives, determiners, verbs, pronouns, etc.). The term *syntactic antecedent* is used in cases of NP-external indexation to refer to the NP indexed by the pronominal target. Finally, the labels *trigger* or *indexation trigger* are used to refer to the entities that activate the use of a certain indexation pattern (i.e., pronouns and nouns) in a given discourse domain.

Indexation patterns can differ depending on the NP-internal or NP-external nature of the indexing targets. Figure 2.1 (courtesy of Östen Dahl) proposes one way of interpreting some major differences between NP-internal and NP-external indexing targets. As shown in figure 2.1, in NP-internal and NP-external indexation referents can be signalled by the indexing targets with or without the intermediation of a head noun or a syntactic antecedent, respectively. In addition, in NP-internal indexation the relationship between indexing targets and referent is mediated by the internal structure of the NP. The indexing target can signal properties of the referent either as lexically/grammatically realised on the head noun or *directly*, that is, without the intermediation of the head noun. Finally, the information to be indexed may flow from the NP to the head noun (as, for instance, in the case of number) and vice versa, from the head noun to the NP (as, for instance, in the case of gender). Similarly, in NP-external indexation, the relationship between referent and indexing target can be direct, if the syntactic antecedent is missing,

or indirect, if it is overtly expressed in the discourse.

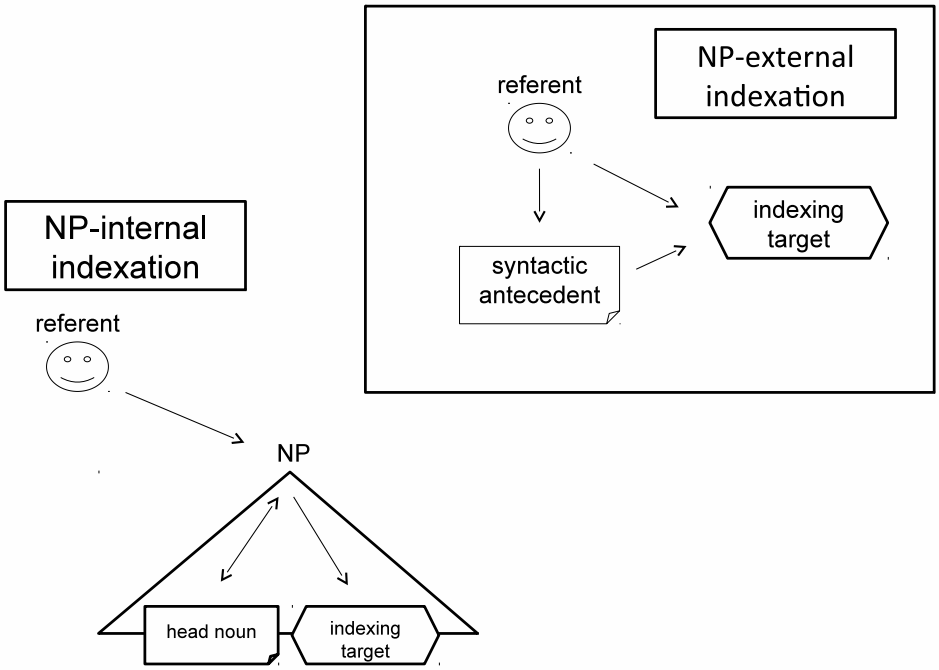


Figure 2.1: NP-Internal and NP-External indexation

In the literature on indexation and agreement phenomena, differences between NP-internal and external indexation have been modelled in the form of a typological hierarchy known as the Agreement Hierarchy. As Croft (2013: 112) puts it, the hierarchy attempts to describe "the degree of syntactic closeness or integration" between different indexing targets and the syntactic antecedent or the head noun. This is discussed in detail in §2.1.2.2.

2.1.2.2 Patterns of gender indexation

One – very often debated – problem in the literature on gender is how to account for those languages, such as English, in which the only evidence for gender distinctions appears on pronouns. In the literature on agreement, pronouns are often defined as non-prototypical agreement targets insofar as they "violate the expectation that agreement targets should share a local domain with their antecedent, preferably the phrase" (Audring 2009). However, in spite of their non-prototypical status, in the literature on (gender) agreement, pronouns are considered to be possible agreement targets (Audring 2009; Corbett 1991, 2006, 2012, 2013a). Within the indexation model introduced in §2.1.2.1, pronominal and NP-internal indexes are also part of one and the same functional domain in the sense that they all function as strategies for signalling reference through the discourse (on the

functional continuum between NP-internal and NP-external indexing strategies, see also Barlow 1992; Corbett 2006; Croft 2013; Siewierska 1999, 2004).

Based on these assumptions, which I also hold in this dissertation, languages like English are considered to be gendered languages, despite their gender system being less pervasive in discourse than gender systems in languages with richer indexation are. In this thesis, gender systems of the English type are singled out through the use of the label *pronominal gender systems*. Pronominal gender systems are crosslinguistically very rare[5] (Audring 2009; Corbett 2013b), and, as shown in the typological survey carried out by Audring (2009), they tend to pattern with strictly semantic principles of gender assignment. Applying Dahl's (2000a) dichotomy between lexical and referential gender, one could think of gender systems of the English type as being referential in nature. In languages with pronominal gender systems, gender indexation signals salient properties of the NP referents, e.g. male vs. female vs. sexually undifferentiated entities, rather than aspects of the lexical semantics of nouns.

Examining gender-related indexation patterns within a language is crucial for establishing the number of genders in that language. On the basis of the observation of independent patterns of indexation, indexation classes (or, in Corbett's terms, "agreement classes") can be established. These are defined as sets of nouns whose indexing targets have the same morphological realisations "under all conditions, that is, if we hold constant the values of other features such as case and number" (Corbett 2013a). Examples (2.1) and (2.2) illustrate the indexation patterns associated with two sets of Italian nouns, one set belonging to the Masculine Gender and the other to the Feminine.

(2.1) The masculine indexation class in Italian (Indo-European, Romance) (constructed example)

(a) *un-o* *zio* *fantastic-o*
INDF-M.SG uncle.M.SG fantastic-M.SG
'a fantastic uncle'

(b) *un* *dolce* *delizios-o*
INDF.M.SG dessertM.SG delicious-M.SG
'a delicious dessert'

(2.2) The feminine indexation class in Italian (Indo-European, Romance) (constructed example)

(a) *un-a* *zia* *fantastic-a*
INDF-F.SG aunt.F.SG fantastic-F.SG
'a fantastic aunt'

(b) *un-a* *nuov-a* *chiave*
INDF-F.SG new-F.SG key.F.SG
'a new key'

[5]The typological distribution of pronominal gender systems is rather striking. In principle, gender systems of the English type "would appear to be the most natural and well-motivated ones," and, also, the easiest to learn (Östen Dahl, personal communication).

Both nouns in example (2.1) trigger the same indexation pattern, a pattern that differs from that of the nouns in (2.2). The two noun sets belong to two independent indexation classes and thus to two different grammatical genders, the Masculine (because many nouns denoting males are assigned to this gender) and the Feminine (because many nouns denoting females are assigned to this gender). Italian is often described in the literature as a language with *overt gender*, that is, a language in which the gender of a noun can be inferred from the morphological appearance of the noun itself. Even though gender assignment in Italian can be often predicted on the basis of the final vowels of nouns, this is true only to a certain extent. The nouns *dolce* and *chiave* in examples (2.1b) and (2.2b), for instance, have the same final vowel and share the same type of number marking – that is, they both end in *-e* when singular and *-i* when plural. However, since they trigger different indexation patterns, they do not belong to the same gender. This is not an exceptional pattern in the language: the *-e/-i* class is a very frequent noun class in Italian, and the nouns belonging to this class can be both feminine and masculine (see Gudmundson 2012 for token frequency counts based on the corpus of spoken Italian LIP, *Lessico di frequenza dell'italiano parlato* (De Mauro et al. 1993)). In sum, examples (2.1b) and (2.2b) show that the covert/overt dichotomy is not a very powerful criterion by which to classify gender systems and that the genders of a language can be identified only by identifying indexation classes.

A less clear-cut, and very much debated, case in the literature on gender indexation is Romanian (Indo-European, Romance). There are three indexation classes in the language, traditionally labelled as *Masculine*, *Feminine* and *Neuter*. These are illustrated in example (2.3), where, for each gender, both singular and plural indexation patterns are given. Notice that the examples show gender indexation on adjectives only. The definite clitics also index the gender of the noun, but, for the sake of simplicity, this is not accounted for in the examples. The indexation patterns illustrated in the examples are summarized in figure 2.2.

(2.3) Gender indexation in Romanian (Indo-European, Romance) (adapted from Corbett 1991: 150)

(a) Masculine Singular

bărbatul e bun
man.the is good.M.SG
'The man is good'

(b) Masculine Plural

bărbaț̦ii sî̂nt bun-i
men.the are good-M.PL
'The men are good'

(c) Feminine Singular

fata e bun-ă
girl.the is good-F.SG
'The girl is good'

(d) Feminine Plural

fetele *sî̂nt* *bun-e*
girls.the are good-F.PL
'The girls are good'

(e) Neuter Singular

scaunul *e* *bun*
chair.the is good.N.SG
'The chair is good'

(f) Neuter Plural

scaunele *sî̂nt* *bune*
chairs.the are good-N.PL
'The chairs are good'

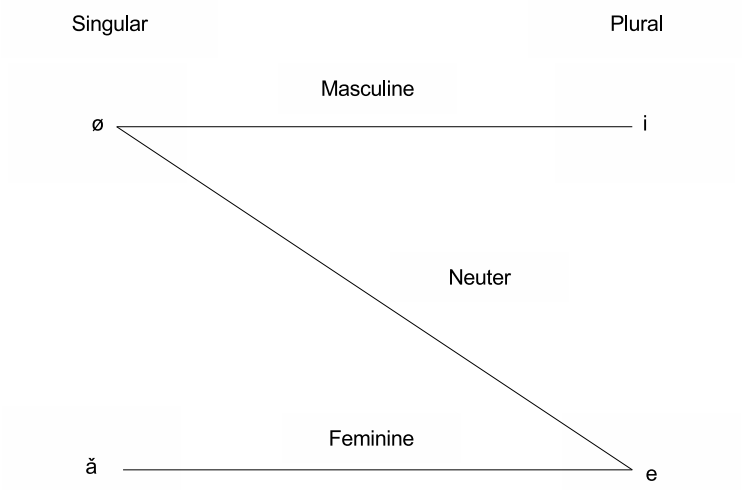


Figure 2.2: The gender system of Romanian (adapted from Corbett 1991: 152)

As shown in example (2.3) and figure 2.2, in Romanian, nouns belonging to the Neuter Gender do not have indexation forms that are uniquely associated with them. On the contrary, they share one of their indexation patterns with nouns assigned to the Masculine Gender and the other with nouns assigned to the Feminine Gender. Since the language has three distinct indexation classes (*ø/-i*, *-ă/-e* and *ø/-e*), it has three genders. For a more detailed account of the gender system of Romanian, see Corbett (1991: 150-154).

In his detailed overview of patterns of gender indexation, Corbett (1991) discusses the case of the *hybrid* nouns, that is, nouns that tend to take indexation for more than one gender. The choice among different genders is constrained by the type of target the nouns are indexed by and follows the predictions made by the Agreement Hierarchy (Corbett 1991, 2006):

attributive > predicate > relative pronoun > personal pronoun

According to Corbett, the configuration of the hierarchy is such that the more rightwards one moves along the hierarchy, the more likely it is for semantic indexation (he uses the term agreement) to override syntactic indexation. Thus non-prototypical indexing targets, such as the personal pronouns, are those that are more likely to trigger indexation based on semantic properties of nouns (see discussion at the beginning of the section). Indexation is defined by Corbett as "semantic" when it is consistent with the biological gender of the NP referent or other semantic properties of the noun; on the other hand, indexation is defined as "syntactic" when it is consistent with the gender as assigned by virtue of morphological and/or phonological rules. One of the most quoted examples of hybrid nouns is the German noun *Mädchen* 'girl', which is formally assigned to the Neuter Gender insofar as it is morphologically marked as diminutive (in German, nouns marked by the diminutive suffix are assigned to the Neuter Gender). The example below shows how targets that rank differently on the hierarchy manifest different types of indexation according to the above-mentioned predictions.

(2.4) German (Indo-European, Germanic) (Corbett 1991: 228)

Schau	*dir*	*dieses*	*Mädchen*	*an,*	*wie*	*gut*	*sie/es*	*Tennis*	*spielt*
look	you	this.N	girl	at,	how	good	she/it	tennis	plays

'Look at this girl, see how well she plays tennis'

Gender indexation with the demonstrative can only be syntactic – *dieses* – whereas with the personal pronoun, speakers can choose between semantic and syntactic indexation – *sie* or *es*.

Following Dahl's (2000a) distinction between *lexical* and *referential* gender (see §2.1.1), conflicts between "semantic" and "syntactic" indexation can instead be seen as conflicts between referential and lexical gender. Gender indexation is referential when it indexes extralinguistic properties of the NP referent, and it is lexical when it indexes aspects of the denotations of nouns viewed as lexical items. These notions, however we label them, play a crucial role in the understanding of some of the grammatical phenomena investigated in this dissertation (see chapter 5 in particular). When accounting for conflicts between possible indexation patterns in this study, the terminology proposed by Dahl (2000a) will be used.

I shall come back to the Agreement Hierarchy in §2.2.4, where similar phenomena are discussed in connection with number.

2.1.3 The function, distribution and diachrony of gender systems

Gender is a highly grammaticalized phenomenon in language: it presupposes rather long evolutionary chains and is generally very stable in the history of language families (Dahl 2004). Given these attributes, at least three questions concerning gender are worthy of discussion:

1. What is the function of gender in grammar and discourse?
2. How are gender systems distributed among the world's languages?
3. How do gender systems arise?

The rest of this section is devoted to the discussion of how answers to these questions have been sought in previous research.

2.1.3.1 The function of gender

When discussing the function of gender in language, scholars usually align themselves with one of the following assumptions:

1. Gender is a useless feature in human communication, and its appearance in grammar is the result of chance (due to grammaticalization) rather than functional necessity (McWhorter 2001; Trudgill 1999).
2. Gender systems are functional systems; their functions are mostly connected with reference tracking and disambiguation (Foley & Van Valin 1984; Greenberg 1978).

A new insight into the understanding of the function of gender has recently been provided by Dahl (2004). According to this view, the idea that gender is devoid of any communicative function does not mesh with its diachronic stability and the frequency of its use in gendered languages. Conversely, viewing gender only as a grammatical device for reference tracking and disambiguation is, according to Dahl, not completely satisfactory either, since this explanation does not take into account the whole range of occurrences of gender in a language (e.g., gender indexation internal to the NP). Rather, reference tracking and disambiguation "can be probably seen as one of several functions of grammatical gender that are all part of a larger scheme of redundancy management" (2004: 202). According to Dahl, gender systems synchronically function similarly to *checksum digit* systems,[6] that is, as devices for error checking in the online process of lexical item selection by speakers. In a language such as French, in which grammatical gender is marked on the article, speakers "know that a masculine article has to go with a masculine noun," and any other combination is perceived as a signal that "something has gone wrong" (Dahl 2004: 202).

[6] Checksum digits are added to bank account and card numbers as a security control device. They are calculated on the basis of various algorithms. For instance, the checksum digit of a code could be the last digit of the sum of its individual digits.

Dahl's hypothesis of gender as a grammatical device for online error checking provides – I believe – an insightful synchronic explanation for the function of gender in language. His hypothesis applies to manifestations of gender in grammar not only at the level of reference tracking and disambiguation but also in terms of indexation patterns internal to the NP. However, this still leaves the puzzle of the function of gender at the level of diachronic explanation unresolved at this point. The issue is discussed in detail in §2.1.3.3.

2.1.3.2 The distribution and stability of gender systems

According to the data in WALS, out of 256 languages, 112 languages, or less than half of the language sample used by Corbett (2013c), have gender. Gendered languages tend to cluster both genealogically and areally – and the occurrence of gender goes hand in hand with the presence of rich morphology. Nichols (1992) describes the genealogically and areally skewed distribution of gender systems as one of their most striking characteristics worldwide. In this respect, she introduces the notion of *class* (i.e., gender) *hotbeds* and *class outliers*. *Hotbeds* represent "areas in which most languages have classes, classes are found in languages of more than one family, and the formal implementation of classes [...] takes more than one form" (Nichols 1992: 130). Conversely, *outliers* are those languages with grammatical gender that are located outside of hotbeds. An example of a class/gender hotbed would be Africa, and an example of class/gender outlier would be the isolate Burushaski (Nichols 1992: 130-131).

Gender is also one of the longest standing features in the history of language families. For instance, within Indo-European, Armenian is the only primary branch in which the gender system of the protolanguage is not reflected by any of the member languages (Dahl 2004: 199). The diachronic stability of gender is directly proportional to areal and genealogical entrenchment. Gender systems tend to be unceasingly inherited but only rarely borrowed. The development of gender is the result of long evolutionary chains whose stages can only be partially detected based on empirical data (see §2.1.3.3). In languages with gender, gender distinctions are very much entrenched in the native speaker's usage. On the other hand, reduction and loss of gender systems by languages appear to be correlated with degree of contact and of adult second language learning (Dahl 2004; Trudgill 1999). This will be made clearer in chapter 7, where the impact of language contact on the complexity of gender systems is discussed more at length.

2.1.3.3 The diachrony of gender

Given that grammatical gender generally presupposes long evolutionary chains, in most cases, it is very difficult to establish how the gender system of a particular language or language family originated. A typical example is the debate over the diachrony of gender in the Indo-European languages: while many hypotheses and theories have been formulated over the years, the issue still remains unresolved.[7] With respect to systems

[7] In this respect, see Matasovič (2004) and Luraghi (2011), two recent contributions that combine comparative reconstruction with crosslinguistic evidence in formulating hypotheses on the origin of

that have emerged relatively recently, more can be said about the development of gender. One of the most relevant and quoted pieces of research on the origin of gender markers is Greenberg (1978). In his paper, Greenberg analyses patterns of grammaticalization involving gender in different branches of the Niger-Congo family and elsewhere in the world as part of a three-stage process. The entire grammaticalization chain is illustrated in figure 2.3.

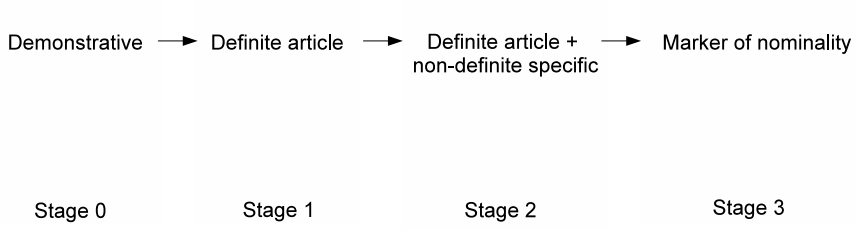


Figure 2.3: The grammaticalization of gender from demonstratives, according to Greenberg (1978)

As shown in the figure, demonstratives are interpreted by Greenberg as the ultimate diachronic source of grammatical gender in the languages of his sample. Demonstratives develop as definite articles when they begin being used to refer to entities that were previously mentioned in the discourse and become obligatory in order to mark a referent as having been identified (Stage 1). The next step in the grammaticalization chain (Stage 2) is taken when the article begins being used with specific but non-identified referents as well. At Stage 2, the absence of the article is restricted to generic contexts only (as in the English sentence *I like flowers* as opposed to *I like the flowers*). An additional step forward is made when one of the two options – usually the co-occurrence of articles and nouns – is generalized to all contexts. At this stage (Stage 3), the article has become a general marker of nominality and the grammaticalization chain has reached its final step of development. The entire process is led, according to Greenberg, by the need for "referential devices of identification" (1978: 78) or, in other words, by the need to track NP referents throughout a given stretch of discourse. Demonstratives play a very central role within this scenario since they constantly foster indexation (Greenberg 1978: 77).[8] A similar pattern of grammaticalization has been proposed by Heine (1997) and Kilian-Hatz & Heine (2010) for the gender system of the Khoe languages. In this case, the grammaticalization chain starts with third person pronouns which evolve as markers of definiteness and eventually as markers of nominality, that is, as bona fide gender markers.

Yet, a crucial question remains unanswered: how do markers of gender become attached to demonstratives (and/or third person pronouns) in the first place? The suggestion made in the literature (Corbett 1991; Dahl 2004; Greenberg 1978) is that certain

gender in the Indo-European language family.

[8] For an overview of demonstratives as a source of grammaticalization of a wide range of grammatical phenomena besides gender, see Diessel (1999a,b).

nouns such as 'person,' 'thing,' 'female,' and 'male' may acquire a classificatory function and then start being regularly used with demonstratives, to the point of fusing with them and thus result in gender marking. This can eventually spread further, for example, to anaphoric pronouns and generate indexation (Dahl 2004: 199). Nouns, then, seem to be the ultimate source of gender systems of this type.

An interesting case of a gender system originating from classificatory nouns is the Australian language Ngang'ityemeri (Australian, Daly, Southern Daly), which has been extensively investigated by Reid (1990, 1997). In Ngan'gytyemeri, multiple strategies of gender marking are available:

- (generic noun)[9] + specific noun + (generic noun) modifiers
- proclitic=Noun + proclitic=Modifiers
- prefix-Noun + proclitic=Modifiers

Diachronically, this system seems to have generated from the use of combinations of generic and specific nouns as a frequent structure of NPs and from the use of generic nouns with noun modifiers as a strategy for reference tracking. Over time, increased frequency of use resulted in morphological reduction and phonological attrition to the extent that certain generic nouns became proclitic words and, in some cases, grammaticalized as prefixes. Synchronically, the distribution of the different strategies of class marking is skewed according to semantics. For example, noun class marking with nouns denoting males, females and human groups is always performed by means of obligatory proclitic words marked on head nouns and noun modifiers. The individual stages of the grammaticalization path from classifier nouns to gender markers are synchronically instantiated by the multiple strategies of class marking listed above. Interestingly, there seems to be a synchronic correlation between the size of a noun class and the degree of grammaticalization of its marker: the larger the number of nouns assigned to a class, the more likely it is for the marker of that class to be a bound morpheme and to trigger indexation on noun modifiers (Reid 1997).

Contrary to Greenberg and Corbett's hypothesis, whereby classificatory nouns are viewed as the ultimate sources of gender, Nichols (1992: 139-142) proposes that in the emergence of gender systems, indexation precedes actual noun categorization processes. Indexation arises as a result of the grammaticalization of hidden – in the sense of not morphosyntactically coded – animacy distinctions or other types of cognitively salient distinctions. Once indexation surfaces, noun classes follow.

Whichever evolutionary pattern is hypothesised for individual gender systems, reconstructing the very early stages of development of gender is in many cases very critical. The notion of *exaptation* (Lass 1990) or, to use Croft's (2000) terminology, *hypoanalysis*,[10] becomes thus useful for explaining the diachronic transitions from whatever functions the sources of gender had at the beginning of the grammaticalization chain - or at

[9] The parentheses indicate optionality.

[10] Both terms refer to processes of functional reanalysis of grammatical constructions. These occur when contextual properties of a construction are interpreted by language users as its inherent syntactic properties, and the unit as a whole acquires new meanings and functions (Croft 2000: 127).

any stage along it - to the functions that full-fledged gender systems display synchronically. In sum, synchronic and diachronic explanations might be hard to relate to each other in the case of longstanding grammaticalization chains as those presupposed by gender systems.

For an overview of paths of grammaticalization of gender and noun class systems across languages of the world, see Aikhenvald (2003: 352-412), in which the emergence of gender in grammar is analysed within a more general discussion of the diachrony of noun classification devices.

2.2 Number

Nominal number is one of the strategies available in human languages to quantify entities (linguistically encoded as nouns or pronouns). I use the label *nominal number* in opposition to *verbal number*. The latter term is used to refer to strategies available in the verbal domain in order to quantify events in terms of the number of times they occur or the number of participants that are involved in them (Veselinova 2013).

Nominal number distinctions align with the following possible semantic cores: one, more than *one*, *two*, *three*, *four*, *a few*, *many*. Conventionally, the grammatical correspondents of such semantic notions are: *singular*, *plural*, *dual*, *trial*, *quadral*, *paucal* and *greater plural*, respectively. Nominal number distinctions can be encoded directly on the nouns by means of affixation, via indexation or by means of plural words. The marking of number on nouns or via indexation is discussed in details in §2.2.3 and §2.2.4. Plural words are especially common among languages that lack inflectional number marking (e.g., in the languages of Southeast Asia). The categorial status of plural words varies substantially across languages: they can be articles or numerals, but they can also constitute a word class of their own or pattern together with other modifying words, most often used to express size and evaluation. To date, the most relevant crosslinguistic survey of plural words is that of Dryer (1989b).

Various types of number systems are found in the languages of the world, and these systems depend on: (1) the interaction between the lexicon and grammar in the encoding of number distinctions, (2) whether or not number marking is obligatory for all nouns and (3) the internal distribution of number values within a language-specific number system. In the next sections, I discuss each of these points in detail. In §2.2.1, I begin with an overview of the countability properties of nouns and noun phrases. Major types of number systems and number values in the languages of the world are discussed in §2.2.2. In §2.2.3, I discuss the interaction between number and animacy; this is followed by an overview of the morphosyntax of nominal number in §2.2.4. Finally, the distribution and patterns of grammaticalization of nominal number are examined in §2.2.5.

2.2.1 Countability properties of nouns and noun phrases

Information about the quantificational properties of nouns can be expressed both grammatically and lexically. The quantificational properties of nouns at the level of lexical semantics are referred to in the literature as *countability properties* and are generally

treated as a manifestation of the polarity between *count* and *mass* (see, among others, Corbett 2000; Koptjevskaja-Tamm 2004). Nouns can be count or mass depending on whether they denote "discrete entities with a well-defined shape and precise limits" or "homogeneous undifferentiated stuff without any certain shape or precise limit" (Koptjevskaja-Tamm 2004: 1067). The same noun can be used both countably or uncountably depending on the context of its occurrence. This is illustrated in example (2.5); the English noun *table* is uncountable in (2.5a) and countable in (2.5b).

(2.5) Allan (1980: 547)

(a) *There's not enough table for everyone to sit at.*

(b) *We need a bigger table.*

Based on this evidence, Allan (1980) claims that countability properties are a characteristic of NPs rather than of nouns as individual lexical items (see also the discussion in Corbett 2000: 84-85.). Yet, he notes that, for instance, in English the countability properties of nouns can be empirically determined on the basis of the relative frequencies of their occurrence as mass or count. Countability preferences can thus be viewed as part of the lexical semantics of nouns that can then be manipulated at the level of construal depending on the discourse context. Corbett (2000) uses the term *recategorization* to refer to the manipulation of the countability properties of a noun at the level of construal (see §2.1.1 for the application of the notion of recategorization to the gender domain).

An important contribution to the understanding of the lexical semantics of nouns in relation to their countability properties is the notion of *boundedness*, which has been discussed in the literature largely within the framework of cognitive grammar (Langacker 1987a,b; Talmy 2000a,b) and conceptual semantics (Jackendoff 1991). Count nouns are bounded by default in the sense that "one cannot divide [their referent] up and still get something named by the same count noun," e.g., *apple* (Jackendoff 1991: 18). Conversely, mass nouns are unbounded in the sense that "one can divide [their referent] up and still get something" named by the same mass noun, e.g., *water* (Jackendoff 1991: 18). According to the literature mentioned above, the notion of boundedness can be applied to the lexical semantics of both nouns and verbs,[11] thus connecting countability properties of nouns and types of event structure. In Jackendoff's (1991) analysis, a second semantic feature is necessary for the characterization of types of nouns in terms of countability properties: *internal structure*. Both *boundedness* and *internal structure* work as binary values. The combination of the possible values with respect to the two parameters gives the following possible types of nouns:

+ bounded, – internal structure = individual entities e.g., *a cat*

+ bounded, + internal structure = groups e.g., *a jury*

– bounded, – internal structure = substances e.g., *milk*

[11] In the verbal domain, the opposition *bounded* vs. *unbounded* would be instantiated by the opposition *perfective* vs. *imperfective* (Langacker 1987a).

– bounded, + internal structure = aggregates (plurals and collectives) e.g., *cats*, *cattle*.

The opposition between mass and count nouns can be reflected in various ways at the level of morphosyntax, for example:

(1) Count nouns may combine quite well with numerals and measure words. On the other hand, in order to count or measure a mass noun, many languages use more elaborated constructions, as in English *three cars* as opposed to *three glasses of water*. The items that are added as a unit of measure for mass nouns are referred to in the literature as *mensural classifiers* (Gil 2013).[12]

(2) Mass nouns often participate in number distinctions in a noncanonical way. In languages whose number system is based on the opposition between singular vs. plural, for example, mass nouns are often either only singular or only plural.

The countability properties of nouns interact with number marking, but the dynamics of such interactions are highly language-specific.

2.2.2 Types of number systems and number values

Not all languages need to obligatorily express nominal number values. In this section, I discuss types of number systems according to the presence or absence of obligatory number marking.

2.2.2.1 General number

Languages that do not encode number distinctions obligatorily have been defined in the literature as having *transnumeral number*, *unit reference* or *general number*. In this dissertation, I use the latter label, *general number*, which was introduced by Andrzejewski (1960) in a study of number marking in Borana-Arsi-Guji Oromo (Afro-Asiatic, Cushitic), and was later adopted by Corbett (2000) in his typological investigation of the category of number. Nouns with general number are by definition outside the system of number distinctions (and are therefore morphologically zero-marked for number); their translation equivalent in languages with obligatory number could be either singular or plural. Normally, in languages with general number, there are additional means used to specify the number value of the noun as singular or plural, if needed. Moreover, in languages with general number, singularity and plurality are generally obligatorily coded on personal pronouns. The frequency of use and the distribution of nouns with general number as opposed to number-marked nouns vary a great deal across languages, as do the strategies of formal marking. It is generally quite rare to have three distinct markers for differentiating among general, singular and plural number with all nouns. What is more

[12] As Gil (2013) puts it, "most or all languages" have mensural classifiers. Overall less frequent are *sortal classifiers*. Sortal classifiers – or *sortal numeral classifiers* – are used when nouns occur in combination with numerals and classify nouns according to inherent properties of the NP referent such as animacy and shape. Sortal numeral classifiers can be both optional and obligatory, and are especially common among the languages of East and Southeast Asia (for an overview, see Gil 2013).

common, instead, is that the general number form patterns, at least for some nouns, either with the singular or with the plural. In Cushitic (Afro-Asiatic) languages, for instance, number unmarked nouns can be: 1) nouns with general number, 2) inherently singular nouns and 3) inherently plural nouns. Plurality with inherently singular nouns is marked by means of plural markers, and singularity with inherently plural nouns is marked by singular markers. In principle, nouns with general number can be marked either as singular or plural, if needed. However, Cushitic languages vary a great deal as to how much overt number marking is allowed with nouns with general number. Number marking in the Cushitic language Kambaata is illustrated in example (2.6).

(2.6) Kambaata (Afro-Asiatic, Cushitic, East Cushitic) (adapted from Treis 2008)

(a) General number noun

haqq
'tree(s)'

(b) Inherently singular noun

meseleta
'girl'

(c) Plural of an inherently singular noun

masal-áakkaáta
girl-PL.F
'girls'

(d) Inherently plural noun

meentú
'women'

(e) Singular of an inherently plural noun

meent-iccúta
woman-SG.F
'woman'

In Kambaata, nouns with general number can sometimes be marked by the singular markers, but the pragmatic trigger of such uses is not very clear (Treis 2008: 140). In Treis' (2008) reference grammar of Kambaata, and more generally, in the literature on Cushitic languages, the singular and plural markers are referred to as *pluratives* and *singulatives*, respectively. This special terminology is used to signal the fact that overt number markers in Cushitic are usually derivational in nature. A more detailed account of the system of number marking in Cushitic languages will be given in chapters 4 and 5.

2.2.2.2 Obligatory number marking and possible number values

Languages like English do not have general number, and speakers need to choose the number value of a noun when they use it (Corbett 2000). As mentioned earlier, the number values that are found in the languages of the world are: singular, plural, dual, trial, quadral, paucal and greater plural. Their distribution is constrained by implicational tendencies. A Number Hierarchy has been formulated on the basis of the observation of such tendencies:

singular > plural > dual > trial

The hierarchy can be seen as equivalent to Greenberg's Universal 34, which he elaborated on the basis of a sample of 36 languages:

> No language has a trial number unless it has a dual. No language has a dual unless it has a plural (1963b: 94).

The Number Hierarchy is structured in such a way that the values to the right imply those to the left. Three values are missing both from the Number Hierarchy and Greenberg's Universal 34: paucals, greater plurals and quadrals. The paucal marks a small number of entities in the real world. It is more common in systems with singular, dual and plural number values than in systems with singular and plural values only. The paucal number is widespread among Oceanic languages (Corbett 2000: 24). The greater plural marks very large numbers of entities in the real world. Fula (Atlantic-Congo, North-Central Atlantic) is an instance of a language with a greater plural (Corbett 2000: 31). The existence of quadral markers has been suggested for a number of Austranesian languages. However, Corbett (2000: 30) argues that for all such cases, there is no clear evidence to claim that quadrals, defined as markers "for referring to four distinct real world entities in the way that trials refer to three" actually exist. One problem with the formulation of the Number Hierarchy then is when and how the the paucal and greater plural should be incorporated. This is discussed in detail by Corbett (2000: 39-42), who ultimately proposes to think about paucal and greater plural as types of number values that differ in nature from the ones appearing in the Number Hierarchy.

2.2.3 Number marking and the Animacy Hierarchy

Typological investigations of nominal number (Corbett 2000; Smith-Stark 1974) have shown that the distribution of plural marking in a language is conditioned by the animacy of NP referents. This is ultimately connected to the countability properties of nouns, whereby animate nouns tend to also be countable nouns. The version of the Animacy Hierarchy that is canonically used to account for typical and atypical strategies of number marking on nouns and pronouns within and across languages is the following:

speaker > addressee > 3rd person > kin > human > animate > inanimate

Haspelmath (2013) elaborates a new version of the Animacy Hierarchy, which he uses to describe the distribution of plural marking on nouns only:

kin > other humans > "higher animals" > "lower animals" > discrete inanimates > nondiscrete inanimates

In contrast to the version used by Smith-Stark (1974) and Corbett (2000), the Animacy Hierarchy proposed by Haspelmath (2013) does not take into account pronouns – because they often display peculiar encodings of plurality – and is based on a more fine-grained classification of noun types in between kinship terms and inanimate nouns.

Both in its original and revised version, the Animacy Hierarchy is concerned with the way in which overt plural marking is distributed according to the semantic properties of nouns. Generally speaking, the likelihood of obligatory plural marking increases the more one moves toward the leftmost end of the hierarchy. In many languages, only human or animate nouns (together with personal pronouns) undergo obligatory plural marking. At the opposite end of the Animacy Hierarchy, inanimate entities and non-discrete inanimate entities, in particular, may be insensitive to number distinctions. The role of animacy in number marking has been explained as the result of cognitive salience: "the distinction between one and more than one is more salient for animates than for inanimates, so that speakers are more likely to make use of available plural markers when they refer to a plurality of animates" (Haspelmath 2013). In §2.2.5, I discuss how such preferential patterns of language use become structurally obligatory as a consequence of grammaticalization.

Finally, it is worth mentioning that, if a language has associative plural constructions of the type "X + those surrounding X", these are often used at least with kinship terms, which occupy the top-ranking position of the Animacy Hierarchy proposed by Haspelmath (2013) (see Daniel & Moravcsik 2013 and Moravcsik 2003 on the typology of associative plural constructions; on the peculiar grammatical properties of kinship terms, see Dahl & Koptjevskaja-Tamm 2001). The same constructions are also often used with proper person names, which, however, do not figure as an independent type on the hierarchy.[13] Associative plural constructions express non-additive plurality: they do not refer to a homogeneous set of referents but rather to the combination of a focal referent (the one referred to by the kin term or the proper name) and a group of referents associated with it.

2.2.4 The morphosyntax of number: indexation

As mentioned in §2.2, nominal number distinctions can be marked via affixation on nouns, plural words or via indexation in the sense discussed in §2.1.2.1. The most common number-indexing targets in the languages of the world are demonstratives and verbs. In general, indexation is described as less central for the definition of the number system of a language than it is for gender. In some languages, however, syntactic manifestations of number are crucial for encoding number distinctions since these are not obligatorily coded on nouns. In languages in which both gender and number indexation coexist, the two systems can either work in parallel or compete to some extent with each other. As

[13] On the opportunity of including proper names in the hierarchy, see Koptjevskaja-Tamm (2002) and Moravcsik (2003).

outlined in chapter 1, interactions between gender and number in the indexation domain are one of the research foci of this dissertation and will be discussed in detail in chapter 5.

As in the case of gender, manifestations of nominal number through indexation are constrained by the Agreement Hierarchy:

attributive > predicate > relative pronoun > personal pronoun

Number indexation can also be either referential or lexical (see the discussion in §2.1.2.2), and the likelihood of referential indexation increases as one moves rightward on the hierarchy. As with gender indexation in the case of *hybrid nouns*, the possibility of mismatches between indexation triggers and targets also exists in the case of number indexation. Examples (2.7) and (2.8) illustrate cases of mismatch for different types of triggers.

(2.7) Corporate nouns in English (Corbett 2000: 189)
The committee have decided

(2.8) Honorifics in Bulgarian (Indo-European, Slavic) (Ljuba Veselinova, personal communication)

Vie	*ste*	*stanal*	*nerven*
You.PL	be.2.PL	become.M.SG	nervous.M.SG

'You have become nervous'

In example (2.7), the noun *committee* denotes a group, that is, a multiplicity of individual entities conceptualized as a unit. Plural indexation with the word *committee* in example (2.7) signals that the multiplicity aspect of the semantics of the NP is highlighted. Singular indexation is preferred to plural indexation by speakers of American English, as opposed to speakers of British and New Zealand English who tend to accept both (Corbett 2000). Example (2.8) shows the co-occurrence of singular and plural indexing targets with the same head noun, when the head noun is plural and denotes an individual in polite address forms. Both the adjective, *nerven*, and the participle, *stanal*, perform a predicative function and show singular indexation. Only the auxiliary *ste* is inflected as plural. Comrie (1975) claims that phenomena of this type can be taken into account by splitting the predicate position of the Agreement Hierarchy (11) into a sub-hierarchy of the type: verb <participle <adjective <noun, where the probability of referential indexation increases as one moves from more prototypical to less prototypical instances of predicate-like words, that is, from verbs to nouns. The two examples illustrate that, similarly to gender indexation, the choice between different patterns of number indexation depends on the types of indexation trigger, the types of target, and the discourse context.

Special constraints in the distribution of patterns of number indexation are also found for quantified NPs, associative plurals and conjoined NPs. In the case of conjoined NPs, languages are biased between preference toward indexation patterns with the conjunct

that is closer to the target and preference for indexation patterns with all conjuncts. Both possibilities can be found in one and the same language; their distribution is usually constrained by the Agreement Hierarchy. Precedence in linear order and animacy also play a role. Indexation triggers that precede their targets are more likely to trigger indexation according to the referential properties of the NP and even more so if animate and/or human. For a more detailed typology of number indexation with conjoined NPs as well as for a discussion of similar matters with quantified NPs and associative plurals, see Corbett (2000: 195-216).

2.2.5 Distribution and diachrony of number systems

2.2.5.1 The distribution of number systems

Nichols (1992) hypothesises the existence of a correlation between the morphological characteristics of languages and the types of attested number systems. Grammatical number marking is, for example, very infrequent among the isolating languages of South-east Asia. The WALS database shows comparable results concerning the distribution of obligatory and non-obligatory strategies of number marking: if one looks at the world map based on the data by Haspelmath (2013), the distribution of obligatory nominal plural marking appears to be quite skewed. Africa and the westernmost part of Eurasia are the areas where this type of marking is most commonly found. Similar results have been attained by Wälchli (2012), who investigates the frequency of occurrence of nominal plural marking in a sample of 82 languages by means of parallel texts. The study confirms that the frequency of occurrence of nominal plural marking is skewed towards Africa and the European region of the Eurasian macro-area. It also shows that the occurrence of nominal plurality tends to align – although not without exception – with the predictions made by the Animacy Hierarchy. In the same study, Wälchli (2012: 255) suggests that attrition in the nominal number domain rarely leads to complete loss. When, as, for instance, in French, number marking on head nouns disappears due to phonological erosion, number-indexation strategies used on demonstratives, articles or even verbs are likely to become more prominent in discourse thus allowing for number distinctions to be still overtly coded through syntax.

2.2.5.2 Sources of number

Grammaticalized number markers can be traced to several sources such as demonstratives, quantifiers, nouns or markers of verbal plurality. Cristofaro (2012) makes a useful distinction between sources of nominal number markers that are etymologically connected with the encoding of plurality and sources that are instead entirely independent of number. The latter come to be used as number markers as an accidental result of grammaticalization. Among the first type of sources, she includes:

(1) Collectives and distributives (expressions like ‘people here and there’), which develop as plural markers in an already existing system of grammaticalized number marking.

(2) Expressions of multitudes: quantifiers such as 'many,' 'a lot,' 'all' or nouns denoting groups of entities – usually humans – such as 'people,' 'men'.

To the above-mentioned sources one could also add markers of nominal plurality that originate from markers of verbal plurality. For instance, Mithun (1988) investigates the evolution of nominal number marking in a variety of North American languages in which the marking of number on the NP (nouns and adjectives) is the result of extension from markers of verbal number. This is not generalised to all nouns but is restricted to highly individuated nouns only.

Among the second type of sources of nominal number marking in Cristofaro's (2012) classification, demonstratives and other deictic or anaphoric entities are found. For instance, Frajzyngier (1997) analyses markers of nominal (and verbal) plurality in a variety of Chadic languages that developed as a result of grammaticalization from demonstratives. Plural marking on nouns is intimately connected to definiteness; deictic markers that initially had scope over entire NPs, and only marked definiteness, later began to mark plurality on nouns.

Different patterns of grammaticalization of nominal number share analogous distributional properties: the overt coding of number distinctions is initially very likely to occur with highly individuated and human or highly animate nouns only. This preference in language use can lead, through increasing frequency, to obligatory marking. Obligatory number marking may then stay bound to high degrees of animacy and individuation or become extended to all types of nouns (Haspelmath 2013). Cristofaro (2012) criticises this approach by claiming that explanations that account for synchronic distributions of language types according to typological hierarchies (in this case the Animacy Hierarchy) cannot always be applied to diachrony. Markers of nominal plurality originate from different and often unrelated diachronic patterns and "the hierarchy is simply a schema which is general enough to subsume all of these processes" (Cristofaro 2012: 8). Along these lines, I shall show in chapter 5 that, for instance, the Animacy Hierarchy may not have the ability to account for the development of number indexation in certain languages of the African macro-area.

2.3 Evaluative morphology

Languages have various means for expressing the semantic values of *big* vs. *small* and *good* vs. *bad*. The markers that morphologically encode these distinctions are labelled *augmentatives* (= bigger size), *diminutives* (= smaller size), *appreciatives* (= something is good), and *depreciatives* (= something is bad). The same marker can express more than one value at once – e.g., small size is often associated with appreciation/endearment or depreciation/contempt – and it is often only on the basis of context that overlapping meanings can be teased apart. Diminutive and augmentative markers belong to the functional and semantic domain referred to as *evaluation*. Morphological evaluative markers constitute a subdomain of the domain of evaluation and are commonly categorized under the label of *evaluative morphology*. They can be found associated with various word classes, e.g., nouns, verbs, adjectives, but nouns seem to be the word class most likely

to undergo evaluative marking (Bauer 1997). In this thesis, I focus on the marking of evaluative morphology on nouns only and, in particular, on those constructions that encode diminutive and augmentative meanings.

The rest of the section is structured as follows: the semantics and pragmatics of evaluative morphology is treated in §2.3.1; in §2.3.2, I provide an overview of the major strategies of evaluative marking and their distribution in the languages of the world before I move on to discuss the diachrony of diminutives and augmentatives in §2.3.3.

2.3.1 The semantics, pragmatics and functions of evaluative markers

One of the seminal crosslinguistic studies on the semantics of evaluative morphology is Jurafsky (1996). His analysis of the semantics of diminutive markers in the languages of the world is based on the notion of *radial category*, following Lakoff (1987). A simplified version of the schema elaborated by Jurafsky as a possible representation of the universal semantic properties of diminutives is provided in figure 2.4.

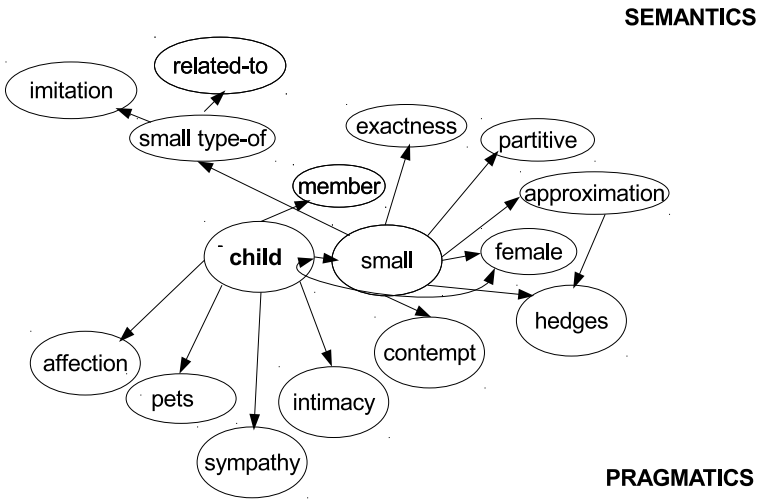


Figure 2.4: Universal semantic and pragmatics properties of diminutives according to Jurafsky (figure adapted from Jurafsky 1996: 542)

The figure can be read both synchronically and diachronically. As explained by Jurafsky, each node is labelled with one of the senses associated with the diminutives, whereas the arrows stand for patterns of semantic change via inference, metaphor, generalization and lamda-abstraction-specification.[14] Jurafsky's schema shows that the semantically

[14] Jurafsky (1996) defines lambda-abstraction-specification as a mechanism of semantic change that ac-

prototypical, and historically original, meaning of the construction is 'child' (see §2.3.3). In addition, the figure maps aspects of the pragmatics of diminutives, that is, the way diminutive markers are used to express the attitude of the speaker towards the referent. Reference to the semantics of augmentatives is made in the paper, but no systematic treatment of the category is carried out.

Another attempt at unifying semantics and pragmatics in the representation of the function of diminutive constructions is that of Wierzbicka (1984). Wierzbicka analyses the categories of diminutive and dispreciative in Polish and Australian English within the framework of Natural Semantic Metalanguage.[15] The different usages of diminutives in the two languages are explained in terms of sociocultural differences between the two speech communities.

Diminutives and augmentatives received special attention within the framework of Morphopragmatics, which was initiated by Dressler & Barbaresi (1994) as a theory and a method for analysing the pragmatic relevance of morphological rules. Within the morphopragmatic approach, the pragmatics of diminutives and augmentatives is viewed as prior and superordinate to semantics.

Grandi's (2002) crosslinguistic study of evaluative morphology in the languages of the Mediterranean area is one of the few existing typological investigations of evaluative constructions. Grandi distinguishes between a *descriptive* and a *qualitative* dimension of evaluative morphology, in which the former refers to the use of evaluatives markers in connection to size variation, and the latter to the use of evaluative markers to express speakers' attitudes toward the NP referents. Körtvélyessy (2012) has a similar understanding of the domain and defines evaluative morphology as the set of morphological strategies that languages use to encode the semantic notion of "less than/more than the standard quantity of substances, qualities, actions and circumstances, with the concept of standard quantity being a relative one." Both models take a unified approach to the semantics and pragmatics of evaluative markers. In this thesis, I use the two models by Grandi (2002) and Körtvélyessy (2012) as guidelines for the description of the semantic properties of diminutives and augmentatives in the languages of my sample.

An interesting comparison can be made between evaluative morphemes and those strategies for the encoding of evaluative meanings that fall outside morphology. Analytic constructions of the type "Adjective encoding size or value + noun" – and vice versa – are generally kept outside the domain of evaluation. However, to a certain extent, these constructions show semantic and functional characteristics that are comparable with those of evaluative morphemes. On a general basis, for instance, the basic meanings of evaluative markers – 'small,' 'big,' 'good,' 'bad,' 'young,' 'old' – correspond

counts for the rise of second-order predicates from diminutive (e.g., approximation, adverbial partitive quantifiers, hedges).

[15] Natural Semantic Metalanguage is a theory of semantics based on the assumption that linguistic meaning, both at the lexical and grammatical level, can be explained by means of *semantic primitives*, or *primes*, and *reductive paraphrasis*. Semantic primitives are a set of simple, undefinable and universal concepts; reductive paraphrasis is a method of semantic analysis that aims at reducing the complexity of language structure to simpler concepts on the basis of semantic primes. Anna Wierzbicka and Cliff Goddard are the main proponents of the theory.

to some of the basic semantic types of prototypical adjectives in Dixon's (1982; 2004) classification: *dimension*, *value*, *age*. In languages lacking evaluative morphology, adjectives that encode the above-mentioned meanings are used in cases in which languages with evaluative morphology would tend to use diminutives or augmentatives. Interestingly, in spite of similar functional and semantic underpinnings, evaluative morphemes and adjectives expressing *size* and *value* are not at all related from a diachronic point of view. As discussed in §2.3.3, there is no evidence for morphological evaluative markers of any type being the diachronic descendents of semantically related prototypical adjectives. The two types of constructions seem to have independent and non-matching histories. Commonalities and differences between affixal and non-affixal encoding of evaluation surely represent a very promising area of research for the understanding of the semantics and pragmatics of evaluation.

2.3.2 Types of marking in evaluative morphology

A valuable generalisation to make when considering types of evaluative markers and their distribution across languages is that diminutives are generally more frequent than augmentatives (Bauer 1997; Dahl 2006; Körtvélyessy & Stekauer 2011). Moreover, certain usages that are commonly associated with diminutives – for instance, their use as quantifiers of uncountable nouns – are never or very rarely encountered with augmentatives.

2.3.2.1 Phonetic iconicity in evaluative morphology: universal or language-specific?

A relevant research trend in the field of evaluative morphology focusses on the existence of iconic patterns of sounds in diminutive and augmentative marking. This tradition of studies goes back to Jespersen (1922) and Sapir (1951), who pointed out the existence of a relation between the high, front vowel [i] and the encoding of small size in a variety of European languages (Jespersen) and in English (Sapir). In Universal #1926 of the *Universals Archive* (Plank & Filimonova 2009), such patterns are claimed to correspond to a universal tendency :

> There is an apparently universal iconic tendency in diminutives and augmentatives: diminutives tend to contain high front vowels, whereas augmentatives tend to contain high back vowels.

The universal is based on Payne (1997).

Further typological investigations, such as Ultan (1978), Bauer (1996), Gregová et al. (2010), Körtvélyessy (2011) test the hypothesis of sound symbolism in evaluative morphology on different language samples. Both vowels and consonants are considered in these studies, the assumption being that palatal (or post-alveolar) consonants – together with high front vowels – are more typical of diminutives than augmentatives. The results of these studies do not confirm the existence of universal patterns of phonetic iconicity in evaluative morphology. The initial hypothesis is thus rejected in all cases. The studies nevertheless show that phonetic iconicity does play a role in the encoding of diminutives

and augmentatives but that the patterns of phonetic iconicity tend to be areal or rather language-specific (for an overview of patterns of phonetic iconicity in the languages of western North America, see, e.g., Nichols 1971).

2.3.2.2 Morphosyntactic encoding of evaluative markers

The morphological encoding of evaluation varies across languages. The following types of marking are attested in the languages of the world (an example for each of the patterns mentioned above is provided):

(a) Affixation

(2.9) Diminutive suffixes in Italian (Indo-European, Romance) (constructed example)

(a) *tavolo*
table.M.SG
'table'

(b) *tavol-ino*
table-DIM.M.SG
'small table'

(2.10) Diminutive prefixes in Italian (Indo-European, Romance) (constructed example)

(a) *appartmento*
apartment.M.SG
'apartment'

b *mini-appartamento*
DIM-apartment.M.SG
'small apartment'

(2.11) Diminutive infixes in Standard Arabic (Afro-Asiatic, Semitic) (Grandi 2002: 220)

(a) *kalb*
'dog'

(b) *k**ulay**b*
dog.DIM
'puppy'

(b) Reduplication

(2.12) French (Indo-European, Romance) (Grandi 2002: 256)

gué-guerre
RED-war
'small fight between people among a group, normally without physical violence' (translation amended by a native speaker)

The distribution of the different types of evaluative markers varies according to language specific properties (e.g., how much morphology there is in a language) and is of course conditioned by the degree of grammaticalization of individual evaluative markers within a given language (see §2.3.3).

In languages with highly elaborated systems of evaluative morphology, several different strategies can be found. Different markers can be found associated with a particular subset of nouns or with the encoding of particular meanings. In addition, evaluative markers can co-occur on the same noun and express (1) extremes in size, whether very small or very large, or (2) the combination of endearment/contempt meanings and size-related meanings. Finally, on a general basis, evaluative markers do not trigger indexation. The only clear exception is represented by those languages in which diminutives and augmentatives are part of the inventory of gender distinctions. This is discussed in detail in chapter 6.

2.3.3 Distribution and diachrony of evaluative markers

2.3.3.1 The distribution of evaluative markers

Evaluative morphology is, in many respects, still an under-investigated phenomenon, and there is no tradition of studies on the distribution of evaluative markers in the languages of the world. A very recent and promising exception to this trend is the work by Körtvélyessy (2012). Based on a sample of world languages stratified according to macro-areas, Körtvélyessy (2012) examines the stability and diffusion of evaluative morphology worldwide. Her ultimate aim is to establish whether evaluative morphology can be counted as one of the defining features of Standard Average European.[16] The data were collected on the basis of a questionnaire designed by the author as well as on the basis of reference grammars. An index calculation formula was then used as an instrument for data analysis. The formula measures the complexity of evaluative morphology within languages, which Körtvélyessy refers to as *Evaluative Morphology Saturation*:

$$\text{EMS} = \frac{(\text{V}_{\text{WF}}+\text{V}_{\text{SC}}+\text{V}_{\text{WC}})}{3}.$$

For each of the sampled languages, the Word Formation Value (V_{WF}) stands for the number of word-formation processes attested for the encoding of evaluation. The value for each language is computed as follows: one point is assigned if a language has one or two word-formation processes in connection with evaluation, two points are assigned if the word-formation values are three or four and so on. The Semantic Category Value (V_{SC}) represents the types of semantic categories (descriptive or qualitative) expressed by evaluative markers, whereas the Word Class Value (V_{WC}) refers to the word classes that undergo evaluative morphology. One point is assigned to each of the semantic categories and word classes that are found in one language. The degree of EMS is calculated

[16] The notion of Standard Average European was introduced by Whorf (1941) in reference to the languages of Europe when viewed as a linguistic area. An inventory of the features shared by the languages of Standard Average European and their relative distribution in the area is presented in Haspelmath (2001). See also Dahl (1990) and the volume by Heine & Kuteva (2006).

independently for diminutives and augmentatives; the total EMS for each language is a sum of the two. The results of the investigation show that Standard Average European features rather complex systems of evaluative marking. Thus, according to Körtvélyessy (2012), evaluative morphology should be included among the defining characteristics of the European linguistic area.

2.3.3.2 The diachrony of evaluative markers

Nouns are the most common source of diminutives and augmentative markers. A nearly universal source of diminutives in the languages of the world is the word for 'child' (Jurafsky 1996), which usually originates as a sort of classificatory noun used to denote the young age of animate entities and is gradually extended to inanimate nouns for which it is used to mark small size in countable nouns and small quantity in uncountable nouns. Another fairly common source of diminutives are affixes that express relational meaning or resemblance and develop as markers of approximation and then as diminutives. This is, for instance, the case of the Italian diminutive suffix *-ino/-ina*, derived from the Latin suffix *-inus/-ina*, which originally meant something like 'related to X' (Grandi 2002). Interestingly, in Jurafsky's terms (1996: 553), the 'related to X' meaning may also be indicative of late stages of grammaticalization of diminutive markers originally expressing size. These markers develop more abstract relational meanings according to the pattern: *small size* > *small type of* > *related to*. This may be indicative of the existence of some sort of bidirectionality in the grammaticalization of diminutives, a possibility that would certainly need to be further investigated. Finally, as pointed out at the end of §2.3.1, and as also noticed by Jurafsky (1996) and Dahl (2006), there is no evidence for affixal diminutives deriving from modifiers meaning 'small.' These can be productively used in analytic constructions of the type "Modifier + noun" (as in English *small* and *little*) but are never attested as sources of affixal diminutives.

The origin of augmentative markers has not been investigated in detail in the literature on evaluative morphology. The exceptions are very few, as in Grandi's (2002) overview of the development of augmentatives in the languages of the Mediterranean area, or in the case of Matisoff's (1992) study of diminutives and augmentatives, in some languages of Southeast Asia. In the latter case, the origin of augmentative markers is said to be the word for 'mother.'

2.4 Gender and its interaction with number and evaluative morphology

Gender, number and evaluative morphology have been introduced in §§2.1, 2.2 and 2.3. In this second and last part of the chapter, I introduce the specific object of study of this dissertation.

As already mentioned in chapter 1, this dissertation has three main research foci: (1) interactions between gender and number; (2) interactions between gender and evaluative morphology; and (3) interactions of gender and grammatical complexity. The following

research questions are addressed within each of the three research foci (a more fine-grained formulation of the research questions is presented in §2.8).

(1) Interactions between gender and number
   - Do gender and number have the same encoding? How are gender distinctions mapped across number values?
   - Do gender and number compete through indexation patterns?
   - Do gender and number have the same relevance to nouns?

(2) Interactions between gender and evaluative morphology:
   - Can *size* be a criterion for gender assignment?
   - How do diminutive and augmentative genders interact with the other gender distinctions of a language?

(3) Interactions of gender and grammatical complexity:
   - Is it possible to measure the grammatical complexity of gender systems?
   - Can interactions of gender with other domains of grammar be accounted for by such a complexity metric?
   - What is the role of these interactions in the overall complexity of a gender system?

Previous studies on the interactions between gender and number are discussed in §2.5, whereas §2.6 focusses on interactions between gender and evaluative morphology. The notion of grammatical complexity and previous studies on the complexity of gender systems are tackled in §2.7.

2.5 The interaction between gender and number: state of the art

That gender and number are engaged in a special relationship with each other is an assumption that linguists from different theoretical backgrounds and with different research agenda share with great confidence.

By examining noun classification systems in a quite diverse sample of languages, de la Grasserie (1898) observed that number functions as a sort of catalyst for noun classification. According to his analysis, this applies both to classifier systems and gender systems, whose function appears to be often related to the quantification of nouns. A similar claim is made in more recent work by Crisma et al. (2011), which investigated three types of noun classification strategies – non-sex-based gender, sex-based gender, and numeral classifiers – in three different genealogical groupings – Bantu, Romance and Sinitic. The purpose of the study was to verify the existence of any common functional ground shared by different noun classification systems. The authors conclude that noun classification strategies in the languages investigated in the paper (particularly Swahili,

Italian, and Cantonese) "play an important role in the process of singling out individuals" and quantifying them (Crisma et al. 2011: 292).

Two parameters that are often used to classify types of interaction between gender and number are *exponence* (cumulative vs. noncumulative) and *syncretism*. These are discussed in detail in the next two sections.

2.5.1 Exponence

In Bickel & Nichols' (2007) model of inflectional morphology, exponence is a dimension of *semantic density*. This can be defined as the amount of meaning that is stored by a morphological unit. Exponence measures semantic density at the level of morphemes, whereas *synthesis* measures semantic density at the word level. Technically, exponence can be cumulative, noncumulative (also referred to as *separatist*) or *portmanteau*. Cumulative exponence is found when two or more grammatical features are encoded by nonsegmentable markers. Similar to cumulative exponents, portmanteau exponents also express more than one category at once. However, the difference between the two types of exponence lies in the fact that in portmanteau exponents "each of the categories expressed corresponds to a separate formative that also exists in the language" (Bickel & Nichols 2007: 188), whereas this is not the case for cumulative exponents. For instance, similar to the French example given by Bickel & Nichols, in Italian, the definite preposition *del* 'of the' is a portmanteau morpheme that has two corresponding separate morphemes in the language: *di* 'of' and *il*, the Masculine Singular Definite Article. On the other hand, on adjectival modifiers, the suffix *-a* encodes feminine and singular cumulatively, but there are no gender-only and number-only corresponding suffixes in the language. Cumulative, noncumulative and portmanteau exponents can coexist in one and the same language and within one and the same grammatical domain. The distribution of such splits in types of exponence is not always random (Plank 1999). According to Plank, the first relevant morphological split is the one between derivational and inflectional features of grammar: derivational features rarely display cumulation, whereas inflectional features are "more prone to licence at least some cumulation, though never with derivational categories" (1999: 292). In his paper, Plank investigates morphological splits between agglutinative and flexive patterns of encoding of case and number, and focusses on splits due to exponence (cumulative vs. noncumulative) and *variance* (morphological alternations within paradigms that cannot be phonologically motivated).

Gender and number often have cumulative exponence. This is asserted in two of the most well-known typological surveys of the two grammatical features:

> Number is of more importance for gender also because it is the category most often realized together with gender. (Corbett 1991: 189)

And, similarly,

> Number is often marked morphologically in this way, cumulated with case, gender or person, depending on the word-class. (Corbett 2000: 145)

To the best of my knowledge, however, there are no studies that systematically survey cumulative exponence of gender and number, and their implications on our understanding of the two grammatical domains. On the other hand, gender and number-related patterns of syncretism have received more attention in the literature. And it is to syncretism that I now turn.

2.5.2 Syncretism

Syncretism is a type of paradigmatic asymmetry whereby certain grammatical distinctions are neutralized or reduced under certain conditioning factors. Patterns of syncretism always feature at least two grammatical domains. The grammatical domain that conditions the occurrence of syncretism is often referred to as the *context* of syncretism. Thus, linguists often talk about syncretism *in the context of case*, syncretism *in the context of person*, or, as is the case in this thesis, syncretism *in the context of number*.

The most recent comprehensive survey of patterns of syncretism in the languages of the world and across different grammatical domains is Baerman et al. (2005). According to this model, syncretism differs from neutralization in that the lack of distinctions is partial in the former and total in the latter. For instance, in Seneca (Iroquoian), Masculine, Feminine and Neuter Gender are distinguished under singular reference; however, only two gender distinctions are preserved under plural reference, as the Feminine and Neuter are conflated. This is an example of syncretism (Baerman et al. 2005: 84). In Russian (Indo-European, Slavic), Masculine, Feminine and Neuter Gender are distinguished only under singular reference: gender is not syntactically relevant under plural reference (Baerman et al. 2005: 28-29). This is an example of neutralization. In languages with two gender distinctions under singular reference, lack of distinction under nonsingular number values will always lead to neutralization. In this dissertation, I do not distinguish between partial or total lack of distinction because both phenomena have similar implications for the relationship between gender and number. I refer to both types of asymmetries as syncretism, but I specify when a certain pattern of syncretism leads to neutralization.

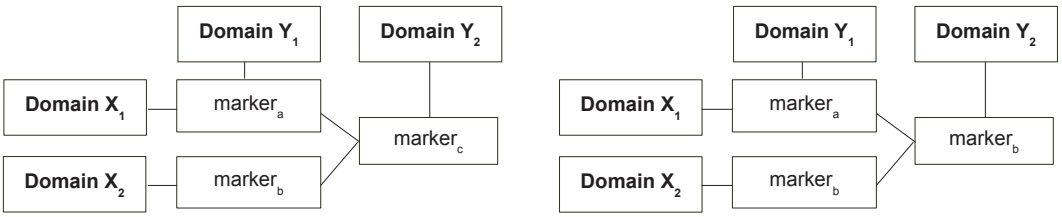
Two types of syncretism can be identified based on how the syncretic forms relate to their non-syncretic counterparts:

(1) The syncretic marker formally differs from each of its component values.

(2) The syncretic marker is formally the same as one of its component values.

These two possible patterns of syncretism are illustrated in figure 2.5. Given two grammatical domains X and Y, and their respective values (X_1, X_2, Y_1, Y_2), X is subject to undergo syncretism in the context of Y when the value of domain Y is 2. This syncretism may result into two alternative patterns depending on whether the value of X in the context of Y_2 differs from (as in 2.5a), or is the same as (as in 2.5b), one of the values of X under Y_1. In the literature on syncretism (see, among others, Baerman et al. 2005; Carstairs 1984, 1987; Carstairs & Stemberger 1988), the pattern in figure 2.5b is also

referred to as *take-over*, since it functions as if one of the values of X under Y_1 *spreads across* paradigmatic cells. In the model of syncretism designed by Baerman et al. (2005), these phenomena are explained in terms of *directional effects* and are formally accounted for by means of *directional rules*.



(a) Syncretism 1 (b) Syncretism 2 (or *take-over*)

Figure 2.5: Possible patterns of syncretism

Previous studies have built on the idea that syncretism is concerned with *markedness relationships*.[17] In the case of gender and number, for instance, it has been often noted that fewer gender distinctions tend to appear with more marked number values. This tendency was formalised as an implicational universal by Greenberg in his Universal 37:[18]

> A language never has more gender categories in nonsingular numbers than in the singular. (1963b: 95)

Greenberg's generalisation is based on a sample of 36 languages and was further tested on a broader dataset (300 languages) by Plank & Schellinger (1997). Plank & Schellinger (1997: 59) argue that, if Universal 37 is taken as having scope over languages in their entirety, its predictive power is stronger than if it is taken as having scope at the word level. At the word level, exceptions tend to be more frequent, both within and across languages. Individual sets of nouns in individual languages may exhibit peculiar markedness preferences with respect to number values. This may come as a result of diachronic accident or as an effect of functional pressure. Examples of diachronic accident are

[17] In this thesis, I follow Haspelmath's (2006) critical reappraisal of the notion of *markedness*. The notion of *markedness* as a property of grammar or an explanatory tool is used in the dissertation only when reviewing studies that explicitly worked with such notions. The words *marked* or *unmarked* are otherwise only used as synonymous with *overtly-coded* and *zero-coded*, respectively.

[18] In Universal 45, Greenberg formulates a similar generalisation about the marking of gender distinctions on singular and plural pronouns: "If there are any gender distinctions in the plural of the pronoun, there are some gender distinctions in the singular also" (1963b: 96). Universal 45 is not discussed in detail in this thesis.

those cases in which gender distinctions enter the nonsingular part of a paradigm when words distinguishing biological gender are grammaticalized as nonsingular number markers. On the other hand, functional pressure can be observed when certain gender distinctions are used to solve gender conflicts that can only arise with nonsingular numbers (Plank & Schellinger 1997: 93). Moreover, it is suggested that with nouns that are semantically associated with nonsingularity, nonsingular number values may be less marked than the singular, and this may also reflect on the way gender distinctions are parsed across number values (1997: 94).

According to Baerman et al.'s (2005) model, when gender undergoes syncretism across number values, gender distinctions are likely to be *realigned* on a semantic basis. For instance, all nouns denoting animate entities might receive the same kind of plural marking irrespective of the gender they are assigned to under singular reference. This effect is referred to as *semanticization of gender distinctions*. Thus, syncretic genders tend to be more semantically transparent than non-syncretic genders. Crosslinguistic evidence is presented in support of this claim both from languages with large and small gender systems (in terms of number of distinctions). I shall discuss this in greater detail in chapter 5.

2.5.3 Cumulative exponence, syncretism and relevance hierarchies

Another classic typological generalisation concerning the relationship between gender and number is Greenberg's Universal 36:

> If a language has the category of gender, it always has the category of number. (1963b: 102-103)

Universal 36 establishes an implicational relationship between the two grammatical domains that, according to Greenberg, is also related to markedness. In this view, markedness relationships exist not only within but also across grammatical domains. Greenberg suggests that markedness relationships across domains can be formalized in the form of *proximity hierarchies*. According to Greenberg, proximate grammatical categories tend to be encoded closer to the root or to the stem of the word they attach to, and are functionally more relevant than non-proximate ones. They also tend to be morphologically more elaborated and crosslinguistically more frequent. Given a grammatical domain, the existence of a less proximate category in one language implies that of the more proximate one. With respect to nominal categories, Greenberg suggests that grammatical gender (as encoded on the indexing targets) and case are less proximate than number. Thus, the presence of gender (and/or case) in one language implies the existence of number (Universal 36). Moreover, phenomena of syncretism are more likely to affect the implying and less proximate categories rather than the more proximate ones (Universal 37, see §2.5.2). Greenberg's notion of *proximity hierarchies* for nominal categories closely resembles the notion of *relevance hierarchy* elaborated later by Bybee (1985) with respect to verbal categories. As far as I know, however, Greenberg's notion of proximity hierarchy has not had any follow-up in the literature.

Twenty years after Greenberg's work, Andrew Carstairs-McCarthy published two studies on case syncretism in which he suggested that a relevance hierarchy of the type proposed by Bybee (1985) for the verbal domain may also exist for nominal features. Carstairs (1987) examines a set of 43 patterns of syncretism (mostly involving case and number) from 16 different languages and shows that syncretic paradigms are almost always also cumulative paradigms. In Carstairs & Stemberger (1988), computational evidence is brought in support of the crosslinguistic data examined by Carstairs (1987). Beside providing evidence for the fact that case syncretism in the context of number is more likely when the two domains have cumulative exponence, Carstairs (1987) and Carstairs & Stemberger (1988) also tentatively suggest that syncretism is one of the ways in which the nominal relevance hierarchy reflects on the structure of paradigms. Number is higher on the relevance hierarchy, and thus it is less prone to undergo syncretism. On the other hand, case is located at a lower level in the hierarchy and is thus more likely to undergo syncretism in the context of number.

Similar conclusions were drawn much later by Vafaeian (2013) in a study focussing on a different grammatical phenomenon: nominal suppletion. In her typological investigation of patterns of adjectival and nominal suppletion, Vafaeian shows that number is the nominal category that is more likely to trigger suppletion in nominal paradigms crosslinguistically, followed by possession, vocative and accusative/ergative case. Since gender is an inherent property of nouns whose grammatical evidence is only displayed through indexation, it falls outside the scope of nominal suppletion (Vafaeian 2013: 123). Taking inspiration from Veselinova (2006), where suppletion on verb paradigms is shown to illustrate the effect of the relevance hierarchy on verbs, Vafaeian (2013) suggests that her results on nominal suppletion might also be considered to reflect a hypothetical relevance hierarchy for nouns. Following this suggestion, number would have a higher position in the hierarchy than the other grammatical features considered in the study.

To summarise, by exploring different grammatical phenomena (syncretism, cumulation and suppletion), and on the basis of independent datasets, Greenberg (1963b), Carstairs (1987), Carstairs & Stemberger (1988), and Vafaeian (2013) suggest the existence of a relevance hierarchy for nouns. These studies also agree that, within a hypothetical nominal relevance hierarchy, number has greater importance to nouns than other features – i.e., grammatical gender and case (Greenberg 1963b), case (Carstairs 1987; Carstairs & Stemberger 1988), possession, vocative and accusative/ergative case (Vafaeian 2013). Apart from Greenberg (1963b), I do not know of any study that has specifically focussed on relevance effects in the relationship between gender (as marked on the indexing targets) and number.

2.5.4 The interaction between gender and number in the languages of Africa

The relationship between gender and number in the African languages has been discussed in the context of individual languages or individual genealogical groupings within the macro-area. Reference to these studies is made throughout the dissertation. In this section, I focus on a few studies that are particularly relevant to the formulation of my

research questions (see §2.8).

In their (2008) paper on Africa as a morphosyntactic area, Creissels et al. survey patterns of areal convergence in African languages by taking into consideration a set of morphosyntactic features that include gender and number. The following claims are made concerning the relationship between the two grammatical domains in the African macro-area:

> (a) Languages devoid of a gender system frequently have a single plural marker with the morphological status of a phrasal affix, and such plural markers tend to be used on a "pragmatic basis", i.e., to be employed only when plurality is both communicatively relevant and not implied by the context, at least in the case of nouns that do not refer to persons [...].
>
> (b) Languages that have gender generally have a morphologically complex plural marking, characterized by a fusion of gender and number markers, and variation in gender and number manifest themselves through morphemes affixed to the head noun and to (some of) its modifiers, in an agreement relationship. In these languages [...] plural markers tend to be present in every NP referring to a plurality of individuals [...]. (Creissels et al. 2008: 119)

Ultimately, what Creissels et al. (2008) claim is that there exists some sort of correlation between presence or absence of grammatical gender and type of encoding of nominal plurality and that this correlation is spread throughout the African macro-area. The major limitation of these claims is that they are not formulated on the basis of the investigation of a stratified sample of African languages. Rather, they are based on the authors' personal knowledge of African languages. As formalized in §2.8, testing the validity of these generalizations on a stratified sample of African languages is one of the goals of this dissertation.

Less broad in scope but general enough to be anticipated here is the question of whether or not gender and number compete in the indexation domain. This question holds particular relevance for those languages in which dedicated patterns of plural indexation are not used with all plural nouns but with a restricted subset of plural nouns. A case in point in the literature on African languages are, for instance, Cushitic (Afro-Asiatic) languages. It is a tradition among experts of Cushitic to consider dedicated patterns of plural indexation as a manifestation of a gender value, the so-called "plural gender" (see Mous 2008: for a general overview). This issue is discussed in detail in chapter 5.

2.6 The interaction between gender and evaluative morphology: state of the art

The relationship between grammatical gender and evaluative morphology has not been as widely investigated as the one between gender and number. A review of the state of

research in this field is thus bound to be short.

In the literature on noun classification, size-related meanings are nearly unanimously mentioned among the possible semantic values of a gender system (Allan 1977; Corbett 1991; Croft 1994; Aikhenvald 2003). However, it is also noticed that, among the most typical semantic underpinnings of gender - i.e., sex, animacy, shape, size - size is the least likely to occur as an independent criterion for the categorisation of nouns. Interestingly, in his seminal work on noun classification systems, Allan (1977: 303) claims that noun classes and gender distinctions based on size *alone* "only appear in African languages [...]."[19]

In languages in which gender marking is functional to the encoding of size, gender assignment is not rigid. Variation in size can be expressed by shifting a noun from one gender to the other (see notions of *free/manipulable* vs. *fixed/rigid* gender assignment and *recategorization* as discussed in §2.1.1). Gender shifts may also express qualitative evaluative meanings, such as *bad* vs. *good*. It has been observed that in languages with sex-based gender, noun class and gender shifts of this type are driven by stereotypical semantic associations between each sex and each size value (Croft 1994; Aikhenvald 2003). For example, many speech communities associate 'female' with 'small' and 'male' with 'big.' Finally, it has been noticed that the extent to which gender can be used to mark physical properties of nouns, among them size, depends on the animacy of nouns (Aikhenvald 2003). Nouns are more likely to be classified according to physical properties if they denote inanimate entities. As mentioned in §2.1.1, Heine (1982) reports that gender shifts of this type are very common in African languages. He also mentions that the possibility of assigning nouns to multiple genders is not totally unconstrained in African languages and that, at least for some nouns, assignment preferences do exist. The nature and scope of these assignment preferences is, however, left unspecified in his paper.

To date, neither the nature of the relationships between gender and evaluative morphology, nor the variables that determine them, have been investigated in a systematic fashion. One of the few exceptions is Grandi (2001), in which the relationships between evaluative morphology and gender, on the one hand, and evaluative morphology and number, on the other, are investigated using data from Greek, Romance and South Slavonic languages. See also Grandi (forthcoming-b) for general typological overview of patterns of interaction between gender and evaluative morphology (as well as evaluative morphology and number). The use of gender shifts as a means of encoding biological gender and size in Moroccan Berber has been recently discussed by Kossman (2014).

2.7 Grammatical complexity of gender systems

The study of language complexity is a very fascinating and controversial field of research within linguistics. For an overview of some of the most recent approaches to the study of language complexity, see Dahl (2004) as well as the different contributions in the volume edited by Miestamo et al. (2008). Two main approaches to the study of language

[19] In Allan's (1977) paper, the noun class systems typical of African languages are referred to as *classifiers*.

complexity currently exist: the *relative* approach and the *absolute* approach. Within the relative approach, language complexity is assessed from the perspective of its users, that is, in terms of processing, learnability and usage (see, among others, Kusters 2003, 2008). Within the absolute approach, which I follow in this dissertation, language complexity is viewed as an objective property of a given system rather than as a measure of the costs and difficulties that the users of that system experience when manipulating it (see, among others, Dahl 2004; Miestamo 2008; Sinnemäki 2011).

Taken in this absolute sense, grammatical complexity is defined "as the number of parts in a system or the length of its description" (Miestamo 2008: 28). Methodologically, Miestamo also suggests, given the practical impossibility to measure the grammatical complexity of a language as a whole, that metrics for measuring grammatical complexity crosslinguistically be *local* (rather than *global*) in their scope. Typologists should focus on individual functional domains and compare them across languages, trying to account for their "minimal, average and maximal complexity" (2008: 31). One of the possible outcomes of an enterprise of this type, he continues on the same page, is to enable typologists to examine "typological correlations between the complexities of different domains" and inspect whether or not "complexity in one domain is compensated by simplicity in another." Miestamo (2008) and Sinnemäki (2011) suggest two major criteria for measuring complexity of individual grammatical domains:

> *The Principle of Fewer Distinctions.* This principle is concerned with the number of obligatory distinctions within a given functional domain: the higher the number of grammaticalized distinctions, the more complex the domain.
>
> *The Principle of One-Meaning–One-Form.* This principle is concerned with the relationship between meaning and form within a particular functional domain. Discontinuous morphemes or portmanteau morphemes represent a typical violation of the One-Meaning–One-Form principle. In the case of discontinuous morphemes, one meaning is distributed over a plurality of forms. In the case of portmanteau morphemes, one form is associated with a plurality of meanings.

Modelling absolute complexity of individual grammmatical domains is still a novel field in language typology. Yet, some work has already been initiated, and gender systems figure as one of the domains under investigation. Audring (2014) proposes a set of cues to assess the grammatical complexity of the gender systems of individual languages. The study is mostly a general programmatic discussion of possible ways of exploring the grammatical complexity of gender and is not based on a stratified language sample. However, different types of gender systems from different areas of the world are referred to throughout the paper. One of the ultimate purposes of Audring's work is to explore whether and, if so, how the absolute complexity of gender systems relates to their learnability and processing from the perspective of language users (i.e., *relative complexity*). Toward this aim, Audring (2014) individuates three dimensions for assessing the absolute complexity of gender systems:

I The number of gender values

II The number and nature of assignment rules

III The amount of formal marking (i.e., number of gender-indexing targets).

The gender system of English would rank low with respect to all three dimensions: it has three gender values, minimal indexation (the only indexes of gender agreement being pronouns), and the assignment rules are both semantic and broad in scope (males are masculine, females are feminine and inanimate nouns – with a few exceptions – are neuter).

Concerning mutual relationships between the three dimensions, Audring (2014) focusses on the way in which I and II interact with III. She argues that pervasive gender indexation (III) is expected to be found in languages with a high number of gender values (I) and/or complex assignment rules (II). This is due to the fact that complexity in formal marking (pervasive indexation) facilitates the processing of grammatical gender in terms of learnability: the more manifestations of gender there are in discourse, the easier it is to learn gender values and assignment rules.

Neither of the two principles suggested by Miestamo (2008) and Sinnemäki (2011), nor the three dimensions of gender complexity suggested by Audring (2014), explicitly tackle the issue of how morphosyntactic interactions between grammatical domains affect the absolute complexities of the individual domains. Audring (2014) mentions cumulative exponence (with number and/or other features of nominal morphology) and syncretism as two factors that increase the absolute grammatical complexity of gender systems at the level of formal marking (dimension III). However, within her model, it is not clear how the occurrence of the two phenomena is to be counted in the assessment of the overall grammatical complexity of gender systems. The issue of language complexity and interactions between grammatical domains, as well as its specific relevance to gender, are extensively dealt with in chapter 7.

Previous studies on the role that cumulation and syncretism play in paradigm structures have mostly discussed their implications on the learnability and processing of language. In §2.5.3, I mentioned two studies by Carstairs (1987) and Carstairs & Stemberger (1988) with respect to the notion of nominal relevance hierarchy. The results of these studies are also interpreted by the authors in terms of language processing. The occurrence of syncretism is said to have a *simplifying* effect on cumulative paradigms in terms of learnability and usage. Similarly, Plank & Schellinger (1997) reflect on the role of markedness in the management of paradigm sizes and of the relationships between inflectional features of grammar:

> [I]nflectional systems ought to remain within certain *limits of complexity*, and the problem that may arise then is how to cut down on them when a language indulges in inflectional elaboration to the extent that these limits are in danger of being exceeded. Quantitatively, the obvious measure to take is simply *not to have as many distinct forms as there are paradigmatic distinctions* – which is the rationale for *syncretism* and defectivation, especially where the exponents of two or more categories are expressed *cumulatively* rather than separately. (Plank & Schellinger 1997: 58; emphasis mine)

The notion of complexity that Plank & Schellinger work with is also of the procedural type: constraints on paradigm sizes are seen as optimising the way speakers handle inflectional morphology.

How can cumulation and syncretism be accounted for within an absolute approach to language complexity? When two obligatory features have cumulative exponence, the overall number of forms in the paradigm gets reduced. At the same time, since one marker is associated with at least two grammatical meanings, the mapping between form and function is less straightforward than in the case of noncumulative exponence (Principle of One-Meaning-One-Form). Ultimately, cumulation decreases linearity in the internal structure of morphemes.

Dahl (2004) examines patterns of syncretism in the context of a broader discussion of the development of nonlinear structures in language (the label *featurization* is used to refer to such processes). According to Dahl, the occurrence of syncretism presupposes the lack of a one-to-one correspondence between surface and underlying language structures ("[i]n general, syncretism implies that the relationship between surface structure and underlying representations is not one-to-one," 2004: 188). The occurrence of syncretism *per se*, however, is not necessarily equal to an increase in absolute complexity (Östen Dahl, personal communication). This interpretation will be followed and developed throughout the dissertation.

Semantic interactions between gender and number and gender and evaluative morphology are, to date, a very poorly investigated field in the study of gender systems complexity. I believe that the notion of gender shift is essential for exploring and classifying these types of interactions. This will be thoroughly discussed in chapter 7.

2.8 Research questions

This dissertation aims to answer the following questions:

(2.13) Research questions concerning gender and number

- Q 1: How common is cumulative exponence of gender and number in the languages of the sample?
- Q 2: What are the formal and semantic factors that trigger gender syncretism in the context of number? Does gender syncretism in the context of number presuppose cumulative exponence?
- Q 3: What are the implications of cumulative exponence and syncretism on the absolute complexity of gender and number systems?
- Q 4: Can these types of interaction between gender and number be seen as a reflex of a nominal relevance hierarchy?
- Q 5: Can gender and number compete through indexation patterns?
- Q 6: Is there any correlation between types of encoding of gender and types of encoding of number?

Q 7: What type of semantic interactions can be found between gender and number?

(2.14) Research questions concerning gender and evaluative morphology

Q 8: How frequently does *size* occur as an independent gender value? How stable and how widely distributed is this phenomenon within genenealogical units?

Q 9: Do the interactions between gender and evaluative morphology differ across types of gender systems and/or strategies of gender assignment (e.g., *sex-based* vs. *non-sex-based* gender systems, or *rigid* vs. *manipulable* gender assignment)?

(2.15) Research question concerning all three domains

Q 10: How do interactions between gender and number and gender and evaluative morphology affect the grammatical complexity of gender systems? Is it possible to measure the role that these interactions play in the absolute complexity of individual gender systems?

In the next chapter, the methods of data collection and analysis that I followed in the attempt to answer these research questions are discussed. Questions 1 to 7 are addressed in chapter 5, whereas questions 8 and 9 are addressed in chapter 6. Finally, question 10 is approached in chapter 7.

2.9 Summary of the chapter

The aim of this chapter has been to introduce the grammatical domains investigated in this dissertation: gender, number and evaluative morphology. Each of the domains has been defined along four dimensions: semantics and functions, morphosyntax, typological distributions and diachrony.

Previous studies on the interactions between gender and number, and gender and evaluative morphology, as well as on the grammatical complexity of gender systems have been surveyed in the second part of the chapter.

Finally, the research questions addressed in the dissertation have been outlined and situated in the context of previous and ongoing debate on the typology of gender systems, their interactions with other domains of grammar and their absolute complexity.

3 Method: language sampling, data collection and organization

The aim of this chapter is to illustrate the sampling methodology followed in the dissertation and to relate it to the research questions outlined in §2.8. The chapter is structured as follows. The sampling methodology designed for this thesis is described in §3.1. In §3.2, I discuss the state of the art on the genealogical classification of African languages, whereas major linguistic areas within Africa are surveyed in §3.3. A practical description of the procedure of language selection is found in §3.4. Data collection and organization are briefly discussed in §3.5. A summary of the chapter is presented in §3.6.

3.1 Sampling methodology

Based on the configuration of its sample, this dissertation may be classified as an instance of what Haspelmath (2012) calls *continent-wide typology*. The label refers to typological studies that restrict their sample space to the continent level, which, in this specific case, is Africa. It does not refer to a specific sample procedure but is merely used to describe a recent trend in typological studies towards continent-based language samples. One could also refer to studies of this type as typological investigations of macro-areas, in Dryer's (1989a; 1992) terms. Dryer describes macro-areas as large geographic groupings consisting of multiple genera. He defines a genus as a genealogical unit "roughly comparable to the subfamilies of Indo-European, like Germanic and Romance" (1989a: 267). The five macro-areas identified by Dryer (1989a) are defined on the basis of continental zones: Africa, Eurasia, Australia-New Guinea, North America and South America.

Sampling is a crucial and thoroughly debated issue in typological research (for an overview of the specific challenges of language sampling in linguistic typology, see Bakker 2011). As in any kind of investigation based on empirical data, the construction of a typological sample must reflect the nature of the research questions a study aims to answer (Bakker 2011: 106). As outlined in §2.8, this dissertation explores the interactions that occur between gender and number, and gender and evaluative morphology. I am interested in accounting for the *types* and *nature* of possible interactions between these domains, as well as their *distribution* and *stability* throughout linguistic areas and genealogical units. Toward this aim, I decided to build my sample with the purpose of combining *intra-* and *intergenealogical* observations of a restricted, and somewhat self-consistent, area of the world. The following considerations led me to elaborate a sampling design of this type:

(1) Via intragenealogical comparison we can map synchronic distributions of linguistic

phenomena throughout closely related languages. On the basis of such distributional data, we can attempt to estimate the stability of the observed phenomena. In the best-documented cases, we may also infer aspects of the diachronic evolution of the phenomena under investigation. This procedure has been extensively practised in typology since its early years (see the study conducted by Greenberg 1980 on word order in Semitic Ethiopian and Iranian languages) but has become more popular within the last couple of decades. Croft (2003: 247) describes intragenealogical typology as the application of the historical-comparative method for typological purposes, and as the basic methodological ground for diachronic typology, which he also refers to as *dynamic typology*.

(2) Via intergenealogical comparison we are able to account for the distribution of a linguistic phenomenon throughout unrelated languages, as well as to illustrate and motivate the limits of crosslinguistic variation with respect to that phenomenon. When practised on a restricted geographic area, this method may lead to the discovery of patterns of areal convergence that cut across genealogical affiliation. Research in areal typology has been very productive in the past two or three decades. As Koptjevskaja-Tamm (2011: 577-578) puts it, one of the major contributions of areal typology to the field of linguistic typology is to show that "it can be of limited value to search for a possible human language without simultaneously investigating its genetically and areally determined manifestations and trying to uncover the possible historical reasons behind this variation."

The dataset that resulted from this combined sample procedure consists of 100 languages selected from different genealogical groupings of Africa.[20] For each of the selected genealogical units, a convenience subsample was created; as a rule of thumb, the number of languages for each subsample was decided on the basis of the size of the individual genealogical units.[21] The availability of descriptive materials on a given language also played a role in the process of language selection. A detailed description of the internal structure of the language sample and the principles that guided language selection is found in §3.4.

The choice of Africa as the sampling area was directly connected to the research questions that the dissertation aims to answer (see §2.8). One of the most important criteria for a language being included in my sample is the presence of grammatical gender in the language. Given that Africa has been defined as one of the world's hotbeds of grammatical gender (see, among others, Corbett 2013b; Nichols 1992: 130), I decided to confine my research to this area of the world. As shall be illustrated in detail in chapter 4, 84 languages in the sample have grammatical gender. The remaining 16 languages do not have grammatical gender. They have been included in the sample as control

[20] The language sample also contains languages such as Hebrew and Maltese, which are actually spoken outside Africa. As Dryer (1989a: 268) points out, all Semitic languages can be seen as part of the same large linguistic area because "their genetic relationships go in that direction."

[21] This method is likely to result in genealogical biases and is thus not suited for large-scale typological samples. See the discussion in Veselinova (2006: fn 2).

languages in order to investigate whether any aspect of the grammatical encoding of number and evaluative morphology correlates with the presence or absence of gender.

Before exploring the details of the structure of the language sample created for this thesis, I give an overview of the current state of the art on the genealogical classification of African languages and the major linguistic areas within Africa.

3.2 African languages: genealogical classification

According to the most recent counts, about 2000 languages are estimated to be currently spoken in Africa (Heine & Nurse 2000; Mous 2003b). These figures need to be taken with caution for more than one reason. First of all, as noticed by Heine & Nurse (2000), many African languages are still underdescribed and others are dying as their last native speakers die. In addition, different figures may result depending on where the line between languages and dialects is drawn, which is often done on the basis of sociopolitical rather than linguistic grounds.

The genealogical classification of the African languages proposed by Greenberg in 1963 has been for several decades undisputed both among africanists and typologists. Greenberg (1963a) divides the languages of Africa into the following four macro-families: Niger-Kordofanian (or Congo-Kordofanian), Nilo-Saharan, Afro-Asiatic, and Khoisan. All further subdivisions within each of the four families are listed in figure 3.1 (next page).

In recent years, and on the basis of extensive historical-comparative investigation, Greenberg's classification has been strongly criticised by specialists of individual languages and genealogical units of Africa. The effects of this debate on the typological community have apparently been relatively minor, considering that some of the canonical typological resources, as the WALS database (Dryer & Haspelmath 2013) and Ethnologue (Lewis et al. 2013) are still quite faithful to Greenberg's classification with respect to the sampling of the African macro-area.

As Dimmendaal (2008: 850) puts it, among the four language families posited by Greenberg, Afro-Asiatic is the one whose classification is the most secure. With respect to the other three macrogroups, it is now believed that more genealogical diversity than the one assumed by Greenberg needs to be posited. In the rest of this section, I provide a short survey of the latest developments in the genealogical classification of the African languages, mostly based on the useful overview by Dimmendaal (2008). I especially consider those aspects of the ongoing debate which have been more crucial in the creation of my own sample.

Most africanists would nowadays agree that one of the weakest and most problematic points in Greenberg's (1963a) classification is the idea of Khoisan as a homogeneous genealogical unit (see, for example, Güldemann 2008). Greenberg believed that the Khoisan family included all the click languages of Africa with the exception of those classified as Bantu (e.g., Xhosa) or Cushitic (e.g., Dahalo). Today, three independent language families and two isolates are posited under the label *Khoisan*. The term *Khoisan* is still used as a label of convenience by which no genealogical reality is intended. Table

- Congo-Kordofanian (= Niger-Kordofanian)
  - Niger-Congo
    - (West) Atlantic
    - Mande
    - Voltaic (= Gur)
    - Kwa
    - Benue-Congo
    - Adamawa-Eastern (= Adamawa-Ubangi)
  - Kordofanian
- Nilo-Saharan
  - Songhai
  - Saharan
  - Maban
  - Fur (Isolate)
  - Chari-Nile
    - Eastern Sudanic
    - Central Sudanic
    - Berta (Isolate)
    - Kunama
  - Koman
- Afro-Asiatic
  - Semitic
  - Egyptian
  - Berber
  - Cushitic
  - Chadic
- Khoisan
  - Southern African Khoisan
    - Northern
    - Central
    - Southern
  - Sandawe (Isolate)
  - Hadza (Isolate)

Figure 3.1: Greenberg's (1963a) genealogical classification of the African languages

3.1 summarises the current state of the art on the genealogical relationships among the languages of the Khoisan area. The classification presented in the table is the one followed by the Glottolog database (Nordhoff et al. 2013) and is based on some of the most recent contributions in the field of historical-comparative reconstruction on the Khoisan area. See appendix A for a complete list of the languages included in the sample from the individual genealogical units listed in table 3.1.

Table 3.1: Genealogical relationships within Khoisan

Genealogical Unit	**Level**
Hadza	Isolate
Sandawe	Isolate
Khoe-Kwadi	Family
Kxa	Family
Tuu	Family

Another important revision of Greenberg's classification concerns the place of Omotic within Afro-Asiatic. When Greenberg elaborated his classification of the African languages, Omotic was considered a subgroup of Cushitic and was commonly referred to as West Cushitic. Fleming (1969) was the first to propose considering the Omotic/Western Cushitic languages as an independent subgrouping directly descending from Afro-Asiatic. Greenberg never adjusted his classification accordingly. Nowadays, the different subdivisions within Omotic are classified either as independent subgroupings within Afro-Asiatic or as independent languages families with no relationship to Afro-Asiatic. By the latter interpretation, shared features with other Afro-Asiatic languages or between languages of different Omotic families are explained as the result of language contact. The Afro-Asiatic membership of Omotic, the genealogical relationships between the different Omotic subgroupings and their possible position within the rest of the family are the topic of a very big, and as yet unresolved, controversy within Afro-Asiatic linguistics (for an overview of different classifications of the Omotic languages, see Amha 2012).

The genealogical relationships within Greenberg's Niger-Kordofanian macrogroup (i.e., Niger-Congo or, following Glottolog, Atlantic-Congo[22]) have also been thoroughly revisited in recent years. Although I cannot go into the details of this debate here, a few major points need to be clarified.

First of all, since the evidence supporting the affiliation of the Mande languages with the Atlantic-Congo family has proved to be scarce, most scholars today prefer to consider these languages as an independent group outside Atlantic-Congo (Dimmendaal 2008: 842).[23]

[22]It is worth mentioning that the two terms – *Niger-Congo* and *Atlantic-Congo* – are not synonymous. Certain subgroupings that are traditionally counted as Niger-Congo (e.g., Mande) are not conceived as part of the Atlantic-Congo group in the Glottolog (see further discussion in this section). The different labelling reflects non-overlapping classifications.

[23]Dimmendaal (2008) also considers the Ubangi languages to be outside Atlantic-Congo, but this is not

Second, the status of the Atlantic languages within Atlantic-Congo is still vigorously debated. The individual subgroupings within Atlantic are so different from each other that, similarly to what has been observed for Khoisan, some scholars have gone as far as to suggest that there is no such thing as an Atlantic genus within Atlantic-Congo. According to these scholars, the label *Atlantic* should be taken as a cover term for a language area rather than for a genealogically coherent grouping (for a brief discussion, see Dimmendaal 2008: 842). This attitude is reflected in recent work by Segerer (2010), who, based on cognacy judgements (mainly) of lexical items, establishes a new genealogical classification of the Atlantic languages. Three major genealogically coherent subgroupings are distinguished: Northern Atlantic, Bak, and Mel (former Southern Atlantic). Gola and Limba, traditionally conceived of as part of the southern subbranch, are classified as isolates within Atlantic-Congo. Accordingly, the term Atlantic is retained only for its areal significance, whereas the affiliation of these languages to the Atlantic-Congo family is unquestioned.

Finally, no agreement has been reached on the internal stratification of the other genealogical units within Atlantic-Congo; their status as members of the family is, however, generally not questioned (Dimmendaal 2008: 843).

The status of Nilo-Saharan as a linguistic unit within Africa is also considered to be very controversial. As in the case of the other macrofamilies, many scholars have harshly criticised Greenberg's lumping approach and have instead enhanced the genealogical diversity of the languages grouped under the *Nilo-Saharan* label. Interestingly, Dimmendaal (2008: 843-844), himself an expert on Nilo-Saharan, defends Greenberg's work on the classification of the Nilo-Saharan languages. He claims that establishing Nilo-Saharan as a language family was Greenberg's "most important contribution to the classification of African languages." Yet, based on new evidence from the most recent descriptive and comparative research, Greenberg's genealogical tree for Nilo-Saharan has also been revised. In particular, two subbranches, Songhai and Koman, are no longer considered as part of the family but as independent genealogical units (Dimmendaal 2008).

To summarise, as conceived today, the genealogical classification of the languages of Africa assumes more diversity than Greenberg did in 1963. This is evident in the map presented as figure 3.2, which illustrates the genealogical classification of Africa proposed by Dimmendaal (2008: 245). As shown earlier in this section, no consensus has been reached yet on the internal stratification of some of the individual genealogical units (e.g., Atlantic within Atlantic-Congo). In fact, more comparative and descriptive work is needed in order to reach a better understanding of the nature of the mutual relationships between individual languages and language groupings within the African macro-area. It is also reasonable to think that in some areas of the continent, genealogical diversity used to be even greater in the past and that its decrease came as a consequence of the expansion of certain speech communities over others (e.g., the expansion of the Bantu languages towards the Khoisan area). An overview of the ecological and sociohistori-

widely accepted; on the other hand, many scholars agree on the idea that Ubangi does not constitute a coherent group within Atlantic-Congo (Harald Hammarström, personal communication).

Figure 3.2: African language families and isolates according to Dimmendaal (2008). Copyright [Dimmendaal (2008)]; this material is reproduced with the permission of John Wiley & Sons, Inc.

cal foundations of the distribution of linguistic diversity within Africa can be found in Dimmendaal (2008: 845-855).

3.3 African languages: large-scale language contact

As observed in the previous section, historical-comparative work on African languages has almost exclusively focussed on establishing genealogical relationships among individual languages and groups of languages within the continent. On the other hand, little has been done on the investigation of large-scale language contact and areal convergence within the African continent (Güldemann 2010; Heine & Nurse 2008).[24] Early attempts at investigating the distribution of linguistic features within Africa as a result of areal convergence are Greenberg (1959, 1983) and Heine (1975, 1976). Greenberg (1959, 1983) attempts to establish a set of features whose distribution identifies the core of Africa as a linguistic area. These features cut across different domains of grammar and the lexicon (phonology, morphology, lexical semantics). As pointed out by Güldemann (2010), one of the major shortcomings of Greenberg's investigation is the great overlapping between the four linguistic areas identified through this method and the four linguistic macrofamilies identified in his (1963a) work on genealogical relationships among African languages (see discussion in §3.2). Heine (1975, 1976) instead focusses on the investigation of areal patterns in the distribution of word order types based on a large sample of African languages. He identifies four major word order types whose distribution is, at in least in part, areally skewed (for an overiew, see Güldemann 2010).

Over the last decade, research on large-scale language contact within the African continental area has received renewed attention. A seminal publication in this sense is the volume edited by Bernd Heine and Derek Nurse in 2008, *A Linguistic Geography of Africa*. The general assumption behind the individual contributions in the volume is that unravelling the dynamics of large-scale language contact within Africa can be crucial for understanding structural properties of African languages and their distribution (Heine & Nurse 2008: 2). The volume is organized as a set of two types of case studies: (1) case studies of individual grammatical phenomena and their distribution throughout Africa (see, for instance, Christa König's chapter on the nominative-marked languages of eastern Africa); (2) case studies of individual linguistic areas and their characteristic features (see, for instance, Tom Güldemann's chapter on the Macro-Sudan belt linguistic area). Two other chapters are more general in their scope and discuss patterns of areal convergence between African languages at the levels of phonology (Clements & Rialland 2008) and morphosyntax (Creissels et al. 2008).

[24] Studies of language contact at the micro-areal level have always been very productive in African linguistics. African speech communities are often highly multilingual, and studies of this type have especially focussed on describing and modelling the social dynamics of language contact among two or more speech communities (see, among many others, the studies by Heine 1970 on African lingua francas and the volume edited by McLaughlin 2009 on multilingualism in African urban areas).

An important contribution to the identification of large-scale linguistic areas within Africa comes from accomplished and ongoing research by Tom Güldemann, who has worked extensively on this topic during the past fifteen years. In his work, Güldemann has focussed on the characterization of specific linguistic areas within Africa (see, for instance, Güldemann 1998 on the Kalahari-Basin area, and Güldemann 2008 on the Macro-Sudan belt) as well as on the overall classification of the languages of Africa in terms of areal rather than geneological groupings. The macro-areal profile of the African content is investigated by Güldemann (2010). Methodologically, the paper takes into account the distribution of two types of linguistic features: typologically rare phenomena that are remarkably frequent in the African macro-area as opposed to the rest of the world (e.g., click phonemes), and typologically less rare features that are nevertheless fairly typical of African languages (e.g., nasal vowels). The following macro-areas are identified by Güldemann (2010) within Africa:

I Berber spread zone

II Chad-Ethiopia

III Macro-Sudan belt

IV Bantu spread zone

V Kalahari Basin

The geographic distribution of the individual areas is shown in the map in figure 3.3 (next page).

Core and peripheral members of each of the five areas are listed in Güldemann (2010) together with a survey of the grammatical features that are shared by the languages of each of the sub-areas. The Berber spread zone and the Chad-Ethiopia zone are the least genealogically diverse: all the languages assigned to these areas are, in fact, Afro-Asiatic. Possible correlations between the natural geography of Africa and the distribution of the five areas throughout the continent are also discussed in Güldemann's paper. For instance, it is pointed out that the northernmost and southernmost areas (i.e., the Berber spread zone and the Kalahari Basin) are less diverse in terms of genealogical composition and number of languages, whereas the highest language diversity is concentrated around the Tropics.

3.4 Sampling procedure

Data collection for this dissertation was conducted on the basis of established genealogical relationships between African languages. Accounting for areal groupings within the African macro-area became especially relevant when analysing the data and interpreting the results of the investigation.

Figure 3.3: Linguistic macro-areas within Africa according to Güldemann (2010). Copyright [Mouton de Gruyter]; this material is reproduced with the permission of Mouton de Gruyter

Since this thesis is mainly a typological investigation of the languages of Africa, I make no claims concerning the status of the genealogical classification of the languages of the macro-area. When establishing the coding conventions for my database, I used the language coding convention of Ethnologue (Lewis et al. 2013) because it is accessible and also widely used by other typological databases (e.g., WALS and Glottolog). With respect to language classification in the strict sense, I follow the classification provided by Glottolog (Nordhoff et al. 2013) because, in my opinion, it is the one that more closely reflects the most recent developments in the genealogical classification of African languages (see discussion in §3.2). The following levels of classification are used in Glottolog: *isolate*, *top-level family*, *subfamily*. The label *subfamily* is used to refer to all levels of classification below *top-level family*. Thus, when comparing Glottolog with WALS, a Glottolog subfamily can correspond either to a subfamily or a genus of WALS. For instance, in Glottolog, Benue-Congo and Bantoid are both labelled as subfamilies. Benue-Congo is counted as a subfamily within Atlantic-Congo (Niger-Congo in WALS) and Bantoid as a subfamily of Benue-Congo. On the other hand, in WALS, Benue-Congo is labelled as a subfamily of Atlantic-Congo and Bantoid as genus. "Practically, all WALS genera exist as subgroups in Glottolog" (Harald Hammarström, personal communication): the genealogical stratifications followed in the two databases are thus largely comparable.

Technically, the sampling procedure that I designed for this thesis approximates what is commonly referred to as a *variety sample* (Bakker 2011; Veselinova forthcoming). Variety samples are created with the purpose of investigating a linguistic phenomenon "in its greatest possible variability" (Veselinova forthcoming). Thus, for a language to be included in a variety sample, it needs to display the phenomenon under investigation in a given study. In this case, the presence of gender is the main condition for inclusion in the sample. However, my language sample differs from a canonical variety sample in that: (1) only one area of the world is investigated; (2) several languages from the same genealogical unit are considered; (3) languages lacking the variable under investigation (i.e., gender) are also included as a control group. I began by looking at those genealogical groupings of Africa whose gender systems had already been investigated in their interaction with number and evaluative morphology. I thus started by sampling Bantu and Cushitic languages. A vast amount of literature is available on the Bantu gender systems and interactions with number, and evaluative morphology have often been discussed. Similarly, the interaction between gender and number in Cushitic languages has been at the core of an intense scientific debate during recent decades. From this first nucleus, the sample was expanded to include:

(1) Additional genealogical groupings characterised by the presence of gender systems similar to the Bantu type (e.g., North-Central Atlantic languages and Sɛlɛɛ[25] within Kwa)

(2) Additional genealogical groupings characterised by the presence of gender systems similar to the Cushitic type (e.g., Chadic, Semitic and Eastern Nilotic)

[25]The spelling *Selee* is used in Glottolog (Nordhoff et al. 2013).

(3) Additional genealogical groupings characterised by the absence of gender (e.g., Mande and Western Nilotic, or Akan and Ewe within Kwa[26]).

A subsample was created for each of the selected genealogical units. These subsamples can be defined as convenience samples. I did not follow any mathematical procedure to establish the exact number of languages that should be selected from each subgrouping. Thus, the language sample created for this thesis cannot serve as a basis for statistical analysis of the *inferential* type, that is, to formulate predictions over preferred typological patterns. The statistical analysis that is applied to the data of the language sample is rather of the *descriptive* type: the frequency distributions of the attested crosslinguistic patterns are computed and presented in the results chapters. In chapter 7, correlation coefficients between the features of my complexity metric are presented. Also in this case, the figures are used as a means for exploring and describing the behaviour of my complexity measure rather than to formulate crosslinguistic generalizations (see chapter 7 for details).

Accessibility of descriptive resources and/or availability of contact with experts of individual languages played a major role in guiding language selection. Bigger subgroupings are generally represented by a higher number of languages. For instance, there are 23 Bantu languages and only 6 Berber languages in the sample. In the Glottolog database, 558 languages are classified as (Narrow) Bantu and 28 as Berber. The number of Bantu and Berber languages included in the sample is thus proportionate to the size of the respective genealogical units. In addition, I tried to organise each of the subsamples in a way that could reflect the internal diversity of a given genealogical unit. For instance, in the case of the Cushitic languages, I made sure to include at least one language from each of the internal sub-groupings within the sub-family: Beja, South, East and Central (see Nordhoff et al. 2013 and appendix A).

The individual genealogical units represented in the sample are shown in table 3.2. The table provides the following pieces of information:

(1) Name of the genealogical unit

(2) Level of genealogical classification, according to Glottolog

(3) Areal grouping that a unit belongs to (if applicable)

(4) Superordinate genealogical grouping that a unit belongs to, according to Glottolog (if applicable)

(5) Number of languages per genealogical unit.

It is worth mentioning that, in the case of the Omotic languages, no superordinate genealogical grouping is claimed in Glottolog. However, since, as discussed in §3.2, the genealogical relationships between the Omotic groups and their affiliation to Afro-Asiatic

[26] Nominal prefixes are attested both in Ewe and Akan as remnants of a now estinct gender system. On the loss of the noun class system in Akan, see, for instance, Osam (1993).

Table 3.2: Genealogical units in the sample

	Genealogical unit	Glottolog level of classification	Areal grouping	Supeordinate genealogical unit	No. of lngs.
	Berber	Subfamily	–	Afro-Asiatic	6
	Chadic	Subfamily	–	Afro-Asiatic	8
	Cushitic	Subfamily	–	Afro-Asiatic	13
	Semitic	Subfamily	–	Afro-Asiatic	7
	Dizoid	Top-level family	Omotic ?	Afro-Asiatic ?	1
	South Omotic	Top-level family	Omotic ?	Afro-Asiatic ?	1
	Ta-Ne-Omotic	Top-level family	Omotic ?	Afro-Asiatic ?	4
	(Narrow) Bantu	Subfamily	–	Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	23
	Defoid	Subfamily	–	Atlantic-Congo	1
	Igboid	Subfamily	–	Atlantic-Congo	1
	Kwa	Subfamily	–	Atlantic-Congo	3
	Mel	Subfamily	Atlantic	Atlantic-Congo	3
	North-Central Atlantic	Subfamily	Atlantic	Atlantic-Congo	7
	Eastern Nilotic	Subfamily	–	Nilotic	3
	Western Nilotic	Subfamily	–	Nilotic	6
	Mande	Top-level family	–	–	4
	Khoe-Kwadi	Top-level family	Khoisan	–	5
	Kxa	Top-level family	Khoisan	–	1
	Tuu	Top-level family	Khoisan	–	1
	Hadza	Isolate	Khoisan	–	1
	Sandawe	Isolate	Khoisan	–	1
Total					100

are still debated issues among specialists, I added *Omotic* as an areal cover term and *Afro-Asiatic* as a superordinate genealogical grouping. A question mark is added next to the two labels to signify that their use is not established.

A complete list of the languages of the sample, the lower levels of classification and sources are provided in table A.1. An alphabetical index of the languages of the sample is presented in table A.2.

3.5 Data collection and organization

Information about the languages to be included in the sample was gathered by attempting at combining three different types of sources:

(1) Reference grammars and other types of descriptive materials

(2) Consultation of experts of individual languages and language families

(3) Consultation of native speakers.

Unfortunately, not all three sources were available for each of the sampled languages. For all the languages in the sample, data were gathered at least on the basis of reference grammars and descriptive materials of various types (e.g., dictionaries, articles in academic journals, teaching resources). Experts of individual languages and language families were consulted for all the genealogical units represented in the sample, except Igboid, Defoid and the Isolate Sandawe. For these languages, only descriptive materials were used. A two-week research visit at the French research institute LLACAN (*Langage, Langue et Culture d'Afrique Noire*) was instrumental in collecting data on the following genealogical groupings: Bantu, Berber, Cushitic, Dizoid, Mel, North-Central Atlantic, South Omotic, and Ta-Ne-Omotic. Only for two of the sampled languages, Amharic (Semitic) and Ewe (Kwa), was it possible to rely on the direct judgments of native speakers. Desalegn Hagos Asfawwesen (Stockholm University) helped me with Amharic and Yvonne Agbetsoamedo (Stockholm University) with Ewe. A complete list of the sources corresponding to each of the languages in the sample is found in appendix A.

The data were collected with the support of a coding sheet that is provided in appendix B. For each of the sampled languages, data collected through the coding sheet were then stored in a relational database. The database (which I call *Gender, Number and Evaluative Morphology in Africa*, henceforth GNEAF) is organised as follows. Basic linguistic and paralinguistic information (e.g., genealogical classification, geographical area, number of speakers, relevant references) is stored for each of the sampled languages. This is linked to information on the grammatical phenomena investigated in the thesis: gender, number, evaluative morphology, and their mutual interactions. The attributes associated with the three grammatical domains are explained in chapter 4, where the gender, number and evaluative morphology systems attested in the language sample are surveyed. Attributes regarding interactions between gender and number, and gender and evaluative morphology, were elaborated on the basis of the research questions that

this thesis aims to answer (see §2.8). A set of attributes is meant to capture the presence or absence of cumulative exponence between gender and number as well as syncretism of gender in the context of number. Another set of attributes is meant to capture relationships between evaluative morphology and grammatical gender. Attributes called "Notes" are used to store language-specific patterns of interactions between gender and number and gender and evaluative morphology that are difficult to capture in the form of pre-set values. Finally, the complexity scores for the gender systems of the sampled languages are also calculated via the database and stored in it.

3.6 Summary of the chapter

The sampling method followed in this thesis was designed to combine intra- and intergenealogical typology. The language sample that resulted from this sampling methodology is biased in at least two ways: genealogically and bibliographically. The sample is genealogically biased because it is built on the basis of genealogical relationships between subsets of languages. The sample is bibliographically biased in that, despite internal subdivisions within genealogical units being reflected by subsamples, these subsamples were created on the basis of available resources and not by means of mathematical methods. These two biases impact data analysis in that the language sample cannot be used for making statistical predictions over preferred typological patterns within and outside the African macro-area. However, making statistical predictions is not the purpose of this investigation. Rather, this investigation aims at to account for the frequency of, the stability of and that distribution of grammatical phenomena concerning the interaction between gender and number, and gender and evaluative morphology. My assumption is that these aims are better attained by looking at multiple subsets of related languages within a self-consistent area of the world. I also assume that a sampling method of this type may allow typologists to observe crosslinguistic patterns that a world-based, balanced language sample could not always detect with the same accuracy (see Iemmolo 2011 and Wälchli 2005 for similar considerations on language sampling).

4 Gender, number and evaluative morphology in the languages of the sample: an overview

4.1 Introduction

In this chapter, I survey the types of gender, number and evaluative morphology systems that are found in the 100 languages of the sample. The typology outlined in the chapter is synchronic. Diachronic data are discussed when useful for explaining peculiar synchronic distributions in the language sample.

Each section within the chapter is devoted to one of the three domains under investigation. §4.2 describes the types of gender systems attested in the languages of the sample; §4.3 classifies the languages of the sample according to nominal number, and an overview of the attested evaluative morphology systems is given in §4.4. A summary of the chapter is found in §4.5.

4.2 Gender

As already mentioned in chapter 3, 84 languages within the sample have grammatical gender. The remaining 16 languages do not have gender and are thus not discussed in this section.

In my database (see appendix B), the gender systems of the individual languages of the sample are classified according to the following criteria:

1. Sex-based and non-sex-based gender systems
2. Number of genders
3. Gender assignment
4. Number of gender-indexing targets
5. Occurrence of gender marking on nouns.

The first three criteria are based on WALS and, more specifically, on the chapters by Corbett (2013b,c,d). Criteria numbers 3 and 4 are my own. Criterion 3, "Number of indexing targets," has been introduced because, together with criteria 1 and 2, it is particularly useful for comparing how gender and number interact through indexation (chapter 5) and for investigating the absolute complexity of gender systems (chapter 7).

Criterion 4, "Occurrence of gender marking on nouns," has been introduced because of its usefulness to the study of exponence of gender in comparison with number (see chapter 5). Additional criteria concern interactions of gender and number and gender and evaluative morphology; these criteria are discussed in detail in chapters 5 and 6.

On the basis of criterion 1, the languages of the sample are classified into two macrotypes: languages with *sex-based* gender and languages with *non-sex-based* gender. This classification is based on Corbett (2013c). The following values are assigned to this feature in WALS and in my own database:

- Sex-based gender
- Non-sex-based gender
- No gender.

Sex-based gender systems are those whose semantic core is based on biological sex. Non-sex-based gender systems are not based on biological sex but rather on some notion of animacy (Corbett 2013c; see also discussion in §2.1.1). Table 4.1 provides the number of languages with sex-based and non-sex-based gender systems; percentages with respect to the languages with gender (Rel. %) and the total number of languages in the sample (Abs. %) are also given. Finally, the distribution of the different system types per genealogical unit is presented in the following way: the name of the genealogical unit is given first; this is followed by the number of languages from that unit that belong to a given type, as compared to (the symbol "/" is used) the total number of languages sampled from that unit. For instance, in table 4.1 the formulation *Chadic (6/8)* next to the value *sex-based* means that six out of the eight Chadic languages of my sample have sex-based gender. The same type of layout is used throughout the dissertation for the tables that summarise typological distributions.

In my sample, the languages with sex-based gender outnumber the languages with non-sex-based gender. A more detailed discussion of the sex-based and non-sex-based gender systems attested in the languages of the sample is found in §§4.2.2 and 4.2.3.

Table 4.1: Sex-based and non-sex-based gender systems in the language sample

	Gender system	**No. of lngs.**	**Rel. %**	**Abs. %**	**Genealogical groups**
	Sex-based	48	57%	48%	Berber (6/6) Chadic (6/8) Cushitic (13/13) Dizoid (1/1) Hadza (1/1) Khoe-Kwadi (5/5) Eastern Nilotic (3/3) South Omotic (1/1) Sandawe (1/1) Semitic (7/7) Ta-Ne-Omotic (4/4)
	Non-sex-based	36	43%	36%	Bantu (23/23) Kxa (1/1) Kwa (1/3) Mel (3/3) North-Central Atlantic (7/7) Tuu (1/1)
	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total		100	100%	100%	

Criterion 2, "Number of genders," classifies languages according to the number of gender distinctions that they display. As discussed in §2.1.2.2, the number of genders in a language is established on the basis of the identification of indexation classes (see examples (2.1) and (2.2) on the indexation classes of Italian). The classification that I follow in my database is based on the one used by Corbett (2013b). Languages are divided in subgroups according to the following cut-off points:

- Two genders
- Three genders
- Four genders
- Five or more
- No gender.

The distribution of number of gender distinctions in the languages of my sample is illustrated in table 4.2. As shown in the table 4.2, 50% of the languages with gender in my sample have two genders, whereas the remaining 50% is split between seven languages with three genders (8.3%), one language with four genders (1.2%) and thirty-four languages with five or more genders (40.5% of the total number of gendered languages). In addition, in the languages of my sample, systems with two gender distinctions are all sex-based with the exception of the Bantu language Bila, whose gender system is based on the opposition between animate and inanimate gender. Gender systems with three distinctions are also sex-based, whereas non-sex-based gender systems pattern with larger systems (from four to more than five distinctions). It is worth mentioning that this tendency cannot be generalised to the whole African continent. In his study of gender in African languages, Heine (1982) points out that the Ubangi languages Zande and Ma – not included in my sample – have a gender system that consists of four noun classes: Masculine, Feminine, Animal and Inanimate. Heine (1982) refers to this type of gender system as a *mixed gender system*, where sex-based and non-sex-based (or "nature-based," as he calls them) criteria of classification co-exist. However, in Heine's typology of the African noun class systems, this type of gender system is claimed to be very rare.[27]

Finally, in the literature on gender, as well as in reference grammars, it is rather common to refer to gender systems with four, five or more than five distinctions as *noun class systems*. This tradition is also followed by Corbett (2013b) and Corbett (2013c), and is discussed in §2.1. In this thesis, I use the cover term *gender* for both larger and smaller systems. The term *noun class* is used to refer to individual markers within a large gender system.

[27] Heine (1982) discusses various criteria for the classification of the gender systems of the African languages. Many of these criteria overlap to a large extent with the ones used in this thesis. See, for instance, his features "Sex-based vs. nature-based gender systems" and "Overt vs. covert noun classes" in comparison with my features "Sex-based vs. non-sex-based gender systems' and "Occurrence of gender marking on nouns."

Table 4.2: Number of genders in the language sample

	No. of genders	No. of lngs.	Rel. %	Abs. %	Genealogical groups
	Two	42	50%	41%	Bantu (1/23) Berber (6/6) Chadic (6/8) Cushitic (13/13) Dizoid (1/1) Eastern Nilotic (1/3) Hadza (1/1) South Omotic (1/1) Sandawe (1/1) Semitic (7/7) Ta-Ne-Omotic (4/4)
	Three	7	8.3%	7%	Eastern Nilotic (2/3) Khoe-Kwadi (5/5)
	Four	1	1.2%	1%	Kxa (1/1)
	Five or more	34	40.5%	34%	Bantu (22/23) Kwa (1/3) Mel (3/3) North-Central Atlantic (7/7) Tuu (1/1)
	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total		100	100%	100%	

The third criterion of classification, "Gender assignment," is based on Corbett (2013d) and attempts at classifying the criteria according to which nouns are allocated to the genders available in one language (for a general discussion of gender assignment in the languages of the world, see §2.1.1). The following values are assigned by Corbett to this feature:

- Semantic assignment
- Semantic and formal assignment
- No gender.

The distribution of gender assignment types in the languages of my sample is shown in table 4.3. The distribution of systems of gender assignment within my sample reflects that of the sample created by Corbett (2013d) for the WALS database, even though the two samples do not overlap much in their composition. Corbett's sample has 49 languages from the African macro-area. Of these 49 languages, 39 have gender: the majority of the gendered languages (34/39) has formal and semantic gender assignment whereas, the gender assignment of the remaining 5 languages is based strictly on semantic properties. In sum, according to both Corbett's sample and my sample, the distribution of systems of gender assignment in the African macro-area is highly skewed towards the mixed type (i.e., formal + semantic).

The fourth and fifth criteria of classification that I use in my database, "Number of gender-indexing targets" and "Occurrence of gender marking on nouns," are concerned with the formal marking of gender in the languages of the sample.

The values assigned to the feature "Number of gender-indexing targets" are:

- One indexing target
- Two indexing targets
- Three indexing targets
- Four or more
- No gender.

The following notions were used as general guidelines for identifying and counting the gender-indexing targets in each language: adjectives (in the sense of dedicated word classes for property-words), demonstratives, determiners, verbs, numerals, copulas, complementizers, adpositions. I did not examine in detail the behaviour of different subtypes within each of these macrotypes of targets. For instance, when I counted demonstratives as one of the possible gender indexing targets in a language, I did not consider different types of demonstratives as different indexes. A more thorough analysis of the behaviour of different types and subtypes of gender-indexing targets would be desirable but falls outside the scope of this investigation. The distribution of number of gender-indexing targets in the languages of my sample is shown in table 4.4.

Table 4.3: Systems of gender assignment in the languages of the sample

	Assignment system	**No. of lngs.**	**Rel. %**	**Abs. %**	**Genealogical groups**
	Semantic	6	7.1%	6%	Bantu (1/23) Chadic (1/8) Cushitic (1/13) Dizoid (1/1) Eastern Nilotic (1/1) South Omotic (1/1)
	Semantic and formal	76	90.5%	76%	Bantu (22/23) Berber (6/6) Chadic (5/8) Cushitic (12/13) Eastern Nilotic (2/3) Hadza (1/1) Khoe-Kwadi (4/5) Kxa (1/1) Mel (3/3) North-Central Atlantic (7/7) Sandawe (1/1) Semitic (7/7) Ta-Ne-Omotic (4/4) Tuu (1/1)
	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
	No information available	2	2.4%	2%	Chadic (1/8) Khoe-Kwadi (1/5)
Total		100	100%	100%	

Table 4.4: Number of gender-indexing targets in the language sample

	No. of targets	No. of lngs.	Rel. %	Abs. %	Genealogical groups
	One	5	6%	5%	Chadic (2/8) Khoe-Kwadi (1/5) Kxa (1/1) North-Central Atlantic (1/7)
	Two	16	19%	16%	Bantu (2/23) Chadic (1/8) Cushitic (4/13) Khoe-Kwadi (4/5) Mel (1/3) South Omotic (1/1) Ta-Ne-Omotic (3/4)
	Three	28	33.3%	28%	Bantu (3/23) Berber (5/6) Chadic (1/8) Cushitic (8/13) Dizoid (1/1) Eastern Nilotic (1/3) Hadza (1/1) North-Central Atlantic (1/7) Semitic (5/7) Sandawe (1/1) Ta-Ne-Omotic (1/4)
	Four or more	33	39.3%	32%	Bantu (17/23) Berber (1/6) Chadic (2/8) Cushitic (1/13) Eastern Nilotic (2/3) Kwa (1/3) Mel (1/3) North-Central Atlantic (5/7) Semitic (2/7) Tuu (1/1)
	No data	2	2.4%	2%	Mel (1/3) Bantu (1/23)
	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total		100	100%	100%	

Table 4.4 shows that even though there is considerable variation within and across genealogical groupings in terms of the number of gender-indexing targets, the distribution is skewed toward either "three" or "four or more indexing targets." The languages of the sample thus tend to have rather pervasive gender indexation.

Finally, it is worth mentioning that only four languages within the sample have a pronominal gender system (for a definition of *pronominal gender system*, see §2.1.2.2): Ju|'hoan (Kxa), Kwadi (Khoe-Kwadi), Mwaghavul (Chadic) and Pero (Chadic). This distribution reflects the crosslinguistic generalisation whereby pronominal gender systems are a very rare phenomenon worldwide (see Corbett 2013b and chapter 2 for discussion). Very little is known about the gender system of Kwadi, a language spoken in Angola that recently became extinct, and the Nigerian language Pero (see Güldemann 2004 for a description of the Kwadi gender system in the Khoe-Kwadi context, and Frajzyngier 1989 for a description of the gender system of Pero). On the other hand, the gender systems of Ju|'hoan and Mwaghavul are well-described and will be discussed in greater detail later in this thesis.

The fifth and last criterion for classifying the gender systems of the languages sample is "Occurrence of gender marking on nouns." The following values are assigned to this feature:

- Gender marking on nouns: Yes
- Gender marking on nouns: No
- No gender.

The distribution of overt gender marking in the languages of my sample is shown in table 4.5 (next page).

Table 4.5 shows that the large majority of the languages with gender in the sample have overt coding of gender on nouns. As mentioned at the beginning of the section, this feature is especially relevant for analysing types of exponence of gender on nouns in comparison with number. This is discussed in detail in §5.2.1.4.

4.2.1 Criteria of classification of gender systems: summary

In this section, I presented the criteria that I use to classify the gender systems of the languages of the sample in my database and throughout the dissertation. Five criteria have been introduced:

1. Sex-based and non-sex-based gender systems
2. Number of genders
3. Gender assignment
4. Number of gender-indexing targets
5. Occurrence of gender marking on nouns.

Table 4.5: Overt coding of gender on nouns in the languages of the sample

	Gender on nouns	**No. of lngs.**	**Rel. %**	**Abs. %**	**Genealogical groups**
	Yes	69	82%	69%	Bantu (23/23) Berber (6/6) Chadic (2/8) Cushitic (8/13) Dizoid (1/1) Eastern Nilotic (3/3) Hadza (1/1) Khoe-Kwadi (5/5) Kwa (1/3) Mel (3/3) North-Central Atlantic (6/7) Semitic (6/7) Ta-Ne-Omotic (3/4) Tuu (1/1)
	No	15	18%	15%	Chadic (4/8) Cushitic (5/13) Kxa (1/1) North-Central Atlantic (1/7) Sandawe (1/1) Semitic (1/7) South Omotic (1/1) Ta-Ne-Omotic (1/4)
	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total		100	100%	100%	

The distribution of the values of each of the five criteria in the languages of the sample has been shown.

Even though the five criteria are equally important, for the sake of simplicity I use criterion 1, "Sex-based and non-sex-based gender systems," as a tool to refer to the gender systems of the languages of the sample throughout the dissertation. Reference to the other criteria is made whenever relevant to the topics under discussion.

The distribution of the sex-based and non-sex-based gender systems within the languages sample is illustrated in the map in figure 4.1 (next page). As shown in figure 4.1, non-sex-based gender systems are localised in a sub-area of Africa that extends southward from the Sub-Saharan region. Sex-based gender systems are more scattered around the continent, but their concentration is higher all throughout northern and central Africa (eastern and central Africa more than western Africa). Isolated sex-based gender systems are also found in southern Africa among the languages of the Khoe-Kwadi family.

I now turn to an overview of the types of sex-based and non-sex-based gender systems attested among the languages of the sample.

4.2.2 Sex-based gender

Of the 48 languages with sex-based gender, 41 have two gender distinctions (i.e., masculine vs. feminine), whereas only 7 languages have three (i.e., masculine vs. feminine vs. neuter/common). Gender systems of the latter type are characteristic of the Khoe-Kwadi[28] and Eastern Nilotic languages. In the Khoe-Kwadi languages, nouns can be assigned to the Feminine, the Masculine or the Common Gender. No noun is assigned to the Common Gender by default: in fact, the third gender is only used when the gender of a noun is left undefined, or in case of conjoined masculine and feminine nouns. The former usage is shown in example (4.1) from Nama, where the noun for 'animal' is marked as common and plural since the NP refers to the animal world as consisting both of male and female referents.

(4.1) Nama (Khoe-Kwadi) (adapted Hagman 1977: 153)

xaḿ-i	*ke*	*'a*	*ǀúrú-ǹ*	*há-ń*	*t*	*k'o-'ao*
lion-3.M.SG	DECL	COP	animal-3.C.PL	all-3.C.PL	of	rule-man

'The lion is the king of all beasts'

Gender distinctions in Khoe-Kwadi languages are expressed by means of cumulative suffixes that also encode number distinctions; gender distinctions are usually overtly coded on nouns (with the exception of Kwadi) or on the last member of the NP. Gender indexation is usually also found on pronouns and verbs. In Kwadi, pronouns are the only targets of gender indexation.

[28] The existence of a third gender – the Common Gender – in Kwadi has been called into question by Güldemann (2004), who claims that the morphological evidence for a third gender in the language is, at best, very weak.

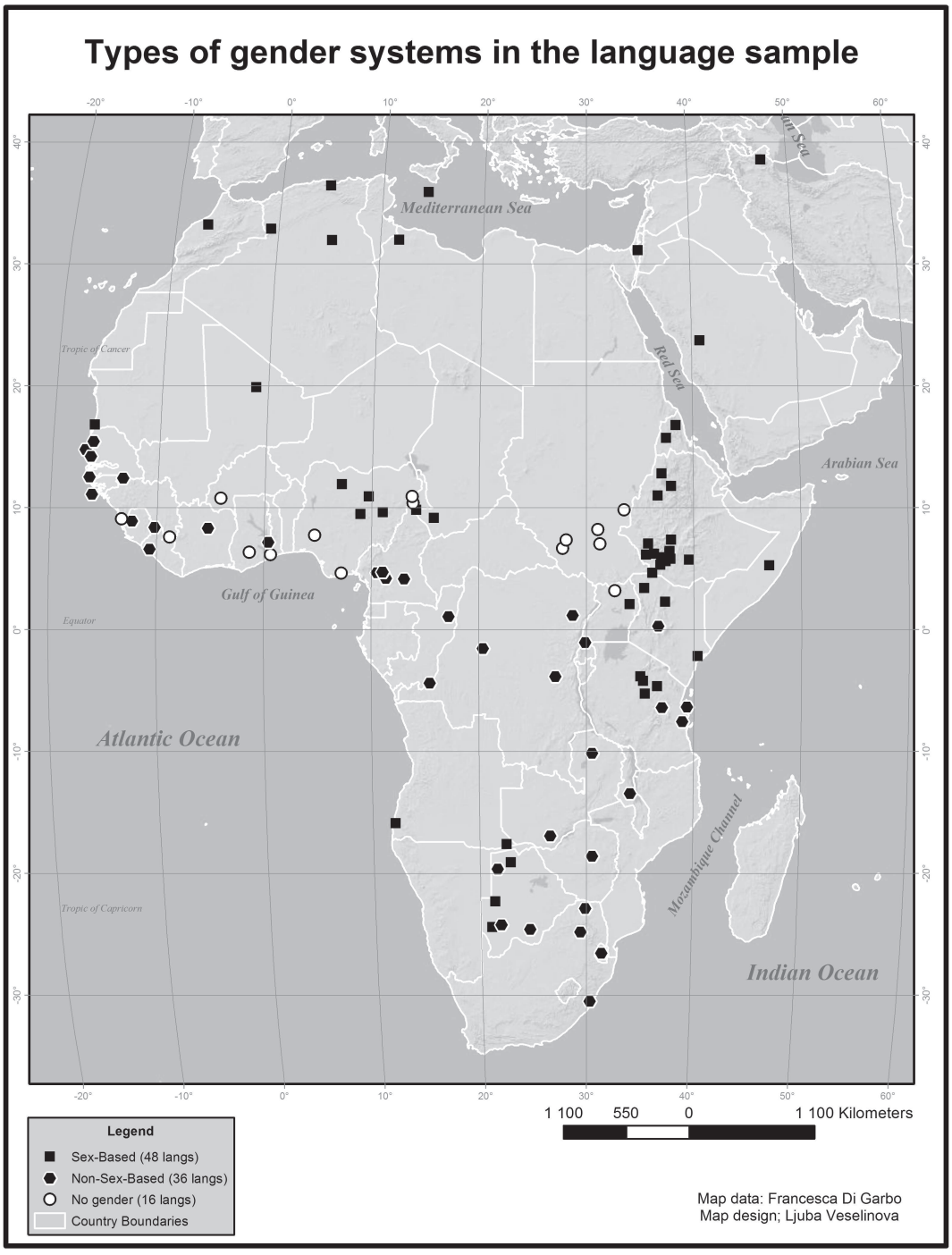


Figure 4.1: Sex-based and non-sex-based gender systems in the languages of the sample

In two Eastern Nilotic languages that are included in the sample, namely, Turkana and Karamojong, nouns can be assigned to the Feminine, Masculine or Neuter Gender. Both in Turkana and Karamojong, the Neuter Gender is, overall, less productive than the Masculine and the Feminine. Masai (Eastern Nilotic) has been classified as having two genders. A third gender is also found in the language, but it is only used to encode location. Payne (1998) refers to the Locative Gender of Masai as a very marginal noun class whose manifestations are only associated with one noun *wwéjì* 'place.' The Locative Gender in Masai qualifies as an *inquorate gender* in Corbett's (1991: 170) terms, that is, as a gender that is "postulated on the basis of an insufficient number of nouns" that are to be counted as lexical exceptions. In all Eastern Nilotic languages, gender is encoded by means of cumulative prefixes that also encode number distinctions. Gender indexation is generally internal to the NP.

Sex-based gender systems with two genders are characteristic of all the Afro-Asiatic branches included in my sample – Berber, Chadic, Cushitic, and Semitic – as well as the Omotic groups (Dizoid, South Omotic and Ta-Ne-Omotic) and the two isolates Hadza and Sandawe. Examples from a language with this type of system, Kambaata (Cushitic), are given in examples (4.2) and (4.3). Example (4.2) illustrates semantically motivated gender assignment, whereas (4.3) illustrates gender indexation on verbs with two inanimate nouns whose gender assignment is not semantically motivated.

(4.2) Kambaata (Cushitic) (Treis 2008: 127)

(a) *hizóo-ha*
sibling-M
'brother'

(b) *hizóo-ta*
sibling-F
'sister'

(4.3) Kambaata (Cushitic) (Treis 2008: 128)

xórb-u	*barcum-íichch*	*aaz-íin*	*afuu'll-itée'u*
ball-F.NOM	chair-M.ABL	interior-M.ICP	sit-3F.PFV

'The ball is [lit. 'is sitting'] under the chair'

In Berber, the Feminine Gender is overtly coded by means of a circumfix. The Masculine Gender is zero-marked. Gender distinctions are generally overtly coded on nouns, pronouns, adjectives and verbs. As for the other Afro-Asiatic branches within the sample, as well as the Omotic groups, those languages in which gender distinctions are overtly coded on nouns usually have suffixal gender markers. Gender marking on nouns is suffixal also in the two isolates Hadza and Sandawe. The number and types of gender-indexing targets vary from language to language and are thus difficult to summarise here. Reference to language-specific patterns of gender indexation is made at different places throughout the thesis (see also table 4.4).

The origin of the sex-based gender of some of the languages considered in this thesis has been object of extensive investigation. In all the documented cases, the grammaticalization of sex-based gender distinctions is connected with deictic (demonstratives or definiteness markers) or anaphoric (personal pronouns) entities. I now turn to a short survey of some of these grammaticalization paths.

In the case of the Eastern Nilotic languages, sex-based gender is an innovation that opposes them to the closely related languages of the Western Nilotic branch, where there is no gender. The gender system of Eastern Nilotic languages originated from prenominal modifiers derived from the noun for 'member/person,' in the case of the Masculine Gender, and from the word for 'girl/daughter,' in the case of the Feminine Gender. The prenominal modifiers later grammaticalized as demonstratives and merged with the demonstrative paradigm inherited from Proto-Nilotic. The new sets of demonstratives initially encoded only deixis but later became gender-indexing targets. A detailed analysis of the origin of gender in Eastern Nilotic languages is found in Heine & Vossen (1983). Similarly, the overt coding of gender distinctions on nouns in Eastern Cushitic languages originated from the grammaticalization of gender-sensitive demonstratives as gender markers (Treis 2008).

The sex-based gender of the Berber languages is the result of reanalysis of markers that were at first used to single out entities within a part-whole type of relationship and only later began to mark biological gender (Mettouchi 2000). Finally, the system of gender marking in the Khoe-Kwadi languages originated from anaphoric pronouns encoding third person reference (on the grammaticalization of gender/number markers in Kxoe, see Kilian-Hatz & Heine 2010).

Typically, in languages with sex-based gender, gender assignment is more semantically transparent with human and animate nouns. This is in line with the generalisations on the typology of gender systems discussed in chapter 2. In chapters 6 and 7, I shall show that a component of semantic predictability in the gender system of many of these languages is related to the encoding of size variation and/or variation in the countability properties of nouns.

4.2.3 Non-sex-based gender

As mentioned before, the non-sex-based gender systems within my sample are in nearly all cases large systems with four, five or more than five gender distinctions.

Ju|'hoan, a Kxa language spoken in Botswana and Namibia, is the only language of my sample with four noun classes. Gender distinctions in Ju|'hoan are never overtly coded on nouns but only appear on anaphoric pronouns, demonstratives and possessee pronouns. These vary according to gender, but no pattern of affixation is identifiable since none of the pronouns appears to be morphologically complex (Heine & König 2011: 154-155). No gender indexation is found on verbs. Five genders are found in !Xóô, a Tuu language also spoken in Botswana. Noun class markers in !Xóô are etymologically related to person pronouns and are overtly coded on nouns, adjectives and verbs (object indexation) by means of suffixes (Güldemann 2000).

If the two systems mentioned above, and the Bantu language Bila (with only two

genders) are excluded, the most robust type of non-sex-based gender system within my sample has more than five distinctions. It is found in 33 out of 36 languages with non-sex-based gender (see table 4.1). These languages, with the exception of the Kwa language Sɛlɛɛ, all belong to the North-Central Atlantic, Bantu and Mel genealogical groups. The noun class systems of the languages of these groups are very similar to each other; these similarities are genealogically motivated since the three groups constitute independent branches of the Atlantic-Congo family. The noun class system of the Bantu languages is, however, considered more conservative with respect to the Proto-Atlantic-Congo system (Katamba 2003; Nurse & Philippson 2003), and the affiliation of a language or a set of languages with the Atlantic-Congo family is often evaluated on the basis of the presence (or absence) of a Bantu-like noun classification system. In this section, I discuss those characteristics of the noun class systems of the Bantu, North-Central Atlantic, and Mel languages that are shared by all three groups. Differences between Bantu, on the one hand, and North-Central Atlantic and Mel, on the other, are discussed separately in §§4.2.3.1 and 4.2.3.2. For the sake of simplicity, I use the label *Atlantic* as a cover term to refer to North-Central Atlantic and Mel. As explained in §3.2, the term does not have any genealogical reality because the two groups – North-Central Atlantic and Mel – are classified as independent branches of Atlantic-Congo.

The prototypical noun class-marking strategies in the Atlantic and Bantu languages consist of the combination of affixes marked on the noun stem and a number of indexing targets, the most frequent of which are verbs, adjectives, pronouns (of different kinds) and numerals. Noun class markers are entirely prefixal in Bantu, whereas in Atlantic, they can be either prefixal or suffixal (see §4.2.3.2). The individual noun classes in Bantu and Atlantic are traditionally described as being either singular or plural. According to the Bantu and Atlantic tradition, pairs of singular and plural class markers form *genders* (see also §2.1). Nouns are assigned by default to a gender/noun class pair or, in the case of nouns that do not participate in number distinctions, to a noun class. In addition, nouns can be assigned to noun classes other than their default class. This process is often referred to in the literature as *noun class shift* or *gender shift* (see §2.1.1). The formation of diminutives, augmentatives, and collectives, as well as personifications of animal nouns, and – in some cases – the encoding of location are all realised by means of processes of this type. It is common among specialists of Bantu and Atlantic languages to define the change from singular to plural classes also as an instance of *noun class shift*. Thus, one could say that the notion of *noun class shift* can be operationalised both in a broad and narrow sense. An example of noun class shift in the broad sense is given in (4.4): the example illustrates singular-plural alternations in the Bantu language Kirundi.

(4.4) Kirundi (Bantu) (adapted from Mel' čuck & Bakiza 1997: 286)

(a) *umun-tu* *mu-bi*
CL1-man CL1-amazing
'an amazing man'

(b) *aba-ntu* *ba-bi*
CL2-man CL2-amazing
'amazing men'

A crucial issue when dealing with Bantu and Atlantic gender systems is their semantic basis. Experts of Bantu and Atlantic languages generally argue in favour of one of the following two hypotheses, both based on synchronic analysis:

1. Gender systems in Bantu and Atlantic are devoid of semantic content. Attempts at reconstructing the underlying semantics of the individual noun classes are neither satisfactory nor convincing (see, for example, Van de Velde 2006: 192-195).

2. Semantically driven gender assignment is still recoverable in contemporary Bantu and Atlantic languages despite that the semantics of some individual noun classes and genders has undergone massive bleaching through time (see, among others Breedveld 1995; Contini-Morava 1997; Denny & Creider 1976; Givón 1971; Katamba 2003; Sagna 2012). A good strategy for depicting semantic productivity in the Bantu and Atlantic noun class systems is to analyse the word-formation processes that are based on class shifts (Desmuth 2000; Sagna 2012).

In this thesis, I consider the second hypothesis the most enlightening one. In fact, even though the semantic motivation for gender assignment is in many cases no longer synchronically transparent, mechanisms of word-formation based on noun class shift are, to a certain extent, grounded in semantics. Based on these premises, two questions are worth asking: (1) does the semantics of the Bantu and Atlantic individual noun class markers condition their grammatical behaviour? and (2) do the different classes have the same status within the system? The two questions are addressed in §§4.2.3.1 and 4.2.3.2, as well as in chapter 6, where the peculiar properties of the evaluative noun classes in Bantu and Atlantic languages are discussed.

4.2.3.1 Noun Classes in Bantu

The gender systems of the Bantu languages are very well studied. Among the most recent contributions, a comprehensive comparative account of the gender systems of a sample of 300 Bantu languages is found in Maho (1999). In this section, I shall survey aspects of the Bantu gender systems that are not shared with North-Central Atlantic and Mel.

Table 4.6 provides a complete overview of the noun class morphology of Kirundi: noun class prefixes on nouns (NCP), gender indexes on adjectives (ADJ), pronouns (PRO), subject (SBJ) and object (OBJ) markers on verbs are noted. Notice that the Bantu noun class markers are usually numbered in such a way that singular prefixes are referred to by odd digits and plural prefixes by even digits. In table 4.6, the singular classes are typed in boldface.

It is believed that the Proto-Bantu system was larger than any system attested in contemporary Bantu languages. This explains the notation of the noun class marker

Table 4.6: Noun classes and indexation patterns in Kirundi (adapted from Meeussen 1959; Mel' čuck & Bakiza 1997)

Class	NCP	ADJ	PRO	SBJ	OBJ
1	*mu-*	*mu-*	*u-*	*a-*	*mu-*
2	*ba-*	*ba-*	*ba-*	*ba-*	*ba-*
3	*mu-*	*mu-*	*u-*	*u-*	*u-*
4	*mi-*	*mi-*	*i-*	*i-*	*i-*
5	*ri-, ø-*	*ri-*	*ri-*	*ri-*	*ri-*
6	*ma-*	*ma-*	*a-*	*ma-*	*ma-*
7	*ki-*	*ki-*	*ki-*	*ki-*	*ki-*
8	*bi-*	*bi-*	*bi-*	*bi*	*bi-*
9	*ny-, ø-*	*ny-*	*i-*	*i-*	*i-*
10	*ny-, ø-*	*ny-*	*zi-*	*zi-*	*zi-*
11	*ru-*	*ru-*	*ru-*	*ru-*	*ru-*
12	*ka-*	*ka-*	*ka-*	*ka-*	*ka-*
13	*tu-*	*tu-*	*tu-*	*tu-*	*tu-*
14	*bu-*	*bu-*	*bu-*	*bu-*	*bu-*
15/17	*ku-*	*ku-*	*ku-*	*ku-*	*ku-*
16	*ha*	*ha*	*ha-*	*ha-*	*ha-*

numbered as 15/17 in the table: in Kirundi, the two Proto-Bantu classes numbered as 15 and 17 merged into one class marker, *ku-*. Similar to Kirundi, none of the attested Bantu languages has preserved the system reconstructed for Proto-Bantu (Katamba 2003). Variation is found in the number of genders that has been inherited or lost in the individual languages of the group.[29] It is generally assumed among Bantu specialists that languages with the most canonical type of system have (a) approximately six individual class markers paired according to singular and plural alternations, and (b) a certain number of classes (more or less up to six) that are not paired according to number distinctions.

In many Bantu languages, gender prefixes can be preceded by a prefix that is traditionally referred to as *augment*, *pre-prefix* or *initial vowel*. The function of the augment varies from language to language and multiple functions can be found in one and the same language. In general, the function of augments is related to definiteness, specificity and/or focus. For an overview, see Katamba (2003) and Maho (1999).

Two types of noun class prefixes can be distinguished in Bantu languages: *replacive* and *additive*. This distinction is particularly relevant for the analysis of the interactions between gender and evaluative morphology but not for the relationship between gender and number. The difference between replacive and additive classes is mainly concerned with the way in which gender is marked on nouns and how this affects indexation pat-

[29] A few languages, like Komo (Bantu, D.30), do not have any gender distinctions at all (Katamba 2003: 109).

terns. In the case of the replacive classes, noun class shifts determine the replacement of the original class marker on the noun as well as on the indexing targets; in the case of the additive classes, the prefix of the class to which the noun is shifted is added to the original prefix and, typically, triggers indexation. As will be shown in detail in chapter 6, diminutive and augmentative class markers can be of the additive type. Example (4.5) illustrates an additive class marker in Herero, Class 7, which is used to derive augmentatives (Herero is not part of my language sample). Only the additive class marker is reflected on the indexing target (in this case the possessive for 'my').

(4.5) Herero (Bantu) (Crisma et al. 2011: 259)

otji-ru-vy	*tj-nde*
CL7-CL11-knife	CL7-my

'my big knife'

The difference between replacive and additive classes already answers the two questions raised at the end of the previous section: (1) Does the semantics of the (Bantu) individual noun classes condition their grammatical behaviour? and (2) Do the different classes have the same status within the system? The status of the individual noun classes within the system is not equal: differences in their grammatical behaviours may be semantically conditioned.

4.2.3.2 Noun classes in Atlantic

In the Atlantic languages, the gender markers can be either prefixal, suffixal or both. Suffixes and prefixes are etymologically related to each other, but the suffixes represent an innovation in the languages of the area; the origin of suffixal noun class markers in Atlantic is briefly described at the end of this section. Table 4.7 illustrates the individual noun class prefixes of the North-Central Atlantic language Bandial, as marked both on nouns (NCP) and indexing targets (in order: determiners, demonstratives, personal pronouns, subject markers, relative pronouns, and adjectives).
The singular classes are typed in boldface, as in the case of Kirundi (table 4.6). Notice that the last three classes, which express location, do not distinguish number. In line with the tradition followed for the Bantu languages, the singular classes of Bandial are referred to by odd digits with the only exception being Class 12, which is also singular but is labelled with an even digit (Sagna 2012). Contrary to what is observed for the Bantu languages, however, there is no conventionalised numbering of the individual noun classes of the Atlantic languages. The numbering of the markers often varies from language to language. Nonetheless, at least for gender markers that are are used to mark human nouns, that is, classes 1 and 2, there is a substantial overlapping in the numbering adopted for the different languages.

The number of the individual noun classes of the Atlantic languages can vary a great deal: as reported by Wilson (1989: 96), for example, Kobiana (North-Central Atlantic) has nearly 40 classes, whereas Nalu (an isolate within Atlantic-Congo and the Atlantic area) has only three. The morphophonology of the markers also varies from language to

Table 4.7: Noun classes and indexation patterns in Bandial (adapted from Sagna 2012)

Class	NCP	DEF.DET	DEM	PRO	SBJ	REL	ADJ
1	*a-*	*øahu*	*um-*	*ø-*	*a-*	*ø*	*a-*
2	*bug-*	*bugagu*	*ubug-*	*bug-*	*gu-*	*g-*	*gu-*
	gu-	*gagu*	*u-bug-*	*bug-*	*gu-*	*g-*	*gu-*
	u-	*wawu*	*ubug-*	*bug-*	*gu-*	*g-*	*gu-*
	e-	*yayu*	*ubug-*	*bug-*	*gu-*	*g-*	*gu-*
	su-	*sasu*	*u-bug-*	*bug-*	*gu-*	*g-*	*gu-*
3	*e-*, *y-*	*yayu*	*uy-*	*y-*	*e-*	*y-*	*e-*
4	*su-*, *si-*, *s-*	*sasu*	*us-*	*s-*	*su-*	*s-*	*su-*
5	*bu-*, *bi-*, *b-*, *ba-*	*babu*	*ub-*	*b-*	*bu-*	*b-*	*bu-*
6	*u-*, *w-* *wawu*	*uw-*	*ø*	*w-*	*u-*	*w-*	*u-*
7	*fu*, *fi-*, *f-*, *fa-*	*fafu*	*uf-*	*f-*	*fu-*	*f- fu-*	
8	*gu*, *ga-*, g-	*gagu*	*ug-*	*g-*	*gu-*	*g-*	*gu-*
9	*ga-*	*gagu*	*ug-*	*g-*	*gu-*	*g-*	*gu-*
10	*mu-*, *mi-*, *ma-*, *m-*	*mamu*	*um-*	*m-*	*mu-*	*m-*	*mu-*
11	*ju-*, *ja-*	*jaju*	*uj-*	*j-*	*ju-*	*j-*	*ju-*
12	*ña-*	*ñañu*	*uñ-*	*ñ-*	*ñu-*	*ñ-*	*ñu-*
13	*t-*	*tatu*	*ut-*	*t-*	*tu-*	*t-*	–
14	*d-*	*dadu*	*ud-*	*d-*	*du-*	*d-*	–
15	*n-*	–	–	*n-*	–	–	–

language. However, languages from the same genealogical unit have, as expected, rather similar noun class inventories. In spite of this diversity in the shape and number of the noun class exponents, virtually every language of the Atlantic area has an *o*-like marker for singular humans, a *ba*-like marker for plural humans, and a *m/ma*-like marker for masses and liquids (Wilson 1989: 98).

An interesting characteristic of gender marking in the Atlantic languages is consonant alternation, whereby the quality of the (initial) consonants of the noun class markers phonetically conditions the quality of the (initial) consonants of the noun stems. Consonant alternation is not found everywhere within the area: it is a prominent feature of North-Central Atlantic languages – such as Maasina Fulfulde, Serer, Nuclear Wolof (although less in the latter language than in the first two) – and of other subgroups, such as the Tenda languages. Only a few traces of consonant alternation are found in the languages of the Mel group. In Maasina Fulfulde, where noun class marking is suffixal, consonant alternation on the initial consonant of the noun stem is the only trace of the now lost noun class prefixes.

As mentioned in the beginning of the section, in the languages of the Atlantic area,

gender marking can be either prefixal, suffixal or both. Suffixal class markers are found, for example, in Kisi (Mel) and in the languages of the Fula subgroup (North-Central Atlantic). The synchronic distribution of the two types of noun class affixes has been explained as the result of a diachronic process of gender renewal, whereby, as mentioned before, the suffixes represent the innovation and the prefixes the original type of marking. A short digression into diachrony may be useful for understanding the distribution of this phenomenon in the languages of the area.

Gender renewal in Kisi, and the Mel languages in general, has been investigated by Childs (1983) using the method of intragenealogical typology (see chapter 3). Childs compares gender marking in Kisi with gender marking in the other languages of the Mel branch (Southern Atlantic, according to his classification) and postulates an areal continuum, ranging from strongly prefixal gender systems – as in Timne – to solely suffixal gender systems – as in Kisi. He also provides evidence for remnants of noun class prefixes in Kisi. In Childs' analysis, the different types of languages attested in this area represent synchronic snapshots of what the diachronic stages of the process of affix renewal may have looked like in the languages that underwent it. According to Childs, suffixal class marking originates from the repetition of noun class indexation after a relative clause or a fronted NP. The expansion of suffix use is accompanied by the phonetic erosion of the older prefixes.

To my knowledge, the origin of the suffixal gender markers in North-Central Atlantic, and Fula languages in particular, has not been investigated in a systematic fashion. A possible explanation for the origin of gender renewal in Fula might be seen in the behaviour of definite and indefinite nouns in the closely related language Serer (North-Central Atlantic) (Guillaume Segerer, personal communication). In Serer, noun class marking is strongly prefixal. However, in addition to their respective class prefixes, definite nouns are accompanied by a clitic that expresses definiteness, proximity and class at the same time.

(4.6) Serer (North-Central Atlantic) (McLaughlin 1992)

(a) *o-tew*
CL1-woman
'woman'

(b) *o-tew=oxe*
CL1-woman=CL.DEF
'the woman'

If the interpretation of the Serer data is correct, the overt coding of definiteness on nouns could be seen as a plausible trigger of the grammaticalization of noun class suffixes in Fula (Guillaume Segerer, personal communication). This hypothesis calls for further investigation.

4.3 Number

In my database (see appendix B), number systems are classified according to the following criteria:

1. Type of nominal number systems in terms of obligatoriness
2. Number values
3. Number of number-indexing targets.

I designed the three criteria based on those aspects of the typology of nominal number systems that are more relevant to the purposes of this investigation. Criterion 1 is used to control if the relationship between gender and number is subject to vary depending on whether or not nominal number is obligatorily marked. Criterion 2 accounts for the types and number of oppositions within a number system and is used to control if these affect in any way the interactions between gender and number. Finally, criterion 3 is crucial for investigating how gender and number interact through indexation. It is worth mentioning that two other features are used to explore the typology of nominal number systems in WALS: "Occurrence of nominal plurality" (Haspelmath 2013) and "Coding of nominal plurality" (Dryer 2013). The first feature accounts for the extent to which plurality is overtly coded on nouns in a given language; the second feature classifies languages according to the morphosyntactic encoding of plurality on nouns. Since the two features are not strictly relevant to the study of the interactions of gender and number, I do not use them as a parameter of classification for the languages of my sample.

On the basis of criterion 1, "Type of nominal number system in terms of obligatoriness," languages are classified into two macrotypes:

(1) Languages with obligatory number: languages in which speakers need to choose the number value of a noun whenever they use it.

(2) Languages with general number: languages in which (some or all) nouns can be outside the system of number distinctions.

This classification is based on Corbett's (2000) discussion of the major systems of number marking in the languages of the world. It is worth mentioning here that Corbett's classification is grounded on the number morphology of nouns rather than on the patterns of number indexation triggered by nouns (see chapter 2 for a general overview). Table 4.8 illustrates the distribution of the two systems in the languages of the sample; the percentage of each type of system with respect to the total number of languages in the sample is also given in the table. For a discussion of general number and obligatory number, see §§2.2.2.1 and 2.2.2.2.

Table 4.8: Nominal number systems in the languages of the sample

	Type	No. of lngs.	%	Genealogical groups
	Obligatory number	71	71%	Bantu (23/23) Berber (6/6) Cushitic (2/13) Dizoid (1/1) Eastern Nilotic (3/3) Hadza (1/1) Khoe-Kwadi (5/5) Kwa (3/3) Mel (3/3) North-Central Atlantic (7/7) Sandawe (1/1) Semitic (6/7) South Omotic (1/1) Ta-Ne-Omotic (2/4) Tuu (1/1) Western Nilotic (6/6)
	General number	29	29%	Cushitic (11/13) Chadic (8/8) Defoid (1/1) Igboid (1/1) Kxa (1/1) Mande (4/4) Semitic (1/7) Ta-Ne-Omotic (2/4)
Total		100	100%	

As shown in table 4.8, the distribution of types of nominal number systems in terms of obligatoriness is genealogically skewed in the sense that languages from the same genealogical units have roughly the same type of number system. Outliers are Awngi and Beja with respect to Cushitic; Bench and Koorete with respect to Ta-Ne-Omotic; and Amharic with respect to Semitic. Awngi and Beja are the only Cushitic languages of the sample with obligatory number. Similarly, Bench and Koorete are the only languages with general number within my Ta-Ne-Omotic sample, and Amharic is the only Semitic language of the sample where general number is productive (see §5.2.1 for nouns with general meaning in Maltese).

Criterion 2, "Number values," identifies the types of number distinctions that are found in a language. The values associated with these features are based on the discussion of the Number Hierarchy outlined in §2.2.2.2, where the distribution of number values across the languages of the world has been considered.

- Singular vs. plural
- Singular vs. plural vs. dual

- Singular vs. plural vs. dual vs. trial
- Singular vs. plural vs. dual vs. paucal.

The distribution of number values in the languages of the sample is shown in table 4.9.

Table 4.9: Number values in languages of the sample

Number values	**Number of lngs.**	%	**Genealogical groups**
singular vs. plural	89	89%	Bantu (23/23) Berber (6/6) Cushitic (12/13) Chadic (7/8) Defoid (1/1) Dizoid (1/1) Estern Nilotic (3/3) Hadza (1/1) Igboid (1/1) Kwa (3/3) Kxa (1/1) Mande (4/4) Mel (3/3) North-Central Atlantic (7/7) Semitic (4/7) Sandawe (1/1) South Omotic (1/1) Ta-Ne-Omotic (3/4) Tuu (1/1) Western Nilotic (6/6)
singular vs. plural vs. dual	9	9%	Chadic (1/8) Khoe-Kwadi (5/5) Semitic (3/7)
singular vs. plural vs. trial	0	–	–
singular vs. plural vs. paucal	2	2%	Cushitic (1/13) Ta-Ne-Omotic (1/4)
Total	100	100%	

As shown in table 4.9, 89% of the sampled languages have a "singular vs. plural" type of contrast. The distribution of the less common systems shows some kind of genealogical and areal skewing. The "singular vs. plural vs. dual" type of opposition is found in all the Khoe-Kwadi languages of the sample as well as in three very closely related Semitic languages – Standard Arabic, Moroccan Arabic and Maltese. The "singular vs. plural vs. paucal" type of opposition is extremely rare within my language sample and only

attested in one Cushitic and one Ta-Ne-Omotic language, Baiso and Koorete. The two languages are not genealogically related (especially if Ta-Ne-Omotic is not viewed as part of Afro-Asiatic; see discussion in §3.2), but they have a long-standing history of contact since they are spoken in the same area of Ethiopia (see §5.5.2 for a more detailed overview).

The third and last criterion of classification of number systems in the languages of the sample is "Number of number-indexing targets." The targets of number indexation in the languages of the sample have been counted following the same method adopted for gender indexation (see §4.2 for a detailed description). The values associated with this feature are:

- One indexing target
- Two indexing targets
- Three indexing targets
- Four or more indexing targets
- No number indexation.

The distribution of number of indexing targets in the languages of the sample is presented in table 4.10. For a general discussion of indexation and nominal number, see §2.2.4. As shown in table 4.10, 60% of the languages of the sample has rather pervasive nominal number indexation (between three and four indexing targets). The relationship between presence of gender and pervasiveness of number indexation is discussed in greater detail in section 5.6. For the sake of comparability, the distribution of different combinations of number of gender- and number of number-indexing targets in the languages of the sample is presented in appendix C.

As in the case of gender, the three criteria for the classification of nominal number systems are equally important. However, for the sake of clarity, I use the opposition between languages with general number and languages with obligatory number as a way to refer to the number systems of the languages of the sample. The geographical distribution of the individual types of number systems within the African macro-area is also illustrated in the map in figure 4.2.

Strategies of number marking in languages with general number are discussed in detail in §4.3.1. In §4.3.2, I provide an overview of a very interesting type of obligatory number system that is found only among the Nilotic languages of the sample and is commonly labelled as *tripartite number system*. Those properties of number marking that correlate with the marking of gender distinctions are accounted for in chapter 5.

Table 4.10: Number of number-indexing targets in the languages of the sample

No. of targets	Number of lngs.	%	Genealogical groups
One	11	11%	Chadic (1/8) Cushitic (1/13) Igboid (1/1) Khoe-Kwadi (1/5) Kwa (1/3) Kxa (1/1) Mande (4/4) North-Central Atlantic (1/7)
Two	21	21%	Bantu (2/23) Chadic (4/8) Cushitic (3/13) Defoid (1/1) Dizoid (1/1) Khoe-Kwadi (4/5) Kwa (1/3) Mel (1/3) South Omotic (1/1) Ta-Ne-Omotic (3/4)
Three	26	26%	Bantu (3/23) Berber (5/6) Cushitic (8/13) Chadic (1/8) Eastern Nilotic (1/3) North-Central Atlantic (1/7) Sandawe (1/1) Semitic (5/7) Ta-Ne-Omotic (1/4)
Four or more	34	34%	Bantu (17/23) Berber (1/6) Chadic (2/8) Cushitic (1/13) Eastern Nilotic (2/3) Hadza (1/1) Kwa (1/3) Mel (1/3) North-Central Atlantic (5/7) Semitic (2/7) Tuu (1/1)
No indexation	6	6%	Western Nilotic (6/6)
No data	2	2%	Bantu (1/23) Mel (1/3)
Total	100	100%	

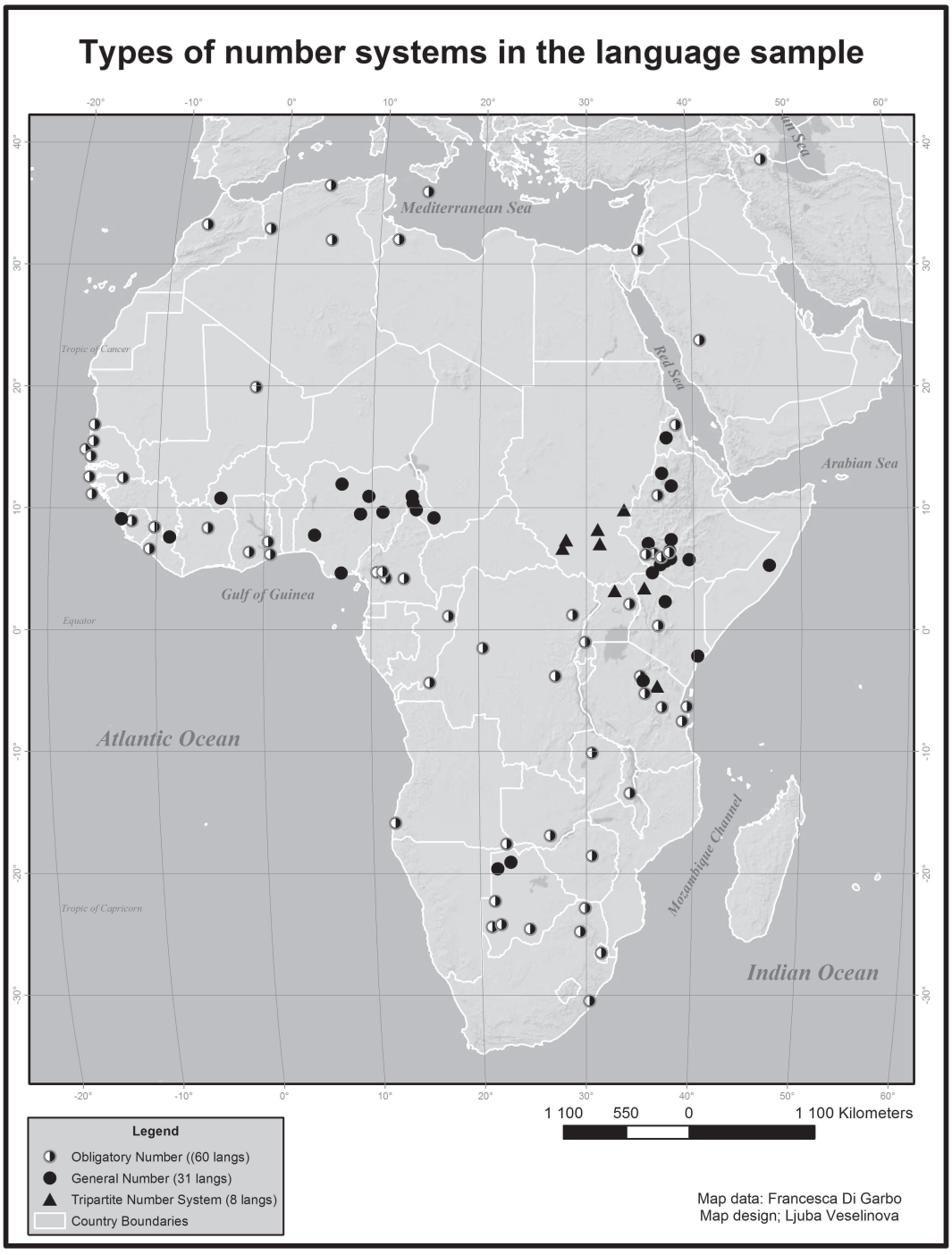


Figure 4.2: General and obligatory number in the languages of the sample

4.3.1 General number

In chapter 2, I pointed out that in a language with general number and a singular/plural opposition, ideally, each noun should have three distinguished forms: (1) the general form, interpretable both as singular and plural; (2) a form interpretable as singular and (3) a form interpretable as plural. However, as pointed out by Corbett (2000), in most languages, this type of tripartite system is often found only with a restricted number of nouns. In other words, the form used for the general meaning with some nouns is also used for the singular or the plural with other nouns. Corbett further points out that the most frequent general number system is the one in which nouns have a form that can be interpreted either as general or as singular, and another form that is interpreted only as plural. This system is labelled by Corbett as "general/singular vs. plural system." The other logically possible system, whereby the plural and the general are encoded in the same way, and there exists a dedicated marker only for the singular ("general/plural vs. singular," following Corbett's terminology), is never attested on its own. Within one language, this latter type of system tends to be restricted to a subset of nouns that are usually semantically inherently associated with plurality (Corbett 2000: 17). The languages of my sample reflect this crosslinguistic pattern.

Among the 29 languages with general number in my sample, only the Cushitic language Baiso has a unique general number form for nearly all nouns. This is shown in example 4.7. Notice that, in addition to Singular and Plural, Baiso also has a Paucal Number.

(4.7) Number distinctions in Baiso (Cushitic) (adapted from Corbett 2000: 11)

(a) *lúban*
lion.GENERAL
'lion(s)'

(b) *lubán-titi*
lion-SG
'a lion'

(c) *luban-jaa*
lion-PAUC
'a few lions'

(d) *luban-jool*
lion-PL
'lions'

In Baiso, there is also a small subset of nouns that are number-unmarked but semantically associated with plurality. These nouns are extensively discussed in §5.5.2.

In the remaining 28 languages with general number, the same form that is used to express general meaning is largely used to encode singular reference as well and is usually number-unmarked. These languages thus display a system of the "general/singular vs. plural" type. Lexical splits, whereby the general form is the same as the singular of certain nouns and the plural of other nouns, are especially common among the Cushitic

languages of the sample. In Cushitic, number-unmarked forms can be interpreted as either general, inherently singular or inherently plural.[30]

Cushitic languages have a rich inventory of morphemes that are used to specify the number value of nouns with general number as well as to form the plural of inherently singular nouns or the singular of inherently plural nouns. These morphemes are often described in reference grammars as word-formation strategies. Number marking in Cushitic is thus seen mainly as a derivational phenomenon. Examples of such word-formation processes are shown in tables 4.11 and 4.12.

Table 4.11: Singular marking in Kambaata, Borana-Arsi-Guji Oromo and Tsamai

Language	**Basic form**	**Singular marking**	**Source**
Kambata	*meentú* 'women'	*meenticcúta* 'woman'	Treis 2008
Borana-Arsi-Guji Oromo	*k'áallitʃ* 'priest(s)'	*k'áallu'* 'the priest'	Andrzejewski 1960
Tsamai	*ʔukaħ-e* 'eggs'	*ʔukaħitte* 'one egg'	Savà 2007

Table 4.12: Plural marking in Kambaata, Borana-Arsi-Guji Oromo and Tsamai

Language	**Basic form**	**Plural marking**	**Source**
Kambata	*meseleta* 'girl'	*masaláakkaáta* 'girls'	Treis 2008
Borana-Arsi-Guji Oromo	*ábba* 'father'	*ábbooti*, 'fathers, paternal uncles'	Andrzejewski 1960
Tsamai	*gurlo* 'cat(s)'	*gurlaɗɗe* 'cats'	Savà 2007

Multiple strategies of nominal plural formation are also found among the Chadic languages. For instance, 40 different plural markers, which can be in turn reduced to a dozen distinct morphological patterns of plural formation, are found in Hausa (Newman 1990, 2000). For an overview of nominal (and verbal) plurality in Chadic, see Newman (1990). Only one plural suffix is found in Bench (Ta-Ne-Omotic), in which the number-unmarked form can be interpreted as either general or singular (Rapold 2006). In Koorete (Ta-Ne-Omotic), number-marked forms are also either general or singular. Overtly marked number values are the Plural and the Paucal (Teketal 2004).

Some of the Western African languages included in the sample, such as Yoruba and Igbo, as well as the four Mande languages Bambara, Dyula, Mann and Susu have quite similar number systems. In these languages, plurality is overtly coded by means of suffixes or clitics that are diachronically related to the third person plural pronouns. Number-unmarked nouns are interpreted as general or singular. In Igbo, general number is restricted to inanimate nouns. With animate nouns, number-unmarked forms can only be interpreted as singular (Carrell 1970; Welmers 1973). These languages have, in general, rather poor morphologies.

[30] Number-unmarked nouns can also be mass or abstract nouns. These noun types can interact with number morphology in special ways. For instance, in Kambaata, mass nouns are rarely marked for number, and abstract nouns are never marked for number (Treis 2014).

In Ju|'hoan (Kxa), nominal plurality can be encoded by means of suppletion, clitical and/or suffixal markers, and associative plurals. None of these strategies is obligatory, and the overt coding of plurality is very frequently omitted in actual language use, especially with inanimate nouns (Heine & König 2011: 166). Interestingly, among the south-eastern dialects of Ju|'hoan, the marking of nominal number is becoming increasingly more systematic and is nearly mandatory for nouns denoting human entities; nonhuman nouns are still generally not marked for number (Heine & König 2011: 372).

In §4.3, I mentioned that in Amharic, number-unmarked nouns can be interpreted either as general or singular and that this is quite exceptional among the Semitic languages of the sample. I also mentioned that in Maltese, there exist nouns that may be described as displaying general number. In descriptive grammars of Maltese, and in line with the tradition of Arabic grammars, these nouns are usually called *collectives*. They are unmarked for number and mostly denote fruits, vegetables and small animals (e.g., insects). Depending on their meaning, they display different countability properties and can be interpreted either as a sort of generic nouns denoting a whole class of entities, as mass nouns or as plurals. Gil (1996) and Corbett (2000) suggest the possibility of analysing collective nouns as lying outside the singular/plural opposition, similarly to general number nouns in Cushitic languages. The status of the collective nouns in Maltese and their relationship with the same category in the Arabic dialects is not discussed in detail in this dissertation.

4.3.2 Tripartite number systems

Tripartite number system is the label conventionally used in reference grammars to refer to the strategies of number marking attested in Nilotic or, more broadly, in Nilo-Saharan languages. Nilotic languages have obligatory number; the overt coding of number distinctions in these languages follows three patterns:

1. singular marking of inherently (i.e., number-unmarked) plural nouns
2. plural marking of inherently (i.e., number-unmarked) singular nouns
3. replacement marking: both singular and plural reference are morphologically overtly coded.

Tripartite number systems are morphologically very elaborated systems in the sense that a variety of markers are available for the individual number values, and the distribution of such markers is both semantically and formally motivated. Examples of tripartite number systems in Western and Eastern Nilotic languages, the two branches of Nilotic represented in my sample, are provided below. The first set of examples is taken from Mabaan, a language of the Western Nilotic branch.

(4.8) Singular formation in Mabaan (Western Nilotic) (Storch 2005: 123)

(a) *ɲɔ́ɔ̀k*
'lice'

(b) *ɲɔ́ɔ́k-cn̰*
lice-SG
'louse'

(4.9) Plural formation in Mabaan (Western Nilotic) (Storch 2005: 123)

(a) *jwɔ́m*
'monkey'

(b) *jwɔ́m-gʌ*
monkey-PL
'monkeys'

(4.10) Replacement pattern in Mabaan (Western Nilotic) (Storch 2005: 123)

(a) *ʔúʌt-à*
worm-SG
'worm'

(b) *ʔúʌt-n̰*
worm-PL
'worms'

Eastern Nilotic languages have grammatical gender (see §4.2.2) and a tripartite number system. Interestingly, the prefixes that mark gender also encode number distinctions. Hence, in these languages, number values are encoded twice on noun stems. In spite of this redundancy, the two strategies are equally active and productive in language use (Dimmendaal 2000: 233). This type of system is illustrated with a set of examples from Turkana.

(4.11) Singular formation in Turkana (Eastern Nilotic) (Dimmendaal 1983: 224)

(a) *ŋi-sikin*
M.PL-breast
'breasts'

(b) *ɛ-sikin-a*
M.SG-breast-SG
'one breast'

(4.12) Plural formation in Turkana (Eastern Nilotic) (Dimmendaal 1983: 224)

(a) *a-ŋasep*
F.SG-placenta
'placenta'

(b) *ŋa-ŋsɛ̀p-a*
F.PL-placenta-PL
'placentas'

(4.13) Replacement pattern in Turkana (Eastern Nilotic) (Dimmendaal 1983: 224)

(a) *e-kk-ut*
M.SG-chicken-SG
'chicken'

(b) *ŋi-kuku-i*
M.PL-chiken-PL
'chickens'

A more thorough discussion of the number systems of Eastern and Western Nilotic languages is found in chapter 5.

4.4 Evaluative morphology

In my database, the criteria that I use to classify types of evaluative morphology systems are:

1. Availability of morphological evaluatives and types of evaluative distinctions
2. Presence or absence of interaction with the encoding of gender
3. Type of marking.

Criterion 1 is the most relevant here, whereas types of interactions with gender (criterion 2) and types of evaluative markers (criterion 3) are more extensively discussed in chapter 6. The following values are associated with criterion 1:

- Diminutive and augmentatives
- Only diminutives
- Only augmentatives
- None
- No data.

The figures relative to criterion 1 are presented in table 4.13. Morphological evaluative markers are found in the majority of the languages of the sample (74%). In most cases (45%), languages have relatively elaborated systems of evaluative morphology in which both diminutive and augmentative values are grammaticalized. If only one type of evaluative meaning is grammaticalized, this will always be the diminutive. Thus, the data from the languages of the sample confirm the typological generalisation according

Table 4.13: Evaluative morphology systems in the languages of the sample

	Evaluative distinctions	**No. of lngs.**	**%**	**Genealogical groups**
	Diminutives and augmentatives	45	45%	Bantu (14/23) Berber (6/6) Chadic (2/8) Cushitic (3/13) Eastern Nilotic (2/3) Hadza (1/1) Khoe-Kwadi (3/5) Kwa (1/3) Mande (1/4) North-Central Atlantic (5/7) Semitic (3/7) Ta-Ne-Omotic (2/4) Western Nilotic (2/6)
	Only diminutives	29	29%	Bantu (7/23) Chadic (1/8) Cushitic (1/13) Dizoid (1/1) Eastern Nilotic (1/3) Khoe-Kwadi (1/5) Kwa (2/3) Kxa (1/1) Mande (2/4) North-Central Atlantic (2/7) Sandawe (1/1) Semitic (3/7) South Omotic (1/1) Ta-Ne-Omotic (2/4) Tuu (1/1) Western Nilotic (2/6)
	Only augmentatives	0	0	–
	None	2	2%	Defoid (1/1) Igboid (1/1)
	No data	24	24%	Bantu (2/23) Chadic (5/8) Cushitic (9/13) Khoe-Kwadi (1/1) Mande (1/1) Mel (3/3) Semitic (1/7) Western Nilotic (2/6)
Total	100	100%		

to which diminutives are the most frequent type of evaluative morphemes (see, among others, Dahl 2006).

The geographical distribution of the types of evaluative distinctions across the sample languages is shown in the map in figure 4.3 (next page). The map in figure 4.3 does not reveal any significant genealogical and areal skewing. On the other hand, as I shall show in chapter 6, the distribution of the types of morphological marking (criterion 3) is, in some cases, areally and genealogically skewed.

Criterion 2, "Presence or absence of interaction with the encoding of gender," accounts for interactions between the morphological encoding of diminutives and augmentatives, on the one hand, and the encoding of gender distinctions, on the other. Evaluative markers appear to be connected with the marking of gender in 51 languages out of 100. Two types of interactions are found in the languages of the sample:

(1) Languages have dedicated diminutive and augmentative genders

(2) Languages do not have dedicated evaluative genders, but gender shifts (see 4.2.3 for a definition of gender shift) are equally used to encode diminutive and augmentative meanings.

Interactions between gender and evaluative morphology, their relation with types of evaluative marking and their distribution throughout the languages of the sample are thoroughly analysed in chapter 6.

Finally, it should be pointed out that the data available on evaluative morphology in the languages of my sample are somewhat less accurate than in the case of gender and number systems. No relevant information on evaluative morphology was found in the sources for 24 languages out of 100. Filling this gap would contribute to a more complete survey of the attested systems, their distribution and their properties.

4.5 Summary of the chapter

In this chapter, I provided an overview of the gender, number and evaluative morphology systems that are attested in the languages of my sample. The features used in my database for the classification of gender, number and evaluative morphology in the languages of the sample have been introduced, and their values described. The distribution of the languages of the sample with respect to each of the features has also been discussed.

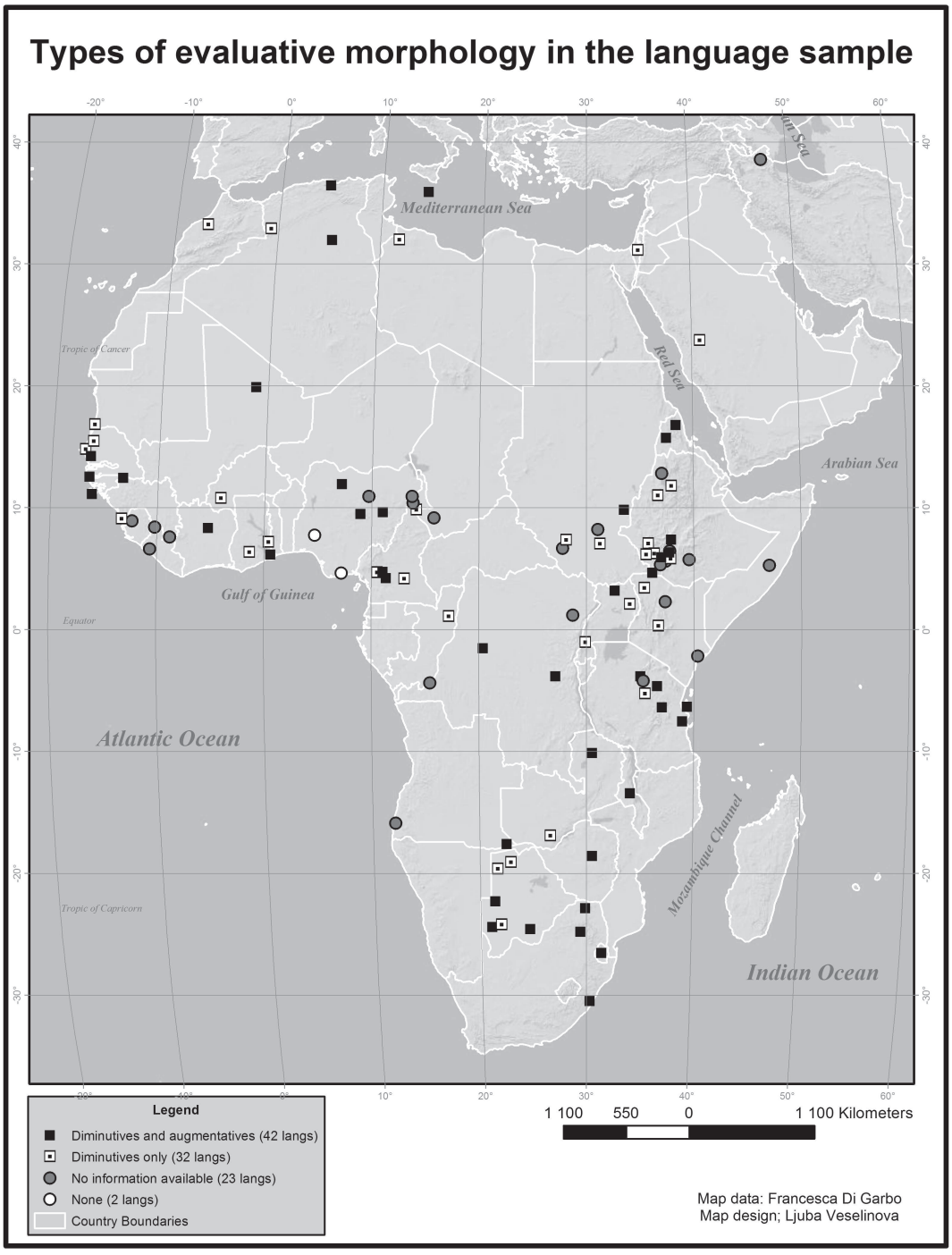


Figure 4.3: Types of evaluative morphology systems in the languages of the sample

5 Gender and number

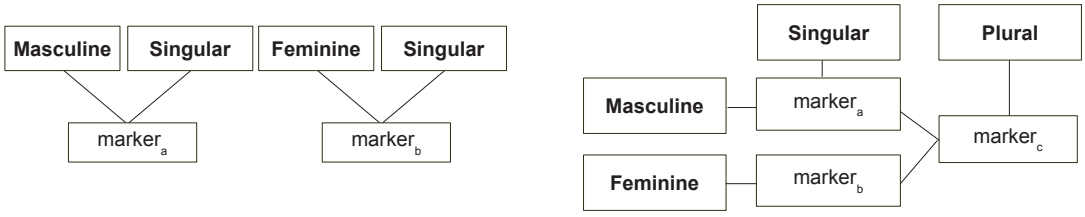
5.1 Introduction

The present chapter illustrates morphosyntactic and semantic interactions between gender and number in the 100 languages of my sample. The research questions addressed in the chapter are listed in (5.1).

(5.1) Gender and number: research questions (same as (2.13))

- Q 1: How common is cumulative exponence of gender and number in the languages of the sample?
- Q 2: What are the formal and semantic factors that trigger gender syncretism in the context of number? Does gender syncretism in the context of number presuppose cumulative exponence?
- Q 3: What are the implications of cumulative exponence and syncretism on the absolute complexity of gender and number systems?
- Q 4: Can these types of interaction between gender and number be seen as a reflex of a nominal relevance hierarchy?
- Q 5: Can gender and number compete through indexation?
- Q 6: Is there any correlation between types of encoding of gender and types of encoding of number?
- Q 7: What types of semantic interactions can be found between gender and number?

The first four research questions, which are concerned with the issues of exponence and syncretism, are addressed in §§5.2, 5.3, and 5.4. As discussed in detail in §2.5.1, *cumulative exponence* is found when the values of at least two grammatical domains – in our case, gender and number – are encoded by nonsegmentable markers. The label *syncretism* is used to describe a phenomenon whereby the values of a particular domain (e.g., in the case of gender, masculine and feminine) share a common morphosyntactic realisation in a certain conditioning environment (e.g., that of plural reference) (Baerman et al. 2005; Corbett 1991). One could think of cumulation and syncretism as having scope on two different dimensions. Cumulation acts on the syntagmatic level, as it has to do with the way in which gender and number are encoded on words in context. Syncretism functions at the paradigmatic level, as it relates to the number of gender distinctions that are available across different number values, and vice versa. An illustration of the two dimensions is provided in figure 5.1, where I represent cumulation and syncretism



(a) Cumulation (b) Syncretism

Figure 5.1: The syntagmatic and paradigatic level of the interaction between gender and number

in a hypothetical language with a sex-based gender (Masculine vs. Feminine) and two obligatory number values (Singular vs. Plural). It is worth mentioning that the schemes proposed in figure 5.1 should not be interpreted as having any diachronic reality. Particularly in the case of cumulation, I do not assume that cumulative encoding of gender and number necessarily presupposes coalescence of two previously distinguished markers. Moreover, even though independent in principle, cumulation and syncretism merge with each other frequently. As observed in §2.5.3, previous studies by Carstairs (1984), Carstairs (1987) and Carstairs & Stemberger (1988) have shown that there exists a correlation between cumulative exponence of case and number, on the one hand, and the occurrence of case syncretism in the context of number, on the other. Similarly, gender- and number-related syncretisms could be expected to be more frequent in languages where the two grammatical domains have cumulative encodings. In this chapter, I test this hypothesis on the languages of the sample.

The question of competing gender and number indexation patterns (research question 5 in (5.1)) is relevant for those languages with sex-based gender where there are either no dedicated plural indexation patterns or the use of dedicated plural indexation patterns is semantically constrained. An example of the first condition would be a language in which plural nouns always trigger the same indexation pattern, for instance, as feminine singular nouns. An example of the second condition would be a language in which a dedicated plural indexation pattern is used, for instance, only when the head noun is a plural noun denoting a human entity, or with personal pronouns with a human antecedent. This issue is discussed in §5.5.

Correlations between the presence of gender and type of number marking (question 6 in (5.1)) are investigated in order to test the claim by Creissels et al. (2008), according to which African languages without gender would tend to display less-elaborated strategies

for the marking of nominal plurality (see also §2.5.4). In §5.6, I try to verify the validity of this claim on the languages of my sample. Finally, semantic interactions between gender and number (question 7 in (5.1)) are discussed in §5.7. A summary of the patterns investigated in the chapter is found in §5.8.

5.2 Cumulation

Ideally, one could think of cumulation as a YES/NO parameter, whereby languages can be classified as either *cumulative* or *noncumulative* with respect to the encoding of two or more grammatical features. This notion of cumulation goes hand in hand with the idea that the morphology of languages can be classified holistically either as agglutinative (and thus noncumulative) or as flexive (and thus cumulative) (for a discussion of the notion of holistic morphological typology see, among others, Haspelmath 2009; Plank 1999). Typological investigations of various grammatical domains have shown that languages rarely, if ever, function so neatly. With respect to exponence, cumulative and non-cumulative strategies may co-exist in one and the same language and within one and the same grammatical domain (see discussion in §2.5.1).

In the case of grammatical features that trigger indexation, such as gender and number, splits may occur between types of exponence on the indexation triggers and on the indexing targets. Since gender is bound to indexation, cumulation between gender and number can truly be assessed only by looking at types of exponence on the indexes. However, as discussed in §5.2.1.4, in those languages in which gender is overtly coded on nouns, and nouns inflect for number, patterns of exponence on nouns are also worth investigating.

One relevant question when investigating types of exponence of grammatical features is how to treat zero-coded values. The problem is discussed by Plank (1999) (see also Bybee 1994 for a study of the meaning of zeroes within tense and aspect systems; McGregor 2003 for a caveat against the proliferation of zeroes in linguistic analysis; and Dahl 2004 for an account of zeroes and grammatical complexity). According to Plank, the way zeroes are interpreted – i.e, "in combination or opposition with something" (1999: 303) – is more relevant than actual morphological segmentability when accounting for presence or absence of cumulation. Ultimately, what is crucial to understand is how zero-coded forms interact with the rest of the paradigm and what types of oppositions there are in the system. In this thesis, cumulation values for each of the sampled languages are controlled for zeroes in the following way. As a general rule, when zero-coded values are in paradigmatic opposition to noncumulative, overtly-coded values, the zeroes are counted as noncumulative exponents. For instance, if language X has zero-marking of the Masculine Gender and the Feminine is overtly coded by means of a marker that does not cumulate with number, the zero-coded Masculine is also counted as noncumulative (examples are discussed in detail in §5.2.1.3).

The following values are associated with the feature "Cumulation" in my database:

1. Cumulative with all indexing targets

2. Cumulative with some indexing targets
3. Noncumulative
4. No gender
5. No number indexation.

Consequently, languages are classified according to total cumulation, partial cumulation and absence of cumulation, the ultimate aim being to identify and attempt to motivate possible distributional preferences in the occurrence of morphological splits related to the exponence of gender and number. The results are presented and discussed in the next section.

5.2.1 Cumulation between gender and number: results

Table 5.1 provides the figures relative to patterns of cumulation between gender and number on indexing targets. Cumulation values, percentages over the number of languages with gender (Rel. %) and over the total number of languages in the sample (Abs. %) are listed together with the distribution of each value throughout genealogical groupings.

As shown in table 5.1, the distribution of the sampled languages with respect to cumulation between gender and number is highly skewed: 81% of the languages with gender have cumulative encoding of gender and number on all indexes. Only two languages within the sample completely lack cumulation between gender and number on the indexes. Finally, none of the languages with gender in the sample completely lacks number indexation (Western Nilotic languages lack number indexation, but they also lack gender).

As mentioned before, the languages of the sample differ a great deal with respect to the number and the types of possible indexes of gender. Such differences are, however, not relevant to the data discussed in this section (see chapter 7 for the relationship between number of indexes and grammatical complexity of gender systems).

Table 5.1: Cumulation between gender and number on the indexing targets

Cumulation value	No. of lngs.	Rel.%	Abs.%	Genealogical groups
Cumulative with all indexes	68	81%	68%	Bantu (23/23) Chadic (5/8) Cushitic (6/13) Dizoid (1/1) Eastern Nilotic (3/3) Hadza (1/1) Khoe-Kwadi (5/5) Kwa (1/3) Kxa (1/1) Mel (3/3) North-Central Atlantic (6/7) Semitic (7/7) South Omotic (1/1) Ta-Ne-Omotic (3/4) Tuu (1/1) Sandawe (1/1)
Cumulative with some indexes	14	16.6%	14%	Berber (6/6) Cushitic (6/13) Chadic (1/8) Ta-Ne-Omotic (1/4)
Noncumulative	2	2.4%	2%	Cushitic (1/13) North-Central Atlantic (1/7)
No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
No number indexation	0	–	–	–
Total	100	100%	100%	

5.2.1.1 Cumulative encoding of gender and number with all the indexes

Cumulation between gender and number on the indexes is found both in languages with obligatory number marking (5.2) and in languages in which number marking on nouns is not obligatory (5.3).

(5.2) Gender and number indexation in Bandial (North-Central Atlantic) (Sagna 2008: 190)

(a) *imbi* *eno* *fi-ttix* *fafu* *fu-mug-i-mux*
PERM if CL7-war CL7:DEF CL7.3SG-kill-2SG.OBJ-DUP
'lit: If it happens that the war kills you' (If you die during the war)

(a) *imbi* *eno* *gu-ttix* *gagu* *gu-mug-ul-mux*
PERM if CL8-war CL8:DEF CL8.3PL-kill-2PL.OBJ-DUP
'lit: If it happens that the wars kill you' (If you (PL die during the wars)

(5.3) Gender and number indexation on demonstratives in Gidar (Chadic) (adapted from Frajzyngier 2008: 319)

(a) *də́f* *ə̀n-ká*
man THIS-M.SG
'this man'

(b) *də́k* *kə́-n-kə́*
woman F.SG-THIS-F.SG
'this woman'

(c) *dí* *ìn-kí*
men THIS-PL
'these men'

As in many other languages of the sample, in Gidar, gender distinctions are neutralized under plural reference (see §5.3 for more details on syncretism of gender in the context of number).[31]

The noun class system of Ju|'Hoan (Kxa) can be considered as a peculiar instance of cumulative exponence of gender and number, and deserves a short note. Ju|'Hoan has four noun class markers:

- Class 1: *ha*
- Class 2: *sì*
- Class 3: *yì*
- Class 4: *ka*

[31] The proximate demonstrative marker in Gidar is *-n*; *ə̀n-* and *ìn-* are two different allomorphs of *-n*.

With animate nouns, the gender markers also serve the purpose of encoding number distinctions. Inanimate nouns are outside the number system; with such nouns, the gender markers do not bear any relationship with the encoding of singular and plural reference. Table 5.2 illustrates the semantics of the gender system of Ju|'Hoan and, in the case of animate nouns, its relationship with the encoding of number distinctions (see Güldemann & Vossen 2000; Güldemann 2000; Heine & König 2011 for a more detailed account of gender in Ju|'Hoan, as well as §5.3.2 for gender syncretism in Ju|'Hoan).

Table 5.2: The semantics of the noun class system of Ju|'hoan (adapted from Heine & König 2013: 156)

Gender	**Number**	**Salient meanings**
CL1/CL2	Singular/Plural	Human nouns
CL1/CL3	Singular/Plural	Many animals, inanimate nouns
CL1	General number	Many plants and plant products, inanimate nouns
CL3	General number	Mostly inanimate nouns
CL4	General number	Inanimate nouns, body part nouns, abstract nouns

5.2.1.2 Cumulative encoding of gender and number with some indexes

As shown in table 5.1, in 16.6% of the languages with gender in the sample, cumulation between gender and number is not found with all the possible targets of gender and number indexation but only with some of them. This pattern of partial cumulation is mainly attested among languages of the Afro-Asiatic family. The number of cumulative and noncumulative indexes can vary a great deal across individual languages.

An interesting system is found in the Berber languages. Gender and number are noncumulative on adjectival modifiers where they are encoded in the same way as on nouns. The Masculine Gender is zero-marked, whereas the Feminine is overtly coded by means of the *t* morpheme. This, depending on the indexing target, can occur as a prefix, infix or suffix. The marking of gender and number on different types of pronouns (personal, possessive, demonstrative) and on the verbs tend to also be noncumulative. However, instances of cumulation are also found. Consider, for instance, the paradigm of the Independent Personal Pronouns in Kabyle, as represented in table 5.3.

Table 5.3: Independent Personal Pronouns in Kabyle (Mettouchi 2000)

Person	Masculine	Feminine	Cumulation value
1SG	*nekk*		–
1PL	*nek^{w}ni*	*nek^{w}n**t**i*	Noncumulative
2SG	*kecc*	*kemm*	Cumulative
2PL	*kunwi*	*kunnem**t**i*	Partially Cumulative
3SG	*neṭṭa*	*neṭṭa**t***	Noncumulative
3PL	*nutni*	*nutenti*	Noncumulative

Idiosyncrasies in inflectional patterns, as the one illustrated in table 5.3, are the result of language-internal developments whose description falls outside the scope of this investigation.

In Cushitic languages, gender and number may not be overtly coded on the same indexes. In 5 of the 13 Cushitic languages of my sample, gender and number are encoded cumulatively at least on the Independent Third Person Pronouns. A more thorough discussion of gender and number indexation in Cushitic is found in §5.5.

In Koorete (Ta-Ne-Omotic), the Demonstrative Modifiers inflect neither for gender nor for number. On the other hand, the Demonstrative Pronouns inflect both for gender and for number, but gender and number values are encoded by separate exponents. This is illustrated in example (5.4).

(5.4) Demonstrative Pronouns in Koorete (Ta-Ne-Omotic) (adapted from table in Teketal 2004: 168)

(a) *há-ʔis-a*
this-M-SG.INDF.ABS
'this'

(b) *haá-ʔis-id̕e*
this-M-PL.INDF.ABS
'these'

5.2.1.3 Noncumulative encoding of gender and number on the indexing targets

Only two languages have been found in the sample in which the encoding of gender and number on the indexes is never cumulative. These are the Cushitic language Beja and the North-Central Atlantic language Wamey. The two languages have different types of gender systems, with Beja having a sex-based gender and Wamey a non-sex-based gender.

In Beja, similarly to the Berber languages, the encodings of the Masculine Gender and the Feminine Gender on the indexes differ from each other only for the presence/absence of the feminine marker *t*. This is added to the individual singular and plural indexes without there being any interaction with the encoding of singular and plural reference (Roper 1928). The paradigm of the Definite Article in Beja is presented in table 5.4 as an illustration of this pattern of encoding.

Table 5.4: The Definite Article in Beja (Roper 1928: 9)

Case	**Masculine**		**Feminine**	
	SG	PL	SG	PL
NOM and VOC	*ū*	*ā*	*tū*	*tā*
ACC and other cases	*o*	ɛ	*to*	*tɛ*

In Wamey, gender and number indexes are also non-cumulative. Figure 5.2 provides the inventory of the singular and plural noun class markers in Wamey, and their corresponding singular-plural pairings (these markers are used both on the targets and the nouns themselves). The noun classes of Wamey have CV syllable structure, whereby C is always a labial consonant (*v* or *w*). The vocalic component of the CV template is, in nearly all cases, the vowel of the corresponding singular class. The ordinal numbers that occur after the noun class prefixes in figure 5.2 indicate the type of consonant alternation that the prefix triggers on the noun stem (see §4.2.3.2 for a definition of consonant alternation in the Atlantic languages). It is worth mentioning that certain class pairs (e.g., 5/6 and 7/8) are identical with each other in terms of their phonological appearance but trigger different patterns of consonant alternation on the nouns and on the indexing targets. They are thus classified by Santos (1996) as different genders.

Concerning the morphology of the plural class markers in Wamey, it has been suggested (Santos 1996; Ferry & Pozdniakov 2001; Pozdniakov 2010, 2013) that the labial component of the plural noun classes – *v-* (and *w-*) – is, in fact, a plural marker that precedes the actual noun class marker. This generalised plural morpheme is built on the model of the plural noun class for human nouns, *və-*. At some point in the history of the language, the labial consonant must have been interpreted as a plural morpheme and was then extended to the entire paradigm, independently of the animacy of the noun. Konstantin Pozdniakov (personal communication) suggests that the phenomenon might have originated from the use of markers of Classes 1 and 2 as generalised anaphoric pronouns. In figure 5.2, the labial component of the plural noun classes is presented in boldface.

In sum, as a result of innovation in the encoding of plural reference, in Wamey plurality is always overtly coded by means of a labial consonant, whereas zero-coding of number is interpreted as expressing singular reference. Gender and number are thus no longer coded cumulatively.

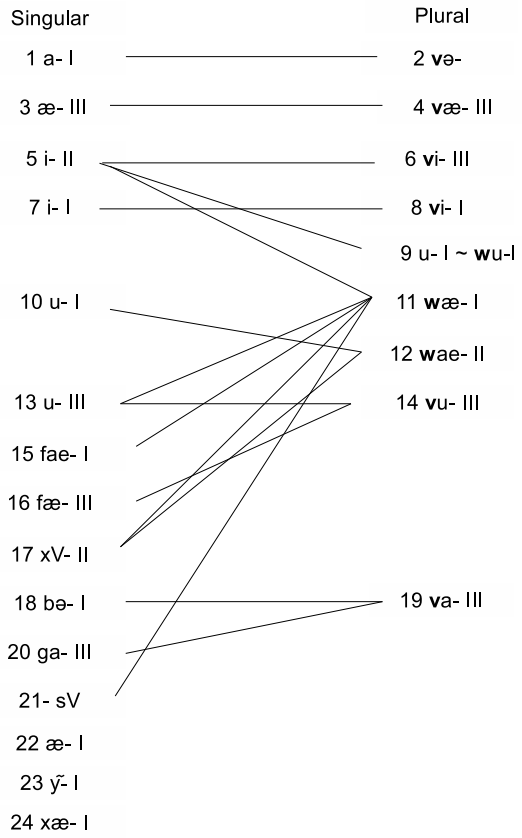


Figure 5.2: The noun class system of Wamey (North-Central Atlantic) (Santos 1996: 145)

5.2.1.4 Cumulation between gender and number on nouns

When gender is overtly coded on nouns (see data and discussion in §4.2) and nouns inflect for number, interesting facts can be observed with respect to the exponence of gender and number values. Even though this phenomenon is not central to the current investigation of gender systems, I tried to account for patterns of cumulation between gender and number on nouns in my database by adding an additional feature: "Cumulation between gender and number on nouns." The following values are assigned to this feature:

1. Cumulative
2. Cumulative SG vs. noncumulative PL
3. Noncumulative with specific gender (G) and number (N) markers
4. Noncumulative
5. No gender marking on nouns
6. No gender.

Table 5.5 provides the figures relative to the distribution of patterns of cumulation between gender and number on nouns across the languages of the sample. As shown in the table, 15 out of the 84 languages with gender in the sample do not have overt marking of gender on nouns and are thus excluded from the count of patterns of cumulation between gender and number on nouns. In slightly more than half of the languages with gender in the sample, the encoding of gender and number on nouns is cumulative. Cumulative encoding of gender and number on nouns is found throughout the individual genealogical groupings represented in the sample. Immediately after come the 17 languages with noncumulative marking, which constitute the 20.2% of the languages with gender. These 17 languages belong to the following genealogical units: Berber, Chadic, Cushitic, Dizoid, North-Central Atlantic and Ta-Ne-Omotic. Eight of these seventeen languages mark number obligatorily on nouns, whereas nine have optional number marking on nouns. The remaining 9.5% of the gendered languages of the sample is characterised by some sort of split system whereby gender and number values on nouns can have both cumulative and noncumulative exponence. In my database, these languages are classified as belonging to one of the two groups in between "cumulative" and "noncumulative" that are listed in table 5.5. These two groups are discussed in detail in the two next following sections.

Table 5.5: Cumulation between gender and number on nouns

Cumulation value	No. of lngs.	Rel.%	Abs.%	Genealogical groups
Cumulative	44	52.4%	44%	Bantu (22/23) Cushitic (2/13) Hadza (1/1) Khoe-Kwadi (5/5) Kwa (1/3) Mel (3/3) North-Central Atlantic (5/7) Semitic (2/7) Ta-Ne-Omotic (2/4) Tuu (1/1)
Cumulative SG vs. noncumulative PL	1	1.2%	1%	Bantu (1/23)
Noncumulative with specific G and/or N markers	7	8.3%	7%	Eastern Nilotic (3/3) Semitic (4/7)
Noncumulative	17	20.2%	17%	Berber (6/6) Chadic (2/8) Cushitic (6/13) Dizoid (1/1) North-Central Atlantic (1/7) Ta-Ne-Omotic (1/4)
No gender marking on nouns	15	17.9%	15%	Chadic (4/8) Cushitic (5/13) Kxa (1/1) North-Central Atlantic (1/7) Sandawe (1/1) Semitic (1/1) South Omotic (1/1) Ta-Ne-Omotic (1/6)
No gender	16		16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total	100	100%	100%	

5.2.1.4.1 Cumulative singular vs. noncumulative plural This is attested only in one language of the sample: Kinshasa Lingala (Bantu). The noun class system of Kinshasa Lingala is a reduced version of the system that is found in the standard variety of the language, Makanza Lingala. Examples (5.5) and (5.6) illustrate the similarities and differences among the two varieties.

(5.5) Makanza Lingala (Bantu) (Bokamba 1977: 184)

(a) *mo-to*
CL1-person
'person'

(b) *ba-to*
CL2-person
'people'

(c) *li-loba*
CL5-word
'word'

(d) *ma-loba*
CL6-word
'words'

(e) *n-dako*
CL9-house
'house'

(f) *n-dako*
CL10-house
'houses'[32]

(5.6) Kinshasa Lingala (Bantu) (Bokamba 1977: 184)

(a) *mo-to*
CL1-person
'person'

(b) *ba-to*
CL2-person
'people'

(c) *li-loba*
CL4-word
'word'

(d) *ba-ma-loba*
PL-CL6-word
'words'

(e) *n-dako*
CL9.house
'house'

(f) *ba-n-dako*
PL-CL10-house
'houses'[33]

The striking difference between Makanza and Kinshasa Lingala is that, along with nouns that do not belong to Gender 1/2, Kinshasa Lingala seems to mark plurality twice: (1) by regular class shift and (2) by adding the prefix *ba-*, that is, the plural class prefix for nouns of Gender 1/2. The plural marker initially used only with human nouns has been generalised as a plural morpheme for all nouns. This has given rise to a situation in which number marking on nouns is somewhat redundant, and gender and number on nouns are, at least in part, noncumulative. The generalised use of Class 2 as a plural marker on nouns is also found in other Bantu languages, where it is usually semantically

[32] In Lingala, as in other Bantu languages, Classes 9 and 10 are not distinguishable on nouns. They do, however, trigger different indexation patterns on the indexing targets.

[33] See footnote 29.

constrained in the sense that it is restricted to animate nouns only (Maho 1999: 134-136). One could speculate that, initially, the use of class 2 as a generalised nominal plural marker in Kinshasa Lingala might have been also semantically constrained. Interestingly, even though the patterns of gender and number indexation of Kinshasa Lingala are quite impoverished, gender and number are still encoded cumulatively on the indexing targets (Bokamba 1977 as well as §7.7).

5.2.1.4.2 Noncumulative with specific gender and/or number markers This pattern is found in the Eastern Nilotic and some Semitic languages of the sample. It only concerns nouns and does not have any parallels on the indexing targets. The Eastern Nilotic languages Turkana, Masai and Karamojong mark number twice. Every noun is marked by prefixes that express gender and number cumulatively. In addition, number is marked by singular and plural suffixes, which do not cumulate with gender. This is illustrated in (5.7) with an example from Turkana.

(5.7) Gender and number marking in Turkana (Eastern Nilotic), same example as (4.13) (Dimmendaal 1983: 224)

(a) *e-kk-ut*
M.SG-chicken-SG
'chicken'

(b) *ŋi-kuku-i*
M.PL-chiken-PL
'chickens'

Such a redundancy in number marking is considered to be the result of historical developments that are internal to the Eastern Nilotic genealogical group. The singular and plural suffixes of the Eastern Nilotic languages are shared with the other languages of the Nilotic genus and are inherited from the protolanguage. On the other hand, the gender/number cumulative prefixes are a later innovation within Nilotic, and are only attested in the Eastern Nilotic branch. A more thorough discussion of the number system of the Nilotic languages is found in §5.6.

In Standard Arabic, Moroccan Arabic, Maltese and Tigre, nominal plurality on nouns can be encoded either by suffixes or by patterns of vowel alternation on the nouns stem (also known as *Broken Plurals*). Broken Plurals in these languages never convey information about gender. Example (5.8) illustrates one type of Broken Plural in Standard Arabic.

(5.8) Broken Plurals in Standard Arabic (Semitic) (Ryding 2005: 146)

(a) *Taalib*
student\[SG]
'students'

(b) *Tullaab*
student\PL
'students'

5.2.1.5 Summary

In this section, I discussed the occurrence of cumulative exponence of gender and number in the languages of the sample (thus addressing question 1 in (5.1)). The data show that the large majority of the languages of the sample have cumulative encoding of gender and number on all indexing targets. Noncumulative, or partially cumulative, encodings of gender and number tend to be found in languages where nominal number marking is not obligatory, but they are also attested in languages with obligatory number (e.g., Berber). As far as exponence of gender and number on nouns is concerned, with the exception of the Berber languages and the North-Central Atlantic language Wamey, noncumulative encodings tend to be found in languages where number marking on nouns is not obligatory but gender is. More hybrid patterns are also attested and are usually the result of language-internal developments.

5.3 Syncretism

Syncretism is a type of paradigmatic asymmetry that features formal identity of cells within an inflectional paradigm (see Baerman et al. 2005 as well as §§2.5.2 and 5.1).

Gender distinctions are particularly prone to undergo syncretism across number values. For instance, the three gender values of language X can be conflated under dual reference but still be distinguished under singular and plural. In addition, the occurrence of gender syncretism may also be restricted by indexing targets. For instance, the demonstratives in language Y may be insensitive to the distinction between the feminine and neuter gender, which is otherwise overtly coded on other indexing targets, such as adjectives and verbs. In this thesis, I mainly investigate instances of syncretism of the first type.

In §2.5.2, I surveyed previous crosslinguistic studies on the occurrence of gender syncretism across number and observed that most of these studies capitalise on the idea that gender and number-related syncretism reflects markedness relationships within and across the two grammatical domains. In particular, I showed that gender distinctions tend to be syncretic in the context of number. In general, gender distinctions are reduced or neutralized under nonsingular number values. I now turn to an overview of the results.

5.3.1 Syncretism of gender in the context of number: results

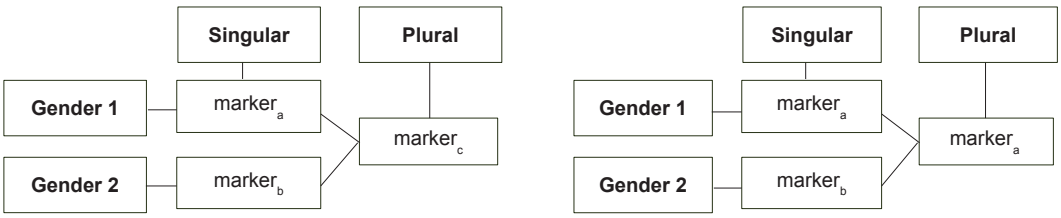
Overall, 67 languages among the 84 gendered languages in the sample have syncretism. In addition, the data suggest that there is a strong correlation between patterns of syncretism between gender and number, on the one hand, and cumulative exponence of gender and number, on the other. This is shown in table 5.6.

The results suggest that in the languages of my sample, the possibility of having syncretism without cumulation is nearly ruled out. The only two languages where syncretism is found in the absence of cumulation are Beja and Wamey (see §5.2.1.3). Otherwise, in my sample, all the languages with syncretism also have either total or partial cumulative encodings of gender and number. As discussed later in §5.4, this is in line with the results of Carstairs (1984), Carstairs (1987), and Carstairs & Stemberger (1988) with respect to case syncretism in the context of number.

Two types of gender syncretism occur under plural reference in the languages of my sample:

(1) The syncretic plural marker is different from each of the correspondent singular markers (Figure 5.3a, examples follow throughout the section)

(2) The syncretic plural marker is the same as one of the correspondent singular markers (Figure 5.3b, examples follow throughout the section).



(a) Syncretism 1 (b) Syncretism 2

Figure 5.3: Types of gender syncretism across number

Patterns of syncretism in which the syncretic marker is formally identical to one of its component values – as in (2) – are generally less frequent than patterns of syncretism in which there is no such formal identity – as in (1). The same crosslinguistic tendency is reflected by the languages of my sample. Syncretisms of the type described in (1) are the most frequent.

Table 5.6: Gender syncretism and its relationship with cumulation in the language sample

Syncr.	Cumulation value	No. lngs.	Rel.%	Abs.%	Genealogical groups
✓	Cum.: all indexes	57	68%	57%	Bantu (22/23) Chadic (5/8) Cushitic (6/13) Dizoid (1/1) Eastern Nilotic (3/3) Kwa (1/3) Mel (3/3) North-Central Atlantic (6/7) Sandawe (1/1) Semitic (5/7) South Omotic (1/1) Ta-Ne-Omotic (3/4)
✓	Cum.: some indexes	8	9.5%	8%	Chadic (1/8) Cushitic (6/13) Ta-Ne-Omotic (1/4)
✓	Noncum.	2	2.4%	2%	Cushitic (1/3) North-Central Atlantic (1/7)
	Cum.: all indexes	11	13.1%	11%	Bantu (1/23) Hadza (1/1) Khoe-Kwadi (5/5) Kxa (1/1) Semitic (2/7) Tuu (1/1)
	Cum.: some indexes	6	7%	6%	Berber (6/6)
No gender	No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total		100		100%	

For instance, in Turkana (Eastern Nilotic), nouns are assigned either to the Masculine, the Feminine or the Neuter Gender. The Masculine and Neuter genders are syncretic under plural reference. The plural marker for masculine and neuter nouns differs from both the corresponding masculine singular and neuter singular markers. This is shown in table 5.7 in which the paradigm of the restrictive markers is provided.[34]

Table 5.7: Gender syncretism in Turkana restrictive markers (adapted from Dimmendaal 1983: 217)

	Singular	**Plural**
F.	*a-*	*ŋa-*
M.	*e-*	*ŋi-*
N.	*i-*	*ŋi-*

On the other hand, gender syncretisms of the type described above in (2), and represented in figure 5.3b, are quite rare. Only 8 of the 67 languages with syncretism in my sample have syncretism of type (2). They are, in order of genealogical affiliation: Baiso (Cushitic), Borana-Arsi-Guji Oromo (Cushitic), Dasaanach (Cushitic), Iraqw (Cushitic), Kambaata (Cushitic), Bench (Ta-Ne-Omotic) and Standard Arabic (Semitic). Moreover, it seems that syncretism of type (2) never exists on its own in a language. In fact, in all the cases where this pattern is attested within my sample, either syncretism of type (1) or no syncretism at all are also found. For instance, in Standard Arabic, plural nouns trigger different types of gender and number indexation depending on the animacy of the NP referent. Two patterns are attested:

(1) Plural nouns denoting human beings do not undergo gender syncretism under plural reference (5.9).

(5.9) Gender and number indexation with plural human nouns in Standard Arabic (Semitic) (Ryding 2005: 126)

(a) *qaadat-un* *ᶜaskariyy-uuna*
PL\leader-NOM military-M.PL
'military leaders'

(b) *al-nisaaʾ-u* *l-mutaqaddim-aatu* *fii* *l-sinni-i*
DEF-women-NOM DEF-avanced-F.PL in DEF-age-GEN
'women of advanced age'

[34] The notion of restrictiveness in Turkana is "allied with, but not exactly equivalent to, definiteness" (Baerman et al. 2005: 84, fn 25).

(2) Nonhuman plural nouns trigger the same indexation pattern as feminine singular (5.10)

(5.10) Gender and number indexation with non-human plural nouns in Standard Arabic (Semitic) (Ryding 2005: 126)

al-funuun-u *l-mu^{c}aaSir-atu*
PL\DEF-art-NOM DEF-contemporary-F.SG
'contemporary arts'

Sometimes, when construed as abstract nouns (Ryding 2005: 126), human collectives also trigger feminine indexation:

(5.11) Gender and number indexation with human collectives in Standard Arabic (Semitic) (Ryding 2005: 126)

al-suluTaat-u *l-ruumaaniyy-atu*
PL\DEF-autorithy-NOM DEF-Roman-F.SG
'the Roman authorities'

The gender and number system of Standard Arabic is thus structured in such a way that gender distinctions under plural reference are maintained only when the referent of the NP is human. Instead of being distributed arbitrarily to one gender or the other – as it happens under singular reference – all nonhuman plural nouns behave in the same way with respect to the indexation patterns that they trigger. Thus in Standard Arabic, the occurrence of syncretism of type (2) is constrained by the semantics of the head noun.

The languages with syncretism of type (2), and the way their distribution competes with that of languages with type (1) syncretism, or no syncretism at all, are surveyed in §5.5 in the context of a discussion on restrictions on the use of dedicated patterns of plural indexation.

5.3.2 Semanticization of gender distinctions under syncretism

One of the typological generalisations on gender syncretism formulated by Baerman et al. (2005) is that in languages with large gender systems, not only are gender distinctions reduced under plural reference, but they also tend to be *consolidated* and *realigned* on the basis of more transparent semantic criteria than the corresponding singular forms. Baerman et al. refer to this phenomenon as *semanticization* of gender distinctions under plural reference (2005: 86). Their generalisation has been tested on the 34 languages with large gender systems in my sample. Only four of these languages confirm this tendency: the Maasina Fulfulde (Nort-Central Atlantic, North), Nuclear Wolof (North-Central Atlantic, North), Kisi (Mel) and Gola (Mel). Two of these languages, Maasina Fulfulde and Kisi, are also part of the sample created by Baerman et al. (2005). In all four languages, the increase in the semantic predictability of gender distinctions in the plural is animacy-driven. The patterns of gender syncretism across number in the Maasina Fulfulde and Kisi were previously discussed by Baerman et al. (2005: 87-89). In the rest of this section, I mostly concentrate on the systems of Gola and Nuclear Wolof.

The gender system of Gola is outlined in table 5.8. Besides the class pairs illustrated in the table, Gola also has one locative class that does not distinguish number.

Table 5.8: Gender and number distinctions in Gola (adapted from Fachner 1990: 10)

Singular	**Plural**	**Semantics of the plural classes**
wo-	aŋ-	Animate
wɔŋ-	maŋ-	
ke-	maŋ-	Inanimate
e-	maŋ	
ke-	le-	Sgv. vs. Clv.[35]

As shown in table 5.8, the marking of nominal plurality in Gola is bound to the opposition between animate and inanimate nouns. Gender distinctions are reshuffled under plural reference on the basis of this clear-cut semantic opposition.

A similar pattern of gender syncretism across number is found in Nuclear Wolof.[36] Nuclear Wolof has eight singular classes and two plural classes, which are illustrated in table 5.9.

Table 5.9: Gender and number distinctions in Nuclear Wolof (adapted from McLaughlin 1992: 200)

Singular	**Plural**
k-	ñ-
b-	y-
g-	y-
j-	y-
l-	y-
m-	y-
s-	y-
w-	y-

Class *k-* and *ñ-* are very restricted in use. Interestingly, the semantically opposite nouns *nit* 'person' and *këf* 'thing' are the only nouns assigned by default to class *k-* when singular. Class *ñ-* is the regular plural for *nit* and, depending on the speakers, is also used as the plural marker for the following highly animate nouns: *jigéen* 'women,' *góor* 'men' and *gaa* 'people.' All other nouns in Nuclear Wolof, including *këf* 'thing,' form their plural in Class *y-* (McLaughlin 1992: 200-201). The massive reduction of the number of class distinctions under plural reference is accompanied, at least for some speakers, by some sort of semantic realignment whereby the most generic human nouns – 'man,' 'woman' and 'person' – receive the same type of plural marking.

[35] This is an extremely rare gender in Gola. Class *le-* expresses collective meanings, and Class *ke-* is sometimes used to derive singulative-like nouns from collective nouns, e.g., *le-kuo* 'a collection of eggs,' *ke*-kuo 'a single egg' (Fachner 1990: 10).

[36] The variety of Wolof referred to in this thesis is the one described by McLaughlin (1992, 1997). In Glottolog (Nordhoff et al. 2013), this is referred to as *Nuclear Wolof*.

In Kisi, the plural classes are also semantically more clear-cut than the singular. What is typologically more exceptional in the gender system of Kisi is that the singular noun classes are more syncretic than the plural. There are three distinct noun classes under singular reference and five under plural. In Baerman et al.'s (2005) typological study of syncretism, Kisi stands out as a fairly unique instance of gender-related syncretism. This applies to my language sample, too (for a more detailed discussion of the gender system of Kisi, see Baerman et al. 2005: 88-89).

In the majority of cases, gender syncretism under plural reference in languages with large gender systems does not seem to work on the basis of semantically transparent criteria. For instance, the North-Central Atlantic language Noon has six genders with six singular classes and two plural classes. Synchronically, the distribution of the two plural classes over the singular is not semantically motivated. Animate nouns receive extra marking on the determiners under both singular and plural reference. Such markers are, however, extraneous to the gender system (Soukka 2000: 65-66). Maho (2003) carried out a survey of syncretic plural classes in the Bantu languages that led to similar results. In his paper, he refers to plural classes that pair with more than one singular class as *polyplural* classes. The use of the syncretic plural classes in Maho's sample is not semantically constrained nor is their distribution specifically restricted to individual areas within the Bantu-speaking world. Thus, syncretism of certain gender distinctions under plural reference can be placed far earlier in the history of the Bantu languages, and its semantic motivation, if any, is impossible to trace (Maho 2003: pp.165-169).

One last interesting case among the languages of the sample with non-sex-based gender is Ju|'Hoan (Kxa). Based on Dickens (1992), Baerman et al. (2005) describe the noun classes of the language as a very rare type of system, where gender distinctions under plural reference outnumber gender distinctions under singular reference, and the latter are semantically more transparent than the former. In fact, as I mentioned in §5.2.1, in Ju|'Hoan, only with animate nouns do gender markers encode number as well. Inanimate nouns have general number, and, for them, gender assignment is not relevant to the encoding of number distinctions. Ultimately, one could suggest that the system of gender indexation of Ju|'Hoan falls outside the scope of syncretism.

Baerman et al. (2005: 84-86) suggest that the semanticization of gender distinctions under non-singular reference can also be found in languages with smaller gender systems. In a tripartite sex-based gender system of the "masculine vs. feminine vs. neuter type," the masculine and feminine genders are generally more likely to undergo syncretism because, on the basis of their respective semantic cores, they constitute a sort of natural class with shared semantic features such as HUMAN or ANIMATE (Baerman et al. 2005: 84). However, the opposite is also found in the languages of the world. For instance, among the languages of my sample, as well as in the sample investigated by Baerman et al., Turkana has a tripartite gender system of the type described above, but the Masculine Gender is syncretic with the Neuter rather than with the Feminine (see Table 5.7).

Beside gender syncretism across number, an interesting case of gender syncretism restricted by target is found among the languages with non-sex-based gender in my sample. The phenomenon is traditionally known as *animate concord* (Wald 1975; Bokamba 1977;

Maho 1999) and occurs when, at least for some targets, gender indexation is determined on the basis of the animacy of nouns rather than on the basis of their morphological properties. As pointed out in §5.3.1, gender syncretism is restricted by targets when two nouns that in principle belong to two different genders trigger the same gender indexation pattern only with some targets. Animate concord is a type of semantically motivated syncretism restricted by targets. It thus triggers semanticization of gender distinctions in Baerman et al.'s (2005) terms. Animate concord is found in 11 of the 35 languages with non-sex-based gender in my sample[37]. These languages differ in the number and types of indexing targets that show animate concord. In urban varieties of Swahili, for example, animate concord is highly pervasive, since it is used with all targets but possessive modifiers (Wald 1975). This is shown in example (5.12): gender indexation on the verb is based on the referential properties of the NP (Class 1 indexation is selected because the NP referent is animate), whereas gender indexation on the possessive modifier is based on the lexical gender of the noun.

(5.12) Swahili (Bantu) (Wald 1975)

Rafiki	*y-angu*	*a-me-fika*
CL9.friend	CL9-of.me	CL1-PRF-arrive

'My friend has arrived'

In Nyanja, animate concord is quite marginal since it is only attested with the Personal Pronouns, and in pragmatically marked contexts (Corbett & Mtenje 1987: 13-14). All 11 languages with animate concord in the sample are either North-Central Atlantic, Bantu or Mel, with the exception of !Xóô, which is a Tuu language. There is a tendency for languages with animate concord to cluster areally (Wald 1975; Childs 1983; Ström 2011). Interestingly, in my sample, animate concord is found among the Languages of Wider Communication[38] (e.g., Lingala, Nuclear Wolof, Swahili, etc.) or among languages that have intense contact with Languages of Wider Communication (e.g., Ndengereko, with respect to Swahili). According to Maho (1999: 123), it is reasonable to think that the presence of animate concord might be underrepresented in reference grammars due to the fact that it is often perceived as a substandard phenomenon.

5.3.3 Summary

In this section, I discussed the occurrence of gender syncretism in the context of number in the languages of the sample. The data from the language sample suggest that the odds of gender being syncretic across number are higher when gender and number values are encoded cumulatively. In addition, they also suggest that patterns of gender syncretism

[37] These languages are: Gola (Mel), Kinshasa Lingala (Bantu), Kisi (Mel), Maasina Fulfulde (North-Central Atlantic), Ndengereko (Bantu), Noon (North-Central Atlantic), Nyanja (Bantu), Swati (Bantu), Swahili (Bantu), Timne (Mel), !Xóô (Tuu).

[38] Languages of Wider Communication are languages that are used by individuals with different mother tongues to communicate with each other (Bokamba 2009; Heine 1970; Mufwene 2003).

across number are usually difficult to motivate from a semantic point of view. There are a few exceptions whereby gender syncretism across number is driven by the animacy of the nouns (§5.3.2).

5.4 Cumulation and syncretism: summary and discussion

Cumulative exponence of gender and number turned out to be very frequent in the languages of my sample and so did syncretism. In addition, I found that syncretism of gender in the context of number largely entails cumulative exponence of the values of the two domains. These results confirm what has already been found by Carstairs (1984), Carstairs (1987), and Carstairs & Stemberger (1988) with respect to patterns of *case* syncretism in the context of number. Syncretism of the values of one grammatical domain (e.g., *gender* or *case*) in the context of another domain (e.g., *number*) is more likely to occur when the two domains have cumulative exponence.

In the rest of this section, I discuss cumulative exponence and syncretism in relationship with language complexity (question 3 in (5.1)) and consider the significance of these findings for the notion of nominal relevance hierarchy (question 4 in (5.1)).

5.4.1 Cumulation, syncretism, and grammatical complexity

As mentioned in §2.7, in this thesis I interpret language complexity in an absolute sense, that is, as an objective property of a given system (Dahl 2004; Miestamo 2008; Sinnemäki 2011). Absolute complexity can be measured on the basis of the number of parts that a system is made up of or the length of its description. One of the assumptions of this approach to grammatical complexity is that the least complex grammatical domain is one in which each form is associated with only one meaning (this is referred to as *Principle of One-Meaning–One-Form* in Miestamo 2008). Within this assumption, cumulation increases the absolute complexity of grammatical domains. A cumulative marker is associated with at least two grammatical meanings, and this determines asymmetries in the relationship between meaning and form.

With respect to syncretism, my results show that syncretic paradigms also tend to be cumulative paradigms that in turn feature an increase in absolute complexity. The presence of syncretism thus *presupposes* absolute complexity but does not necessarily *add* further complexity to the structure of inflectional paradigms. These notions are operationalised in chapter 7, where a metric for gender complexity is proposed (see also Dahl: 188, and the discussion in §2.7).

5.4.2 Syncretism and relevance hierarchy for nominal features

A few studies (Carstairs 1987; Carstairs & Stemberger 1988; Greenberg 1963b; Vafaeian 2013), chronologically quite distant from each other and conducted on the basis of independent datasets, have suggested that there exist asymmetries in the behaviour, and mutual interactions, of different features of nominal morphology (see §2.5.3 for an introduction). Such asymmetries have been interpreted as pointing at the existence of a

relevance hierarchy for nominal features similar to the one proposed by Bybee (1985) for the verbal domain. The above-mentioned studies focus on two phenomena:

(1) Syncretism, based on the assumption that the most relevant features are the least likely to undergo syncretism (Carstairs 1987; Carstairs & Stemberger 1988; Greenberg 1963b).

(2) Suppletion, based on the assumption that the higher a feature is on a relevance hierarchy, the more likely it is to trigger suppletion (Vafaeian 2013, based on the results of Veselinova 2006 on patterns of verbal suppletion).

The results of these studies converge in pointing to number as the highest ranking feature of a possible nominal hierarchy for nouns. Greenberg (1963b), Carstairs (1987) and Carstairs & Stemberger (1988) show that number – as opposed to grammatical gender (Greenberg 1963b) and case (Greenberg 1963b; Carstairs 1987; Carstairs & Stemberger 1988) – is the grammatical feature which is less likely to undergo syncretism. Similarly, Vafaeian (2013) shows that number is the grammatical feature that more frequently triggers suppletion in nominal paradigms (followed by possession, accusative/ergative case, and vocative).

The investigation of syncretism conducted in this thesis continues the line of research described in (1). In particular, the results discussed in this section provide further support for the argument of Greenberg (1963b) with respect to gender and number. Number is not likely to undergo syncretism in the context of gender, whereas the opposite is very frequent. Thus the relationship between the two features is strongly asymmetrical. If these asymmetries are interpreted as pointing to a relevance hierarchy, then the results of this investigation suggest that nominal number has higher relevance to nouns than gender.

5.5 Split plural indexation systems

As observed throughout chapter 2, manifestations of grammatical gender are bound to indexation, whereas nominal number may (but needs not to) be manifested through indexation. In §§5.2 and 5.3, I showed that if a language has both gender and number indexation, the two are likely to interact via exponence, syncretism or both. In this section, I consider those rare cases in which languages with sex-based gender exhibit a split in the indexation patterns triggered by plural nouns (question 5 in (5.1)). I suggest referring to these systems as *split plural indexation* systems. In general, split plural indexation systems function in such a way that:

(1) Some plural nouns trigger the same indexation pattern as either feminine singular or masculine singular nouns, or the same indexation pattern as their correspondent singular.

(2) Some plural nouns trigger an indexation pattern that differs both from feminine and from masculine. Given that this indexation pattern is exclusively used to signal plurality, I refer to it as Dedicated Plural Indexation (henceforth, DPI).

Eight languages of my sample have split plural indexation systems. They all have bipartite sex-based gender and belong to different subgroupings of the Afro-Asiatic language family or to Ta-Ne-Omotic. They are (in order of genealogical affiliation): Miya (Chadic), Baiso (Cushitic), Borana-Arsi-Guji Oromo (Cushitic), Dasaanach (Cushitic), Iraqw (Cushitic), Kambaata (Cushitic), Bench (Ta-Ne-Omotic) and Standard Arabic (Semitic). In these languages, the distribution of DPI depends on the type of indexation trigger or the type of indexing target. Restrictions on the use of DPI based on the type of target are only found in the Cushitic languages Kambaata and Dasaanach. These languages are discussed in §5.5.2.3. Restrictions on the use of DPI based on indexation trigger types are found in the remaining six languages. They are constrained either by the animacy or by the lexical plurality of nouns. Restrictions based on animacy are discussed in §5.5.1; restrictions based on the lexical plurality of nouns are discussed in §5.5.2. In the following sections, I will show that these phenomena, as rare as they are, can be very relevant for understanding how gender and number indexation systems come to interact and to be mutually constrained.

5.5.1 Animacy-based uses of Dedicated Plural Indexation

In §5.3, I surveyed patterns of gender syncretism in the context of number and analysed the distribution of plural indexation in Standard Arabic. In Standard Arabic, DPI is only used with plural human nouns, that is, with nouns denoting entities that are located at the top of the Animacy Hierarchy. When the plural noun is nonhuman, the feminine singular indexation pattern is used. Consider the following examples, already discussed in §5.3.1.

(5.13) Indexation with human plural nouns in Standard Arabic (Semitic) (Ryding 2005: 126) (same example as (5.9))

(a) *qaadat-un* *ᶜaskariyy-uuna*
leader\PL-NOM military-M.PL
'military leaders'

(b) *al-nisaaʾ-u* *l-mutaqaddim-aatu* *fii* *l-sinni-i*
DEF-women-NOM DEF-avanced-F.PL in DEF-age-GEN
'women of advanced age'

(5.14) Indexation with nonhuman plural nouns in Standard Arabic (Semitic) (Ryding 2005: 126) (same example as (5.10))

al-funuun-u *l-muᶜaaSir-atu*
DEF-art\PL-NOM DEF-contemporary-F.SG
'contemporary arts'

In the nonstandard dialects of Arabic, the distribution of DPI is fuzzier than in the standard. Belnap & Shabaneh (1994), Belnap (1999) and Corbett (2000) investigate variation in the indexation patterns selected by plural NPs in Cairene Arabic. They

show that, even though in Cairene Arabic DPI is more likely to be used when the NP refers to a plurality of humans, animals or individuated entities, or when the target is distant from its trigger, the indexation patterns associated with plural NPs are still relatively unpredictable. Furthermore, based on a corpus of Old Arabic texts from the 6th and 7th century, the studies by Belnap & Shabaneh (1994) and Belnap (1999) show that in Old Arabic the use of feminine indexation with non-human plural NPs was also less predictable (and less frequent) than in Standard Arabic. The more categorical use of feminine singular indexation with non-human plural NPs is thus described by the authors as an innovation of Standard Arabic, not equally shared by the nonstandard varieties (i.e., New Arabic). In Moroccan Arabic, the only nonstandard dialect of Arabic included in my sample, DPI with plural NPs is described as the most frequent option (Harrell 1965: 158). Feminine indexation is rarely used and is described by Harrell as an idiosyncratic pattern. In those varieties of Moroccan Arabic where feminine singular indexation with plural NPs is more widespread, it is always in free variation with DPI (Harrell 1965: 158).

Animacy-based restrictions on the use of DPI in Miya (Chadic) have been previously investigated by Schuh (1989), Corbett (1998), and Corbett (2000). In Miya, overt plural marking on nouns is obligatory only with animate nouns and, in such cases, it always triggers plural indexation. This is shown in (5.15).

(5.15) Plural indexation with animate nouns in Miya (Chadic) (Corbett 2000: 72-73, quoted from Schuh (1989))

nýkin *dzáfə*
this.PL man.PL
'these men'

With inanimate nouns, number marking on nouns is optional and plural indexation never occurs. Even when plurality is overtly coded, inanimate nouns trigger the same indexation as their correspondent singular (i.e., either masculine or feminine). This is shown in (5.16).

(5.16) Indexation with plural-marked inanimate nouns in Miya (Chadic) (Corbett 2000: 72-73, quoted from Schuh (1989))

nákən *víyaúwawáw*
this.SG.M fireplace.PL (*víyaúwawáw* 'fireplace' is masculine)
'these fireplaces'

Thus in Miya, the overt coding of plurality on inanimate nouns is optional, and its syntactic manifestation via plural indexation is ungrammatical. Interestingly, the two inanimate nouns for 'thing,' *kutə* and *ham*, and the inanimate interrogative pronoun *maa*, 'what?' always trigger plural indexation in Miya (Schuh 1989: 179). They are inanimate but control plural indexation because they are lexically associated with plurality.

Lexical plurality is the factor that controls the distribution of DPI with plural nouns in some languages of the Cushitic group. This is discussed in detail in the next section.

5.5.2 Dedicated Plural Indexation in Cushitic

5.5.2.1 Introduction

In this section, I examine the uses of DPI in the Cushitic languages of the sample. Before examining in detail the indexation patterns associated with plural NPs in Cushitic languages, the main characteristics of their gender and number systems are outlined.

There are 13 Cushitic languages in my sample. They all have grammatical gender and distinguish between masculine and feminine gender. Gender assignment is largely arbitrary except for nouns denoting humans and animate entities: in general, nouns denoting male entities are assigned to the masculine gender and nouns denoting female entities to the feminine gender. As for nominal number, all Cushitic languages distinguish between singular and plural. In addition, the number system of Cushitic languages can be classified according to two main parameters:

(1) Obligatoriness

(2) Indexation patterns associated with plural nouns.

In many Cushitic languages, nominal number marking is not obligatory (see also §4.3.1). Nouns with general number, inherently singular nouns or nouns that are overtly marked as singular trigger indexation according to their gender, that is, either masculine or feminine. On the other hand, plural nouns can trigger different indexation patterns in different languages. In some languages, DPI is used with all plural nouns, whereas other languages have split plural indexation systems, and DPI is used with only a subset of plural nouns. In order to investigate the uses of DPI in the Cushitic languages of my sample, I considered only those indexing targets in which plural indexation is in opposition with masculine and feminine indexation. In all the languages of the Cushitic sample, this opposition is found at the very least on the verbs and/or the third person plural pronouns.[39]

Finally, in Cushitic languages, as in many other languages of the world, nouns can be plural either morphologically or lexically. Morphological plurals are nouns for which plurality is overtly coded by means of plural morphemes. Lexical plurals are nouns that are not marked for number but are associated with plurality by virtue of their lexical semantics (Acquaviva 2008). The following types of nouns are particularly likely to exhibit lexically plurality: abstract nouns, nouns denoting activities involving multiple actions or participants, collectives, body parts nouns, nouns denoting diseases, expressions of location and place names, masses, nouns denoting internally complex objects, and expressions of time and time intervals. Lexical plurals are both animate and inanimate and thus feature discontinuous areas of the Animacy Hierarchy.

The number of lexical plurals in the Cushitic languages of my sample varies. For instance, according to my sources, there are at least 70 lexical plurals in Iraqw, 19 in Borana-Arsi-Guji Oromo, 12 in Rendille and 11 in Baiso. These are listed in appendix D.

[39] In many Cushitic languages, adjectives are reduplicated when they modify plural nouns (Mous 2008). This pattern is not taken into account in my study because its distribution does not compete with that of masculine and feminine indexation.

The noun lists provided in the appendix are organised according to semantic subfields. The lists have been elaborated on the basis of previous literature on lexical plurals in other language families of the world (see Kibrik 1992 on lexical plurals in Daghestanian languages and Koptjevskaja-Tamm & Wälchli 2001 on lexical plurals in the languages of the Circum-Baltic area).

In some of the Cushitic languages of my sample, lexical plurals and morphological plurals can trigger different indexation patterns. This is discussed in detail in the following subsections.

5.5.2.2 The distribution of DPI in the Cushitic sample

The marking of nominal plurality is obligatory only in two languages of my Cushitic sample: Awngi and Beja. I refer to these languages as *Group 1* languages. These languages, which will not be examined in detail here, have generalised uses of DPI: all plural head nouns – together with the third person plural pronouns – trigger an indexation pattern that differs from both masculine and feminine indexation.

In the remaining 11 languages of my Cushitic sample, the marking of nominal number is not obligatory. Two of these eleven languages, Kambaata and Dasanaach, have rather idiosyncratic uses of DPI and are thus discussed separately in §5.5.2.3. The remaining nine languages differ in the distribution of the DPI in more systematic ways, which are discussed in the rest of this section.

In Dahalo, Dirasha, Qimant, Somali and Tsamai, DPI is used with third person plural pronouns, morphological plurals and lexical plurals. I refer to these languages as *Group 2* languages. Plural indexation with lexical plurals and morphological plurals in Tsamai is illustrated by (5.17).

(5.17) Indexation with lexical plurals and derived plurals in Tsamai (Cushitic)

(a) Lexical plurals (Savà 2005: 51)

gore *žiʔ-e*
people eat-3PL.UNM
'The people ate'

(b) Morphological plurals (Savà 2007: 10)

ħezg-aɗɗe *li-ɠ-e*
star-PL go.out-3PL.UNM
'The stars went out'

Languages of this type represent the dominant pattern in my Cushitic sample.

Baiso, Iraqw and Rendille have split plural indexation systems. They differ from each other with respect to the cut-off points of these splits. In Baiso, Iraqw and Rendille, morphological plural nouns trigger the same indexation pattern as either masculine (Baiso) or feminine nouns (Iraqw, and Rendille). I refer to these languages as *Group 3* languages. In Iraqw, DPI is reserved to the Third Person Plural Pronouns, the lexical plurals and some morphologically plural nouns (Mous 1993). In Rendille, DPI has a rather broad

range of uses: it is found with Third Person Plural Pronouns, lexical plurals and most of the nouns that are morphologically marked as plural. The only exception is the plural suffix *-Ce/-nye*, which triggers feminine singular indexation (Oomen 1981: 60). A more elaborated system is found in Baiso.

In Baiso, as in all other Cushitic languages with non-obligatory number, general number nouns and morphological singular nouns trigger indexation according to their gender. The morphological plural nouns trigger the same indexation pattern as masculine nouns (Corbett 2000; Corbett & Hayward 1987). This is illustrated in examples (5.18) and (5.19), where the indexation patterns associated with a masculine and a feminine noun under general, singular and plural reference are illustrated.

(5.18) Number indexation with masculine nouns in Baiso (Cushitic) (adapted from Corbett 2000: 181)

(a) *lúban* *hudure*
lion.M.GENERAL slept.M
'The lion(s) slept'

(b) *lubán-titi* *hudure*
lion.M-SG slept.M
'The lion slept'

(c) *luban-jool* *hudure*
lion-PL slept.M
'The lions slept'

(5.19) Number indexation with feminine nouns in Baiso (Cushitic) (adapted from Corbett 2000: 182)

(a) *kimbír* *hudurte*
bird.F.GENERAL slept.F
'The bird(s) slept'

(b) *kimbír-titi* *hudurte*
bird.F-SG slept.F
'The bird slept'

(c) *kimbir-jool* *hudure*
bird-PL slept.M
'The birds slept'

As shown in the two examples, whenever a noun is marked by the plural suffix *-jool*, it triggers the same indexation pattern as masculine singular nouns.

Baiso is the only Cushitic language of my sample to have Paucal as a number value.[40] The paucal suffix *-jaa* is used to specify reference to a small group of entities. As the

[40] A paucal marker is also found in the Ta-Ne-Omotic language Koorete, which is spoken in the same area where Baiso is spoken. It has been suggested that the morphological encoding of paucality in Koorete – but not the marker as such – is the result of borrowing from Baiso (Teketal 2004).

Singular and the Plural, it is not an obligatory number value. It has been suggested (Corbett & Hayward 1987: 17-18) that, historically, the paucal suffix in Baiso resulted from the reanalysis of an originally plural suffix. Interestingly, the paucal suffix in Baiso always triggers the same indexation pattern as the Third Person Plural Pronouns, that is, DPI. This is illustrated in (5.20), where the same nouns as in (5.18) and (5.19) are used.

(5.20) DPI with paucal nouns in Baiso (Cushitic) (adapted from Corbett 2000: 181-182)

(a) *luban-jaa* *hudureene*
lion-PAUC slept.PL
'A few lions slept'

(b) *kimbir-jaa* *hudureene*
bird-PAUC slept.PL
'A few birds slept'

Corbett (2000: 216-217) refers to the indexation patterns associated with paucal-marked nouns in Baiso as an instance of semantic indexation (or "semantic agreement" in Corbett's own terms): smaller groups are more individuated than larger and thus more likely to trigger semantically justified indexation (that is, DPI).

DPI in Baiso is also used with a small subset of seven morphological plural nouns. These are listed in table 5.10.

Table 5.10: Morphologically plural nouns taking plural indexation in Baiso (adapted from Corbett & Hayward 1987: 13)

Number-unmarked form	**Gender**	**Plural-marked form**	**Meaning**
baal	M	*baalallo*	'feather, leaf'
suul	M	*suulallo*	'nail, claw'
fer	M	*fererroo*	'finger, toe'
gilib	M	*gilboo*	'knee'
nébe	F	*nebebboo*	'ear'
aayo	F	*aayoos*	'mother'
ilkoo	–	*ilkool*	'tooth/teeth' [41]

Corbett & Hayward observe that the nouns in table 5.10 denote entities that are likely to occur in pairs or in small groups: "one talks of *two knees* and *two ears* more frequently than of larger numbers, and similarly *nails*, *fingers* and *mothers* are found in smaller rather than larger groups" (1987: 13). Accordingly, I propose to classify the indexation patterns associated with these nouns also as an instance of semantic indexation or, to use the terminology introduced by Dahl (2000a), referential indexation.

Finally, in Baiso DPI is used with the lexical plural nouns listed in table 5.11.

[41] The noun for 'tooth/teeth' is the only one that triggers plural indexation both when number-unmarked and when marked as plural. See also table 5.11 and Corbett & Hayward (1987: 13).

Table 5.11: Lexical plurals in Baiso (Corbett & Hayward 1987: 9)

Semantic groupings	**Nouns**
Animate collectives	*saé* 'cattle'
Body parts	*ilkoo* 'tooth, teeth'; *kalaljaa* 'kidneys'; *lukkaa* 'foot, feet, leg(s)'; *iloo* 'eye(s)'; *ogorroo* 'hair'; *moo* 'hips, lumber region'
Objects coming in pairs	*keferoo* 'sandals'
Mass nouns	*eenoo* 'milk'; *soo* 'meat'; *udú* 'faeces'

As mentioned before, similar to Baiso, Iraqw and Rendille, Borana-Arsi-Guji Oromo also has split plural indexation. In this language, however, DPI is only found with Third Person Plural Pronouns and the lexical plurals. Morphological plurals always trigger the same indexation as feminine singular nouns. I refer to this language as *Group 4*. Examples (5.21) and (5.22) illustrate indexation with morphological and lexical plurals in Borana-Arsi-Guji Oromo.

(5.21) Indexation with morphological plurals in Borana-Arsi-Guji Oromo (Cushitic) (adapted from Stroomer 1995: 51)

shift-aa-ni *taani* *k'awwee* *d'aan-te*
bandit-PL-SBJ those guns shoot-3F.PST
'Those bandits shot guns'

(5.22) Indexation with lexical plurals in Borana-Arsi-Guji Oromo (Cushitic) (adapted from Stroomer 1995: 51)

bisaani *ibidda* *d'aamf-uu* *didani*
water fire estomguish-NOM refuse-3PL.PRS
'The water refuses to extinguish the fire'

It is worth mentioning that, even though they have the same endings, the two nouns – *shiftaani* and *bisaani* – are analysed in different ways in my source. *Shiftaani* is described as a morphological plural, marked by the plural suffix *-aa* and the subject marker *-ni*, whereas *bisaani* is described as a lexical plural noun, morphologically nonsegmentable.

To summarise, the Cushitic languages of the sample differ in their use of DPI. Most languages have generalised uses of DPI. In other languages, the use of DPI is constrained by the type of linguistic entity that triggers indexation. Figure 5.4 represents the type of plural indexation in each of the languages analysed in this section. The figure is based on the group subdivisions outlined throughout the section:

Group 1: Awngi, Beja

Group 2: Dahalo, Dirasha, Qimant, Somali and Tsamai

Group 3: Baiso, Iraqw, Rendille

Group 4: Borana-Arsi-Guji Oromo

Languages of groups 1 and 2, which differ from each other in the obligatoriness of number marking but share the same indexation patterns for plural nouns, are represented together in the figure. As shown by the figure, in the Cushitic languages with split plural indexation, the distribution of DPI is constrained by the lexical plurality of nouns.

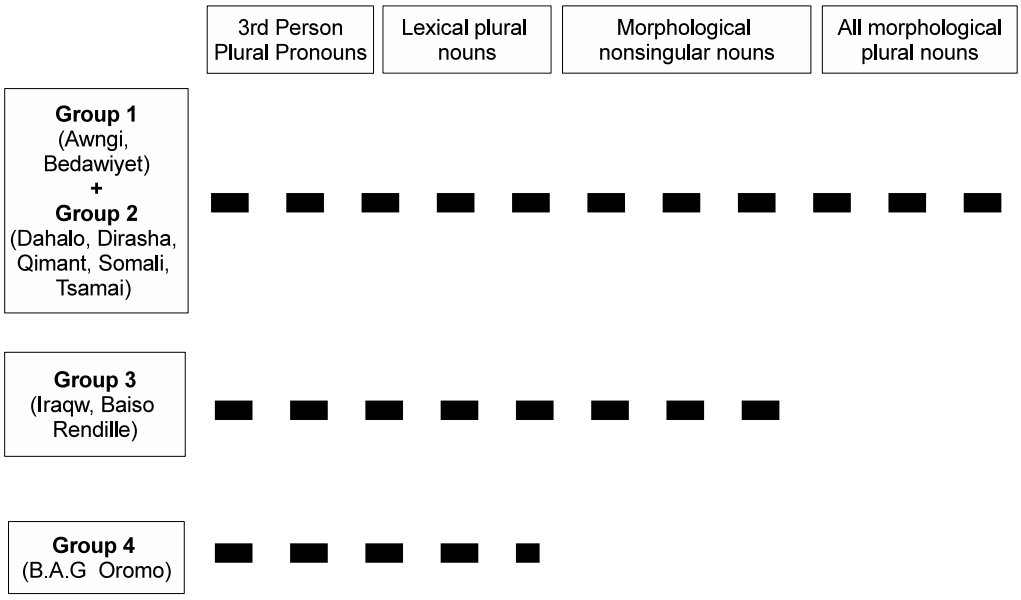


Figure 5.4: Triggers of DPI in the Cushitic sample

The geographic distribution of the four groups represented in figure 5.4 is shown in the map in figure 5.5.

It is noteworthy that Beja, one of the two languages with obligatory number, is spoken at the very edge of the Cushitic-speaking area, as represented in my sample. The peculiar status of Beja within the Cushitic language family is briefly discussed by Mous (2012: 345), who describes the language as "quite deviant compared to the rest of the Cushitic languages." Awngi, on the other hand, together with the other languages of the Agaw subgroup, is still understudied and little is found in the literature with respect to its status within the language family.

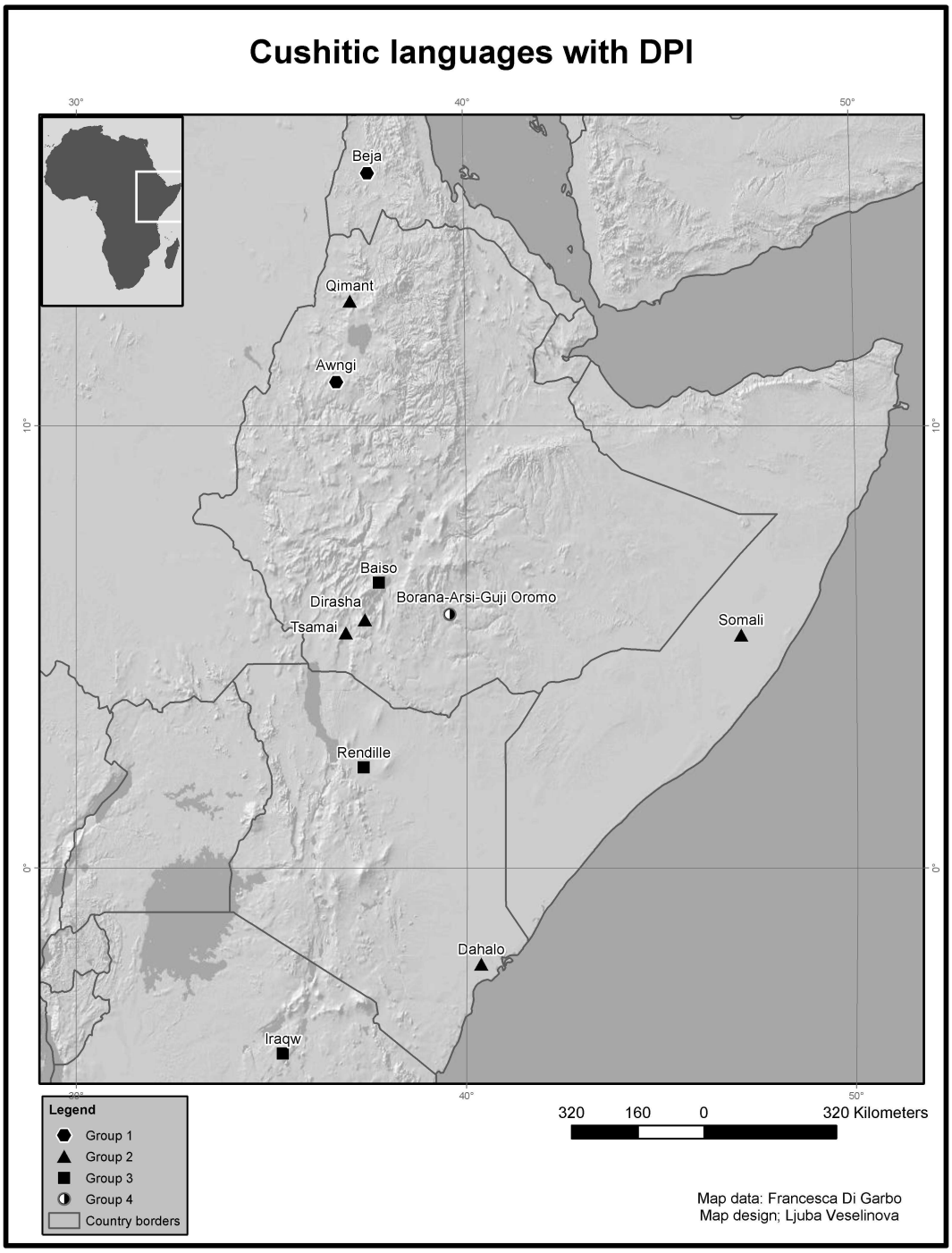


Figure 5.5: DPI in the Cushitic sample

5.5.2.3 Two peculiar languages within the Cushitic sample: Kambaata and Daasanach

The use of DPI in Kambaata and Daasanach is very marginal.

In Kambaata, the (Independent) and (Dependent) Personal Pronouns as well as the Anaphoric Demonstratives are the only indexing targets that inflect for number. Otherwise, plural indexation triggers in Kambaata (lexical plurals, morphological plural nouns and the Third Person Plural Pronouns themselves) always trigger the same indexation patterns as feminine singular nouns on all targets. Verbal indexation with plural nouns is shown in (5.23).

(5.23) Feminine indexation with plural nouns in Kambaata (Cushitic) (Treis 2014: 131)

alí	*wud-í*	*yabúr-r-assa*		*aaqq-án-t-iyan*
upper.M.GEN	side-M.GEN	lip-PL-F.NOM-3.PL.POSS		take-PASS-3.F.PCO-DS
hor-i-ssá	*inq-áak-ant*		*hínn*	*y-itoo'u*
all-M.GEN-3.PL.POSS	teeth-PL-F.NOM<n>		smirk	say-3.F.PFV

'the upper lips were grabbed (lit. 'taken') and the teeth of all of them smirked'

The example also shows two occurrences of the Dependent Third Person Plural Pronoun *-ssa*. In both occurrences – *yabúrrassa* and *horissá* – the suffix is interpreted as a possessive pronoun. In Kambaata, when the Dependent Personal Pronouns are attached to nouns, they always function as possessives (Treis 2008: p.348).

Personal Pronouns (Dependent and Independent) as well as the Anaphoric Demonstratives are the only targets of plural indexation in Kambaata. When the antecedent is a lexical plural or a morphological plural, the plural form of these pronouns is selected. The Independent Personal Pronouns are only used when the antecedent is human whereas the Anaphoric Demonstratives are used with all other types of antecedent (Treis 2014).

The use of DPI in Dasaanach is even more marginal than it is in Kambaata. There are no grammaticalized Third Person Pronouns, and plural nouns in Dasaanach always trigger the same indexation as masculine singular nouns (Tosco 2001). The generic nouns *mí/mú* 'man,' *máa* 'man, person,' *minni* 'woman' and *gáal* 'people' are used as sort of anaphoric words with human antecedents (Tosco 2001: 212). Interestingly, it is only with these anaphoric words that an incipient pattern of plural indexation is found, since *gáal* 'people' will be used when the human antecedent is plural. It is not clear what type of anaphoric words are used with nonhuman antecedents. Tosco (2001: 227) only mentions that the generic noun *ʔée* 'thing' can be used when an NP lacks an overt nominal head.

5.5.3 Discussion

In this section, I considered the issue of whether or not gender and number compete through indexation (see question 5 in (5.1)) and focussed on a small subset of the languages of the sample with split plural indexation. In these languages, the use of DPI is constrained by the lexical semantic properties of the indexation triggers and is in

complementary distribution with masculine and feminine indexation (these languages all have bipartite sex-based gender).

In §5.5.1, I considered the case of Standard Arabic and Miya, in which the use of DPI is constrained by the animacy of the NP referent. These types of semantic restrictions on the use of DPI go hand in hand with the well-documented role that the Animacy Hierarchy plays on the overt coding of nominal plurality on nouns across the languages of the world (see Corbett 2000; Haspelmath 2013; Smith-Stark 1974, as well as §2.2).

In §5.5.2, I investigated the uses of DPI in Cushitic. In Cushitic languages with split plural indexation systems, the Animacy Hierarchy does not seem to play a major role in predicting the types of indexation triggers that are more likely to select DPI. In these languages (with the exception of Kambaata and Dasaanach), at the very least the third person plural pronouns trigger DPI. In addition to these pronouns, the most common triggers of DPI are the lexical plurals, that is, nouns that are morphologically unmarked for number but inherently associated with plurality in virtue of their lexical semantics. In languages with less restricted or non-restricted uses of DPI, the next available indexation triggers after the lexical plurals are nouns that are morphologically marked as nonsingular. If read from the bottom to the top, Figure 5.4 suggests that the number of triggers of DPI grows somewhat incrementally across the Cushitic languages of the sample. I have found no languages in which morphological plurals trigger DPI, but lexical plurals do not do so. Similarly, I found no language in which morphological singular nouns trigger DPI.

The Cushitic data analysed in this section suggest that the Animacy Hierarchy may not be the only tool to explain the distribution of optional and obligatory number marking on the indexing targets (see also Cristofaro 2012 on the limitations of the Animacy Hierarchy in understanding the grammaticalization of nominal number). More difficult is it to establish the diachronic significance of these data. The situation attested in some of the Cushitic languages of the sample could be interpreted as the result of the expansion of the triggers of DPI from the personal pronouns to the lexical plurals, and from there to the morphological plurals. On the other hand, the opposite path, i.e., a reduction of the domain of use of DPI, is also a plausible diachronic scenario. A more thorough historical-comparative investigation of the Cushitic languages would be needed in order to understand the diachronic development of this type of system. This lies outside the scope of the present investigation.

Finally, an interesting interplay between gender and number indexation is found in Bench (Ta-Ne-Omotic), where the distribution of DPI closely resembles the situation observed in some of the Cushitic languages of the sample. In Bench, plural-marked nouns normally trigger masculine indexation. DPI can be found only with "indicative final, polar question, imperative, and relative verbs" as well as with demonstratives (see Rapold 2006: 540), but it is always optional. Rapold (2006: 540-541) describes DPI in Bench as an instance of semantic indexation that competes with morphological indexation (masculine and feminine). The distribution of DPI in Bench is described as both semantically and pragmatically constrained, but not much is said in my source about the exact nature of these constraints.

5.6 Presence of gender and type of number marking

In this section, I discuss if and how the presence of gender correlates with any aspect of the nominal number marking system that is found in a language (see last research question in (5.1)). As explained at the beginning of the chapter, this research question was elaborated in order to verify whether two generalisations on morphosyntactic properties of African languages formulated by Creissels et al. (2008) can be applied to the languages of my sample. The two generalisations are as follows:

> (a) Languages devoid of a gender system frequently have a single plural marker with the morphological status of a phrasal affix, and such plural markers tend to be used on a "pragmatic basis," i.e., to be employed only when plurality is both communicatively relevant and not implied by the context, at least in the case of nouns that do not refer to persons [...].
>
> (b) Languages that have gender generally have a morphologically complex plural marking, characterized by a fusion of gender and number markers, and variation in gender and number manifest themselves through morphemes affixed to the head noun and to (some of) its modifiers, in an agreement relationship. In these languages [...] plural markers tend to be present in every NP referring to a plurality of individuals [...] (Creissels et al. 2008: 119).

The claims made in (b) describe a very common pattern in the languages of the sample that have obligatory number. In fact, the properties enumerated in (b) – cumulation (fusion) of gender and number markers, and marking of gender and number on nouns and through indexation (agreement) – have been at the bulk of this chapter.

The language type described in (a) is also attested in the languages of my sample but has not been considered in detail so far. Of the 16 genderless languages of my sample, 10 have strategies of nominal number marking that very closely reflect the generalisation expressed in (a). These languages are: Yoruba (Defoid), Igbo (Igboid), Bambara (Mande), Susu (Mande), Mann (Mande), Mandinka (Mande), Hdi (Chadic), Mina (Chadic), Akan (Kwa) and Ewe (Kwa). Let us consider Yoruba as an illustration of this type of system. The description of the number system of Yoruba is based on Ajíbóyè (2010) and Bamgbose (1966).

In Yoruba, there is no grammatical gender, the marking of nominal plurality is optional with all nouns. Number-unmarked nouns can be interpreted either as singular or plural. When speakers want to specify that the referent of a NP is plural, several strategies are available: the plural word *àwọn* (formally identical to the Third Person Plural Pronoun), the plural morpheme *wọ̀n* (which can only be used to pluralize demonstratives), quantifiers such as *púpọ̀*, 'many,' numerals, and, finally, when the noun occurs together with an adjectival modifier, adjectival reduplication. Thus, from a morphosyntactic point of view, the only indexes of the number values of the NP referent are the Third Person Pronouns and the Demonstratives. When plurality is overtly coded on

the demonstratives, the whole NP is obligatorily interpreted as having a plural referent. Plural marking with the plural word *àwọn* is illustrated in (5.24).

(5.24) Plural marking in Yoruba (Defoid): the plural word *àwọn* (Ajíbóyè 2010: 143)

Mo	*kí*	*àwọn*	*ọkúnrin*	*tí*	*ó*	*wà*	*níbẹ̀*
1SG	greet	PL	man	that	RPbe	there	

'I greeted the men that were there'

One property that seems to be shared by each of the ten languages mentioned above is the lack of pervasive systems of number indexation. Table 5.12 illustrates how many number-indexing targets are found in these languages.

Table 5.12: Number-indexing targets in Creissels et al.'s (2008) type (a) languages

Number of number indexes	**Number of languages**	**Genealogical groups**
1	6	Kwa (1/3); Igboid (1/1) Mande (4/4)
2	4	Kwa (1/3); Chadic (2/8); Defoid (1/1)

The figures in table 5.12 show that all ten languages classifiable as type (a) in the classification proposed by Creissels et al. (2008) have very minimal number indexation with only one or a maximum of two number indexes. In addition, if a language has one number index only, this is the third person pronouns.

The remaining six genderless languages in the sample are the Western Nilotic languages Acoli, Belanda Bor, Dinka, Luwo, Mabaan and Nuer. They all lack grammatical gender but are characterised by a rather elaborated system of number marking on nouns that also serves a classificatory function. This type of number system is known in the literature as the *tripartite number system*[42] because three types of nouns exist in these languages (see also the discussion in §4.3.2):

(1) Nouns where only singular reference is overtly coded

(2) Nouns where only plural reference is overtly coded

(3) Nouns where both singular and plural reference are overtly coded (also known as *replacement pattern*).

[42]Tripartite number systems with classificatory functions are also found in the three Eastern Nilotic languages of the sample – Turkana, Masai and Karamojong. In these languages, the tripartite number systems co-exist with cumulative encoding of gender and number. This was discussed in §5.2.1.4.2.

Since all nouns, both those that are overtly coded for number and those that are not, are associated with a number value, the tripartite number system of Western Nilotic is an obligatory number system. The three patterns described above are exemplified in examples (5.25), (5.26) and (5.27). These examples have already been discussed in §4.3.2, where a general description of tripartite number systems was outlined.

(5.25) Singular marking in Mabaan (Western Nilotic), same example as (4.8) (Storch 2005: 123)

(a) *ɲɔ́ɔ̀k*
'lice'

(b) *ɲɔ́ɔ́k-cn̲*
lice-SG
'louse'

(5.26) Plural marking in Mabaan (Western Nilotic), same example as (4.9) (Storch 2005: 123)

(a) *jwɔ́m*
'monkey'

(b) *jwɔ́m-gʌ*
monkey-PL
'monkeys'

(5.27) Replacement pattern in Mabaan (Western Nilotic), same example as (4.10) (Storch 2005: 123)

(a) *ʔúʌt-*
worm-SG
'worm'

(b) *ʔúʌt-n̲*
worm-PL
'worms'

Dimmendaal defines the tripartite number systems spread throughout Nilotic and, more broadly, Nilo-Saharan languages as "a classificatory technique" (2000: 214), whereby nouns with analogous morphophonological or semantic properties result in similar number morphology. The same interpretation is proposed by Storch (2005, 2007) in her investigation of the structure, properties and diachrony of NPs in Western Nilotic. She defines the morphological strategies that Western Nilotic languages use to categorise nouns in semantically homogeneous groups as *nominal classifiers* following Aikhenvald's (2003) sense of the term. Beside signalling number distinctions, these markers express the following meanings: biological gender, animacy, tactile perception, shape and size. Even though the tripartite number system is preserved in all Western Nilotic languages, the classifier-like morphemes through which it functions are not equally spread among the languages of the group. Storch (2005, 2007) suggests that the systems attested in all contemporary Western Nilotic languages are a reduced version of what the original protosystem must have looked like. This reduction is explained by Storch as the result of a general propensity towards the simplification of the number marking system, and towards the generalisation of a few singular and plural markers to all nouns, independently of their lexical semantics. These generalised number markers are called by Storch

imperialistic number markers (2005; 2007). The inventory of the singular and plural markers in three of the Western Nilotic languages included in the sample – Mabaan, Dinka and Luwo – is given in appendix E.

Interestingly, similarly to the ten languages examined above, in the Western Nilotic languages of the sample number indexation is also minimal or null (Storch 2005).

To summarise, the languages of my sample generally reflect the generalisations proposed by Creissels et al. (2008) with respect to gender and number in African languages. However, what my data suggest is that the grammatical phenomenon that tends to be affected the most by the co-presence of gender and number is indexation. Languages in which gender and number are both attested tend to have very pervasive indexation systems. On the other hand, if only number is present, indexation patterns tend to be very poor, and number distinctions are mostly marked at the phrasal level (i.e., only one marker per NP is found). This pattern is illustrated in figure 5.6 (see also appendix C for a more detailed overview).

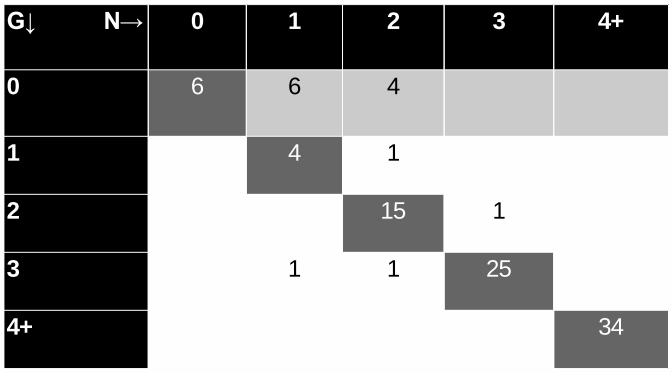


Figure 5.6: Gender- and number-indexing targets in the languages of the sample

Two relevant patterns can be detected in figure 5.6: (1) the majority of the languages of the sample have pervasive indexation systems both for gender and number; (2) the majority of the languages of the sample have the same number of gender- and number-indexing targets. However, with respect to (2), it is worth mentioning that *same number of gender- and number-indexing targets* does not always mean *same type of indexing targets* or *same type of marking strategy*. For instance, in Cushitic languages number distinctions on adjectives tend to be signalled by means of reduplication patterns, whereas gender distinctions tend to be signalled only by means of affixation. A systematic of overview of such differences in the distribution of type of indexing targets and types of marking strategies per domain constitutes a promising area of research; however, this falls outside the scope of this investigation.

To conclude, in light of the data examined in this section and in the rest of the chapter, I suggest reformulating the generalisations by Creissels et al. (2008) in the following way:

(a_1) At least in African languages, the development of rich and pervasive nominal indexation systems seems always to involve both gender and number. These systems exhibit a strong tendency for cumulative exponence and related interactions (i.e., syncretism).

(b_1) When gender marking is absent, number marking tends to remain optional and follow the principle "one marker per NP."

5.7 Gender shifts and semantic interactions between gender and number

In §2.1.1, I discussed the phenomenon whereby in certain languages, nouns can be assigned to more than one gender depending on the way the NP referent is construed, and I briefly examined previous studies on the topic. I proposed to use the label *manipulable gender assignment* to refer to the possibility of assigning nouns to multiple genders and the label *rigid gender assignment* to refer to those languages in which nouns are invariably assigned to one gender. As also observed in §2.1.1, the process whereby nouns can be assigned to genders others than their default is referred to in this dissertation as *gender shift*.

In this section, I discuss a set of data from my language sample that suggests that, in some languages, gender assignment can be manipulated in order to indicate variation in the countability properties of nouns. The distribution of this phenomenon in the languages of the sample is shown in table 5.13.

Table 5.13 shows that gender shifts for the encoding of variation in the countability properties of nouns are attested in 41.6% of the gendered languages in the sample. The distribution of the phenomenon is, however, highly skewed from a genealogical point of view. Gender shifts for the encoding of variation in the countability properties of nouns are attested only among the following genealogical groups: Bantu, Berber, Eastern Nilotic, North-Central Atlantic and Semitic. The majority of the languages displaying this phenomenon are Bantu, followed by Berber, North-Central Atlantic, Semitic and Eastern Nilotic, where the phenomenon is attested in only one language, Turkana.

The rest of this section is organised as follows. In §5.7.1, I discuss data from the Bantu and North-Central Atlantic languages because the two groups have a rather similar gender system and are also genealogically related. In §5.7.2, I consider the occurrence of this phenomenon in Berber, Semitic and in the Eastern Nilotic language Turkana. Similarly to Bantu and North-Central Atlantic languages, these languages also share the same type of gender system (i.e., sex-based gender).

Table 5.13: Distribution of the use of gender shifts for the encoding of variation in the countability properties of nouns

Use	No. of lngs.	Rel.%	Abs.%	Genealogical groups
Present	35	41.6%	35%	Bantu (21/23) Berber (6/6) Eastern Nilotic (1/3) North-Central Atlantic (4/7) Semitic (3/7)
Absent	45	53.6%	45%	Bantu (2/23) Cushitic (13/13) Chadic (6/8) Dizoid (1/1) Eastern Nilotic (2/3) Hadza (1/1) Khoe-Kwadi (5/5) Kwa (1/3) Kxa (1/1) Mel (1/3) North-Central Atlantic (2/7) Sandawe (1/1) Semitic (4/7) South Omotic (1/1) Ta-Ne-Omotic (4/4)
No data	4	4.8%	4%	North-Central Atlantic (1/7) Mel (2/3) Tuu (1/1)
No gender	16	–	16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total	100	100%	100%	

5.7.1 Gender shifts, number/countability in Bantu and North-Central Atlantic

The two obligatory number values in most North-Central Atlantic and Bantu languages are singular and plural. The individual noun classes of these languages are associated either with singular or with plural reference. As observed in 4.2.3, in the most regular cases, the singular classes have systematic correspondents among the plural classes. Examples (5.28) and (5.29) illustrate gender and number marking in Swahili (Bantu) and Bandial (North-Central Atlantic).

(5.28) Swahili (Bantu)(Thompson & Schleicher 2001: 39)

(a) *m-tu*
CL1-person
'person'

(b) *wa-tu*
CL2-person
'people'

(5.29) Bandial (North-Central Atlantic) (Sagna 2008: 275)

(a) *a-rokka*
CL1-worker
'worker'

(b) *u-rokka*
CL2-worker
'workers'

Nouns with low countability properties in Bantu and North-Central Atlantic languages are always assigned to a noun class and often do not undergo number distinctions. Moreover, since noun class values are always cumulated with number values, uncountable nouns are either lexically singular or lexically plural. In Swahili, nouns denoting abstract notions are assigned to Class 11 and lack a plural counterpart. An example of a Class 11 abstract noun is given in (5.30).

(5.30) Swahili (Bantu)(Contini-Morava 2000: 6)

u-singizi
CL11-sleep
'sleep, sleepiness'

In Bandial, nouns denoting liquids and masses are typically assigned to the plural class *mu-*, Class 10, and do not have a singular counterpart. An example of a Class 10 mass noun is given in (5.31).

(5.31) Bandial (North-Central Atlantic) (Sagna 2008: 272)

mu-u
CL10-salted.water.from.the.river
'salted water from the river'

If the gender systems of the Bantu and North-Central Atlantic languages are interpreted as rigidly grounded in the notion of gender as singular-plural pairing, lexical singular and lexical plural nouns should be represented as displaying defective paradigms. As shown later in this section, other irregularities can be observed in these systems.

Tables 5.14 and 5.15 illustrate the regular singular-plural pairings of Swahili and Bandial. Nonpairing noun classes – as the locative classes[43] – are excluded from the table.

Table 5.14: Noun class and singular-plural pairs of Swahili (adapted from Contini-Morava 2000)

Class	**Singular**	**Class**	**Plural**
1	*m-*	2	*wa-*
3	*m-*	4	*mi-*
5	*ji-*	6	*ma-*
7	*ki-*	8	*vi-*
9	*n-*	10	*n-*
11	*u-*	6	*ma-*
11	*u-*	10	*n-*

Table 5.15: Noun class and singular-plural pairs of Bandial (Sagna 2012: 136)

Class	**Singular**	**Class**	**Plural**
1	*a-*	2	*bug-*
3	*e-*	4	*su-*
5	*bu-*	6	*u-*
7	*fu-*	8	*gu-*
9	*ga-*	10	*mu-*
11	*ju-*	6	*u-*
12	*ñu-*	6	*u-*

In addition to the usages discussed so far, in the Bantu and North-Central Atlantic languages of my sample, noun class marking can be manipulated to encode non-obligatory

[43] Many Bantu and North-Central Atlantic languages have noun classes that are used to encode location (locative classes are also found in the languages of the Mel group). The locative noun classes have rather peculiar properties as compared to others. For instance, they have no basic members if not the word for 'place' and tend to trigger noncanonical patterns of subject marking, also known in the literature as *locative inversions* (for an overview of this matter, see Maho 1999). In a sense, they can be seen as rather nonprototypical members of the gender systems of the North-Central Atlantic, Mel and Bantu languages (see, for example, Marten 2010: 252, on locative classes in Bantu). The locative classes are generally insensitive to number distinctions.

number values such as *collectives*, *distributive plurals* or *singulatives*. In Bandial, for instance, several class markers can be used to form collectives from nouns that otherwise have regular singular-plural noun class pairings. One of these markers is *ba-*, a singular class that is defined as a subclass of Class 5 and is described as semantically associated with assemblages (Sagna 2008, 2011). When nouns with other regular class pairs are assigned to Class 5, the resulting meaning is what Sagna (2011) defines as *diminutive collective*, that is, a noun that denotes "a group of small entities as a unit" (Sagna 2011: 243). This is shown in example (5.32).

(5.32) Bandial (North-Central Atlantic) (Sagna 2011: 243)

(a) *e-vval*
CL3-stone
'stone'

(b) *si-vval*
CL4-stone
'stones'

(c) *ba-vval*
CL5-stone
'pile of stones'

Class 4 is the regular plural of the noun for 'stone' (5.32b). When assigned to Class 5 (5.32c), the noun denotes a collection of stones.

Similarly, in Swahili the noun class prefix *ma-*, Class 6, has several usages: (1) it is the regular plural correspondent of Class 5; (2) it is the default noun class for lexical plurals denoting masses and liquids; and (3) it is used to form the collective of nouns that form their regular plural in other classes. The latter use is illustrated in example (5.33).

(5.33) Swahili (Bantu) (Shadeberg 2003: 84)

(a) *simba*
CL9.lion
'lion'

(b) *ma-simba*
CL6-lion
'a pride of lions'

In Swahili, nouns assigned to Class 9 when singular are usually assigned to Class 10 when plural. When assigned to Class 6, as in (5.33b), they are interpreted as collectives.

The phenomena illustrated in examples (5.32) and (5.33) show that in some Bantu and North-Central Atlantic languages class assignment can be manipulated in order to modify the countability properties of nouns and encode non-obligatory number values. These phenomena thus challenge the idea that the gender system of these languages can

be understood solely on the basis of the singular-plural pairings. In the North-Central Atlantic and Bantu languages where these phenomena are attested, the possibility of manipulating the countability properties of nouns is a fundamental property of the gender system (see also Acquaviva 2008: 41-42).

5.7.2 Gender shifts, number/countability in Berber, Semitic and Eastern Nilotic

The feminine marker of the Berber languages can be used to encode variation in the countability properties of nouns. In particular, uncountable nouns are marked by the feminine marker (and trigger feminine indexation) when a singulative interpretation is intended for the noun. This is illustrated in example (5.34), from the Berber language Nefusi.

(5.34) Nefusi (Berber) (Beguinot 1942)

(a) *ettefâḥ*
'apples' (collective)

(b) *t-attefâḥ-t*
F-apples-F
'one apple'

The circumfixing of the feminine marker to the collective noun *ettefâḥ* results in the formation of a noun that is semantically singular and grammatically triggers feminine indexation. This noun is countable, and, as such, it can undergo regular plural marking, as illustrated in example (5.35).

(5.35) Nefusi (Berber) (Beguinot 1942)

t-attefâḥ-în
F-apples-F.PL
'apples' (plural)

The singulative function of the feminine marker *t* is attested in all the Berber varieties of my language sample. As shall be illustrated in §6.4, in the Berber languages of my sample, gender assignment can also be manipulated to encode diminutive and augmentative meanings. Inherently masculine nouns are shifted to the Feminine Gender when a diminutive interpretation is intended for the noun (see example (6.22)). Similarly, inherently feminine nouns are shifted to the Masculine Gender when an augmentative interpretation is intended for the noun (see example (6.23)). Specialists of Berber languages point out that the interpretation of the feminine marker as either feminine proper, singulative or diminutive depends on the interplay between the lexical semantics of the noun, the discourse context and the speaker's attitude towards the NP referent (see, among others, Penchoen 1973b and Mettouchi 2000 with respect to Tachawit and

Kabyle, respectively). Mettouchi (2000) also suggests that the diminutive and singulative function of the *t* marker in the Berber languages was diachronically prior to the feminine proper function. As pointed out already in §4.2.2, Mettouchi's idea is that the original function of the marker was to identify an entity with respect to a reference point, the relationship between the two (the entity and the reference point) being inherently hierarchical and roughly of the part-whole type. According to Mettouchi, only at a later stage, and starting with animate entities, was this polarity reanalysed as a "masculine vs. feminine" type of opposition. The diminutive and augmentative uses of the gender markers in the Berber languages are discussed in detail in §6.4.1 (on the mutually exclusive relationship between the different meanings of gender shifts – biological gender, size and countability –, see also Kossman 2014).

Similarly to the Berber languages, in Classical and Standard Arabic, Moroccan Arabic and Maltese, the feminine marker functions as a singulative when used with uncountable nouns, both masses and collectives. Example (5.36) illustrates how in Classical Arabic, the interpretation of the suffix *-AT* varies based on the countability properties of the base noun.

(5.36) The suffix *-AT* and the countability properties of nouns in Classical Arabic (Semitic) (adapted from Hämeen-Anttila 2000: 602)

(a)	*kalb-* 'dog (male)'	*kalb- + AT* 'bitch, female dog'
(b)	*ḥamām-* 'pigeons'	*ḥamām- + AT* 'one pigeon'

In (5.36a), the base noun is animate and countable. It is interpreted as denoting both female and male dogs, and syntactically behaves as masculine, in the sense that it triggers masculine indexation. When this base noun is marked by *-AT*, the resulting meaning can only be feminine and the noun triggers feminine indexation. On the other hand, in (5.36b), the base noun is uncountable and syntactically behaves as a masculine noun. When this base noun is marked by *-AT*, the result is a singulative, countable noun that triggers feminine indexation and can be regularly pluralized. Interestingly, masculine and feminine gender with the noun in (5.36b) can be encoded only periphrastically, that is, by using the modifiers for 'male' and 'female' together with the noun for 'pigeons/pigeon.' This is shown in example (5.37).

(5.37) Biological gender with collectives and singulatives in Classical Arabic (Semitic) (adapted from Hämeen-Anttila 2000: 602)

(a) *dhakar- al-ḥamām-* 'the male of pigeons'

(b) *al-ḥamā+at adh-dhakar-* 'one male pigeon'

In this thesis, I follow the tradition of most descriptive grammars of Arabic dialects in which the collective, uncountable form is considered the basic form as opposed to the singulative-marked form. Elie Wardini (personal communication) suggests that at least in the spoken varieties of Arabic, the singulative-marked forms may actually be considered the basic forms. Following this suggestion, the noun *ḥamāmat* 'pigeon' should

be interpreted as an inherently feminine countable noun analogous to the nonsingulative noun *qiṭṭat* 'cat.' The uncountable noun *ḥamām-* 'pigeons' should rather be interpreted as the result of some sort of subtractive morphological rule, whereby presence vs. absence of overt gender marking is a means for manipulating the countability properties of nouns. Wardini's hypothesis could be verified on the basis of behavioural data from native speakers of different Arabic dialects. This research line lies, however, beyond the primary scope of the present investigation.

What is more important to underline here is that, as in the Berber languages, the above-mentioned Semitic languages productively use manipulation of gender assignment as a means for encoding variation in the countability properties of nouns. Uncountable nouns – both masses and collectives – tend to behave as masculine nouns syntactically. When marked by the feminine marker, these nouns behave syntactically as feminine and are semantically interpreted as countable and singular. While the semantic denotation of the noun stays unchanged, the construal of its countability properties changes radically when the indexation class changes.

Similarly to Mettouchi's (2000) proposal on the diachrony of gender in the Berber languages, it has been proposed that the gender system of the Arabic dialects was also originally non-sex-based. If this is the case, it is reasonable to hypothesise that the feminine marker originated from "a more complex system of classes within which the category of number has to be included as well [...]" (Moscati 1964: 86).

The feminine marker is also used to encode diminution in Moroccan Arabic and Maltese, similarly to the Berber languages, but not in Standard Arabic (for a more detailed description of evaluative morphology in the Semitic languages and its relationship with gender, see chapter 6).

In Turkana (Eastern Nilotic), feminine and masculine nouns can be shifted to the Neuter Gender to encode variation in the countability properties of nouns or in the size/age of the NP referent. The interpretation of shifts to the Neuter Gender largely depends on the countability properties of the base noun. If the base noun is countable, the meaning resulting from the gender shift is diminutive; if the base noun is uncountable, shifts to the Neuter Gender are used to encode singulative meanings. Examples are discussed in §6.4.1. See also §7.7 for a discussion of the grammatical complexity of gender in Turkana.

5.7.3 Gender shifts and semantic interactions between gender and number: summary

In this section, I examined the use of gender shifts as a strategy for encoding variation in the countability properties of nouns in languages with manipulable gender assignment. This was counted as an instance of semantic interaction between gender and nominal number. The phenomenon is attested in 34 of the 84 gendered languages is the sample but its distribution is highly genealogically skewed. The role that this type of interaction plays in the absolute complexity of gender systems is discussed in detail in chapter 7.

5.8 Summary of the chapter

The purpose of this chapter has been to provide a systematic overview of patterns of interaction between gender and number in the languages of the sample along the following dimensions: exponence, syncretism, competing indexation patterns, correlations in type of marking, and semantic interactions through gender assignment. The research questions addressed in the chapter are listed in (5.1).

Both cumulative exponence of gender and number were found to be very frequent in the languages of the sample. In addition, the occurrence of gender syncretism in the context of number was found to correlate with the presence of cumulative paradigms for the encoding of gender and number.

The presence of cumulative exponence of gender and number has been interpreted as an increase in the absolute complexity of gender and number systems. Furthermore, the results of the investigation, in line with what has been claimed in previous literature, have been interpreted as suggesting that, in a hypothetical relevance hierarchy for nouns, number would need to be ranked higher than gender.

The data from the language sample seem to disprove the claim by Baerman et al. (2005) according to which syncretism leads to the semanticization of gender distinctions across number values (particularly nonsingular number values). In the majority of the cases attested in the sample, syncretism of gender in the context of number does not yield any semantically transparent reorganisation of gender values. In the few cases in which semanticization does occur, it is always animacy-driven.

The question of how gender and number interact through patterns of indexation was tackled by considering those (rare) cases in which a dedicated pattern of plural indexation (DPI) exists, but its uses are restricted by the semantic properties of the indexation triggers. Two types of semantic restrictions were found: (1) restrictions based on the animacy of the indexation triggers, whereby only human or animate nouns trigger DPI, and (2) restrictions based on the inherent plurality of the indexation triggers, whereby lexical plurals are one of the few available triggers of DPI besides the third person plural pronouns. These results suggest that the Animacy Hierarchy might not be the only valuable predictor of the use of number indexation in a language. They also suggest that there might be an opportunity for revising existing models pertaining to the grammaticalization of nominal number. One could perhaps think of a split whereby animacy is the best predictor for the use of obligatory number marking on nouns (as suggested by Corbett 2000, Haspelmath 2013, and Smith-Stark 1974), whereas both animacy and lexical plurality play a role in constraining the distribution of number indexation. Further research is needed in order to verify the crosslinguistic validity of this hypothesis. A commonly acknowledged major functional divide between gender and number is that *gender* is a *lexical feature*, whereas *number* is *referential*. The Cushitic data discussed in this chapter suggest that competing indexation patterns between gender and number tend to be associated with nouns for which this functional divide is not as clear-cut, that is, with the lexical plurals. The crosslinguistic relevance of these phenomena should also be further investigated.

In §5.6, I investigated possible correlations between presence/absence of gender and

type of number marking by using the generalizations of Creissels et al. (2008) as a reference. I showed that, as a rule, the languages of my sample reflect these generalisations. In addition, I showed that, a relevant implication of the co-presence of gender and number is the existence of very pervasive indexation systems. In languages without gender, number indexation is null or, at best, very poor. In addition, the languages of the sample tend to exhibit the same number of gender and number indexing targets.

Finally, in §5.7, I discussed a particular type of semantic interaction between gender and number, that is, the use of gender shifts to encode variation in the countability properties of nouns. I showed that in languages where nouns can be assigned to multiple genders depending on the construal of the NP referent, gender shifts can be used to mark uncountable nouns as countable and singular.

6 Gender and evaluative morphology

6.1 Introduction

In this chapter, I discuss the interactions between gender and the morphological encoding of evaluation in the languages of my sample. Within the grammatical domain of evaluation, only diminutive and augmentative markers are considered. The following research questions are addressed in the chapter:

(6.1) Research questions concerning gender and evaluative morphology (same as (2.14))

Q 1: How frequently does *size* occur as an independent gender value? How stable and how widely distributed is this phenomenon within genenealogical units?

Q 2: Do the interactions between gender and evaluative morphology differ across types of gender systems and/or strategies of gender assignment (e.g., *sex-based* vs. *non-sex-based* gender systems, or *rigid* vs. *manipulable* gender assignment)?

Two kinds of interaction are found in my data: (1) interactions between grammatical gender (both sex-based and non-sex-based) and evaluative morphology, and (2) interactions between evaluative morphology and the encoding of biological gender (male vs. female) in languages with non-sex-based gender or without grammatical gender. A major part of the present chapter is devoted to the analysis and discussion of the first type of interaction. This is discussed in §§6.2, 6.3 and 6.4. The second type of interaction, which was observed in a smaller number of languages, is shortly discussed in §6.5. Absence of interaction between the two domains is discussed briefly in §6.6. A summary of the chapter is provided in §6.7.

6.2 Gender and evaluative morphology: overview of results

Two major kinds of interactions between grammatical gender and evaluative morphology are found in the languages of the sample:

(a) If a language has non-sex-based gender, which generally implies a high number of gender distinctions (see chapter 4), diminutives and augmentatives are part of the inventory of gender distinctions. Nouns are generally not assigned to the diminutive and augmentative genders by default but only when a diminutive or augmentative interpretation is intended. I refer to these languages as Type 1.

(b) If a language has sex-based gender, which generally implies two or three gender distinctions (see chapter 4), a change in the default gender of a noun (e.g., from

masculine to feminine, from feminine to masculine or from masculine/feminine to neuter) may express diminutive or augmentative meanings. I refer to these languages as Type 2.

The distribution of each type of interaction within the languages of my sample is summarised in table 6.1.

Table 6.1: Interactions between grammatical gender and evaluative morphology

Type	**No. of lngs.**	**Rel. %**	**Abs. %**	**Genealogical groups**
Type 1	23	27.4%	23%	Bantu (15/23) Kwa (1/3) North-Central Atlantic (7/7)
Type 2	28	33.3%	28%	Berber (6/6) Chadic (2/8) Cushitic (3/13) Dizoid (1/1) Eastern Nilotic (3/3) Hadza (1/1) Khoe-Kwadi (3/5) Semitic (3/7) Sandawe (1/1) South Omotic (1/1) Ta-Ne-Omotic (4/4)
No interaction	13	15.5%	13%	Bantu (6/23) Chadic (1/8) Cushitic (1/13) Kxa (1/1) Semitic (3/7) Tuu (1/1)
No data	20	23.8%	20%	Bantu (2/23) Chadic (3/8) Cushitic (9/13) Khoe-Kwadi(2/5) Mel (3/3) Semitic (1/7)
No gender	16		16%	Chadic (2/8) Defoid (1/1) Igboid (1/1) Kwa (2/3) Mande (4/4) Western Nilotic (6/6)
Total	100	100%	100%	

Table 6.1 shows that languages of Type 2 are slightly more frequent than languages of Type 1. The two types of interactions are discussed in two separate sections: Type 1 is discussed in §6.3, whereas Type 2 is the focus of §6.4.

As mentioned in §2.1.1, Corbett (2013a) proposes to refer to Type 2 phenomena as instances of *recategorization*. In addition, as observed in 4.2.3, both Type 1 and Type 2 phenomena are often described in the literature as *gender shifts*. Since the notion of *recategorization* and related terminology has not yet been established (Greville Corbett, personal communication), the notion of *gender shift* – or, depending on the system, *noun class shift* – is used in this dissertation to refer to both Type 1 and Type 2 phenomena. To date, very little has been done to explore the crosslinguistic distribution of size-related (and value-related) gender shifts (see discussion in §§2.1.1 and 2.6). To my knowledge, this dissertation provides the first extensive account of the crosslinguistic distribution of this phenomenon over a large sample of languages from one of the world's macro-areas.

6.3 Type 1: Diminutive and augmentative genders

As outlined in §6.2, in the languages of Type 1 (23 out of 100; see table 6.1), diminutives and augmentatives are part of the inventory of gender distinctions. In my analysis, I count as diminutive and augmentative markers internal to a gender system either the singular-plural class pairs or, if no number distinctions can be found, the individual classes that are used to express diminutive and augmentative meanings. Within my sample, this type is attested among the North-Central Atlantic and Bantu languages as well as, to a certain extent, in the Kwa language Sɛlɛɛ (table 6.1). Typically, the formation of diminutives and augmentatives is done by means of gender shift. Speakers shift nouns from their default gender to the diminutive or the augmentative genders. As a results, nouns trigger different indexation patterns (the diminutive and augmentative indexation patterns) and are themselves marked in a different way. Examples (6.2) and (6.3) illustrate the mechanisms of noun class shift for diminutive and augmentative formation in a Bantu and a North-Central Atlantic language.

(6.2) Tonga (Bantu) (Carter 2002: 21)

(a) *mu-sankwa*
CL1-boy
'boy'

b *tu-sankwa*
CL12-boy
'small boy'

(6.3) Wamey (North-Central Atlantic) (Santos 1996: 160)

(a) *ì-ñí́*
CL5-elephant
'elephant'

(b) *bə-ỹí*
CL18-elephant
'big elephant'

It is relatively common for a Type 1 language to have several diminutive and augmentative noun class markers. This is discussed in §6.3.1 and illustrated with examples from North-Central Atlantic and Bantu languages. In addition, the diminutive and augmentative genders may trigger a special type of marking on nouns: this is discussed in §6.3.2 with examples from Bantu and North-Central Atlantic. §6.3.3 provides an insight into diachrony and discusses the renewal of the evaluative genders in the southeastern Bantu languages. Finally, in §6.3.4, I discuss evaluative genders in the Kwa language Sɛlɛɛ, where the use of diminutive gender markers is obligatorily combined with the use nominal diminutive suffixes.

6.3.1 Distribution of diminutive and augmentative genders and general characteristics

Within my sample, the languages of Type 1 differ in the number of diminutive and augmentative genders. For instance, some languages may have only one diminutive and one augmentative gender, whereas others may have more than one diminutive gender and more than one augmentative gender. When accounting for the distribution of diminutive and augmentative genders in my database, I only differentiate between "One" and "More than one" diminutive gender and "One" and "More than one" augmentative gender. However, within my sample, languages with more than one diminutive or more than one augmentative always have maximum two genders per evaluative type. The only exception is Venda (Bantu), which has three diminutive genders (but only two augmentative genders) (Poulos 1990). The distribution of diminutive and augmentative genders in the languages of Type 1 is summarised in table 6.2.

Table 6.2 shows that none of the attested distributional patterns is strikingly dominant. However, the most frequent pattern is the one with one diminutive gender and no augmentative gender and is attested in 7 out of 23 languages. Interestingly, in 6 of the 23 Type 1 languages – Bandial (North-Central Atlantic), Bemba (Bantu), Bidyogo (North-Central Atlantic), Lega (Bantu), Shona (Bantu) and Wamey (North-Central Atlantic) –, the system of augmentative marking is richer than the diminutive (more class distinctions are available). Out of these 23 languages, 5 have one diminutive and one augmentative gender. The combination of more than one diminutive and more than one augmentative gender is found only in two languages of the sample (the Bantu languages Lega and Venda). Finally, a system with more than one diminutive gender and no augmentative genders is only attested in two languages (the Bantu language Chiga and the Kwa language Sɛlɛɛ). Overall, the augmentative genders within my sample tend to be less differentiated than the diminutive genders from a morphological point of view. If one of the evaluative genders in a language lacks plural marking, this is the augmentative – as in Kirundi (Bantu) or Serer (North-Central Atlantic). Moreover, as shown in table 6.2, languages may have diminutive gender(s) but lack augmentative gender(s) – as in

Table 6.2: Distribution of diminutive and augmentative genders in Type 1 languages (23/100)

	Dim.	**Aug.**	**No. of lngs.**	**Abs. %**	**Genealogical groups**
	More than one	More than one	2	8.7%	Bantu (2/23)
	More than one	One	1	4.4%	North-Central Atlantic (1/7)
	One	More than one	6	26%	Bantu (3/23) North-Central Atlantic (3/7)
	One	One	5	21.7%	Bantu (4/23) North-Central Atlantic (1/7)
	More than one	None	2	8.7%	Bantu(1/23) Kwa (1/3)
	One	None	7	30.5%	Bantu (5/3) North-Central Atlantic (2/7)
Total			23	100%	

Nuclear Wolof (North-Central Atlantic) or Noon (North-Central Atlantic). The opposite is not attested in the languages of the sample: all the languages with augmentative genders have at least one diminutive gender (see also §4.4 for more general tendencies in the distribution of diminutive and augmentative markers in the languages of the sample).

If a language has more than one evaluative gender per evaluative meaning (that is, more than one diminutive and/or augmentative gender), the different available markers may differ from each other in terms of distributional preferences. For instance, they may specialise in the encoding of different size nuances, i.e., 'small' vs. 'tiny' or 'big' vs. 'huge.' This is the case in the Bantu language Lega, where the diminutive marker *ka-*, Class 12, encodes small size, whereas the Diminutive Class *si-*, Class 19, encodes extremely small, or tiny, size. The uses of the two prefixes are illustrated in (6.4).

(6.4) Lega (Bantu) (Botne 2003: 430)

(a) *mu-ntu*
CL1-person
'person'

(b) *ka-ntu*
CL12-person
'small person'

(c) *sị̀-ntu*
CL19-person
'tiny person'

Another possibility is that some of the individual diminutive and augmentative classes are used with restricted types of nouns as in Wamey (North-Central Atlantic), where the augmentative marker *ga-*, Class 20, is used with nouns originally assigned to Classes 1 and 3 only (example (6.5)), whereas the Augmentative Class *bə-*, Class 18, is used to form the augmentative of nouns from any other class (example (6.3)).

(6.5) Wamey (North-Central Atlantic) (Santos 1996: 160)

(a) *à-sæ̀n*
CL1-man
'man'

(b) *ga–sæ̀n*
CL20-man
'big man'

Such patterns of differential marking are usually neutralized under plural reference. Thus, in Lega, the Diminutive Plural Class correspondent to the Diminutive Singular Classes 12 and 19 is Class 13, which corresponds to the marker *tu-*; likewise, in Wamey, the Augmentative Plural Class correspondent of the Augmentative Singular Classes is *va-* (Class 19).

Synchronically, two types of evaluative nouns classes can be distinguished:

(1) Noun classes that are extremely polysemous and productive in language use, and are also used to derive diminutives and augmentatives when nouns assigned by default to other classes are shifted to them. I call these Extended Evaluative Classes (henceforth EECs)

(2) Noun classes that are *only* used to derive diminutives and augmentatives. Generally, there are no (or very few) nouns that are assigned by default to any such classes; these essentially function as word-formation strategies. I call these Only-Evaluative Classes (henceforth OECs).

Languages can have both EECs and OECs or only one of them. Bidyogo (North-Central Atlantic) and Shona (Bantu) provide a very good illustration of the coexistence of the two types of evaluative classes.

In Bidyogo, Class E- and KO-[44] are a good example of EECs. Class E- is one of the most frequently used and one of the most bleached from a semantic point of view (it is

[44] In his grammar of Bidjogo, Segerer (2002) uses capital letters to refer to the phonological realisation of a class marker; this orthographic convention is also used in the glosses as a means for referring to the individual noun classes and the indexation patterns associated with them. The noun classes are not numbered.

also the noun class to which loanwords are assigned by default). Nouns assigned to class E- are most commonly shifted to class KO- to encode number distinctions. However, the interpretation of the two class markers in terms of number distinctions is subject to vary. With some nouns, E- works as the singular and KO- as the plural, whereas with other nouns the two markers work in exactly the opposite way. E- and KO- also function as evaluative markers: the polarity between the two classes is maintained, but, again, the semantic values associated with each class vary according to the type of noun that they classify and depending on the context. Thus, E- is both a diminutive and an augmentative and so is KO-. Consider the following examples:

(6.6) Bidjogo (North-Central Atlantic) (Segerer 2002: 103)

(a) *kɔ-kɔɲ*
KO-palm.leaf
'palm leaf'

(b) *ɛ-kɔɲ*
E-palm.leaf
'small palm leaf'

(c) *ɛ-man*
E-rice
'rice'

(c) *kɔ-man*
KO-rice
'grain of rice'

In opposition to Class E- and KO-, Class BA- is a perfect example of OEC. It is only used to derive augmentatives, and there are no nouns that are inherently assigned to it.

(6.7) Bidjogo (North-Central Atlantic) (Segerer 2002: 125)

(a) *jɔ-kɔ*
JO-house
'house'

(b) *ba-kɔ*
BA-house
'big house'

In Shona, singular augmentatives can be encoded either by shifting a noun to Class 5 or to Class 21. Class 5, (6.8), is one of the most productive and least semantically transparent noun classes in the language. It is thus another good example of an EEC. In Fortune's (1955) analysis, the underlying form of Class 5 is *ru-*, but the only phonetic realisation of the prefix is the voicing of the initial consonant of a noun stem – if unvoiced – as in (6.8). On the other hand, Class 21, (6.9), is used only to form augmentatives and is thus an instance of OEC. Interestingly, Classes 21 and 5 trigger the same indexation (Fortune 1955: p.105).

(6.8) Shona (Bantu) (Fortune 1955: 77)

(a) *mu-kwasha*
CL1-son-in-law
'son-in-law'

(b) *gwasha*
CL5.son-in-law
'big son-in-law'

(6.9) Shona (Bantu) (Fortune 1955: 104)

(a) *rw-eŋga*
CL11-frying.pan
'frying pan'

(b) *z-eŋga*
CL21-frying.pan
'large frying pan'

Nouns assigned by default to Class 5 are shifted to Class 21 to form their augmentatives. The opposition between Class 5 and Class 21 in the encoding of augmentatives is neutralized under plural reference: plural augmentatives are always in Class 6.

Interestingly, with respect to diminutive formation, Shona only has OECs. Diminutives are encoded by shifting nouns from their default genders to Class 12, *ka-*, for diminutive singular, and Class 13, *tu-*, for diminutive plural. These classes are exclusively used for diminutive formation and have no default members (Fortune 1955: 94-95).

The evaluative classes of Swahili (Bantu) are closer to the EEC type. Shifts to Classes 7, *ki-*, and 8, *vi-*, are used for the formation of singular diminutives and the plural diminutives. Classes 5, *ji-*, and 6, *ma-*, are used for the formation of singular augmentatives and plural augmentatives. The two class pairs encode evaluative meanings only as a result of gender shift. Otherwise they are two of the most productive genders in Swahili. Based on the description by Contini-Morava (2000), a wide range of nouns, covering a wide range of semantic fields, is assigned by default to these genders. In the case of the 5/6 class pair, there is apparently no connection between the semantics of the nouns assigned by default to this class pair and the augmentative meanings that this encodes via gender shift. On the other hand, Contini-Morava describes the semantics of the nouns assigned by default to the 7/8 pair as being generally associated to "smallness of size" (Contini-Morava 2000: 10).

6.3.2 Multiple or additive class marking with diminutives and augmentatives

An interesting morphological property characterises the evaluative classes of some Bantu languages: nouns that are shifted to the evaluative classes do not necessarily lose their original class markers. On the contrary, they retain them and combine them with the evaluative class markers, which ultimately control indexation. The phenomenon has

already been introduced in §4.2.3.1, where I discussed the distinction between *additive* and *replacive* types of class markers in Bantu. In this section, I examine a couple of examples from Bemba (Bantu) and Kikuyu (Bantu), which illustrate this phenomenon. Two patterns are attested in these languages:

1. Nouns from Classes 1 and 2 (mostly humans), and 3 and 4 (generally plants and some animals) maintain their original class prefix when shifted to the diminutive class. This is shown in examples (6.10a) and (6.10b).

2. The original prefix of the noun is not retained. This applies to the majority of the nouns undergoing diminution and is shown in examples (6.10c) and (6.10d).

(6.10) Kikuyu (Bantu) (Stump 1993: 8-9)

(a) *mû-raata*
CL1-friend
'friend'

(b) *ka-mû-raata*
CL12-CL1-friend
'small friend'

(c) *i-rima*
CL5-hole
'hole'

(d) *ka-rima*
CL12-hole
'small hole'

Multiple class marking is found associated with the marking of evaluative morphology elsewhere in Bantu. According to Kavari & Marten (2009), for instance, multiple class marking in Otjiherero is associated with the marking of diminutive, augmentative and locative meanings, as well as with some kinds of plural formation.

A systematic survey of multiple class marking in Bantu is still missing, and the occurrence of the phenomenon in individual languages is usually explained in morphophonological terms. At a higher level of abstraction, the phenomenon of multiple class marking is an interesting indicator of the fact "that the lexical use of noun classes is at least in some instances morphologically distinguished from the derivational uses of classes" (Crisma et al. 2011).

I did not find any example of multiple class marking in the North-Central Atlantic and Mel languages of the sample. An arguable case is found in Bandial (North-Central Atlantic). In Bandial tales, animal names are personified by means of the prefix *ja-*. The prefix is analysed by Tendeng (2000) as a combination of the diminutive class marker *ju-* – Class 11 – and the class marker *a-*, which is characteristic of nouns denoting humans. Further on, the author states that the two class prefixes merge as the result of a phonological rule of vowel deletion, whereby /u/ is dropped before the *a-* prefix.

Semantically, the diminutive class marker added to the marker for humans – Class 1 – functions as a sort of attenuator: animals are not just construed as human but as "slightly human." If Tendeng's (2000) hypothesis is accepted, then personification constructions in Bandial can be seen as an instance of multiple class marking. The explanation offered by Tendeng has received mixed reactions in the literature. For instance, Sagna (2008: 227) suggests that the cancellation of /u/ before *a-* may not be a synchronically productive phonological rule of the language. Example (6.11) illustrates the pattern of personification of animal names in Bandial.

(6.11) Bandial (North-Central Atlantic) (Sagna 2008: 228)

já-munduno	*a-jeg-or*	*me*	*bú-sol*	*maa...*	
CL11b-Hyena	CL1.3SG-turn-RECP	SUBORD	CL5a-back	like	this

'Every time Hyena turns back like this'

In addition, Tendeng's argument seems to be further weakened by the fact that in Bandial, diminutivized human nouns do not retain their default class prefix, as illustrated by example (6.12).

(6.12) Bandial (North-Central Atlantic) (Sagna 2008: 255)

(a) *a-ñ̃ñ̃il*
CL1-child
'child'

(b) *ju-ñ̃ñ̃il*
CL11-child
'baby'

Following Sagna's (2008) convention, the alleged "diminutive + human class" combination is treated as an allomorph of the Diminutive Class and is referred to as Class 11b. Personified nouns trigger Class 1 indexation, which could be explained as an instance of referential indexation.

6.3.3 The renewal of evaluative morphology in the southeastern Bantu languages

In §6.3, I showed that two types of evaluative classes can be found in the languages of Type 1: the EECs and the OCSs. The former are noun classes with a very productive range of uses and very loose semantics, which are also used to encode diminutives and augmentatives. Conversely, the OECs are noun classes that are only used to express evaluative meanings. In this section, I focus on the diachronic stability of the OECs in the Bantu languages. In fact, there is evidence to believe that, at least in some Bantu languages, the OECs are diachronically less stable than the EECs. When the system of noun class marking of a Bantu language erodes, these classes – often together with the locative classes – disappear. In such cases, either only the EECs survive – as in Swahili

– or new evaluative markers develop. These new evaluative markers may be completely independent of grammatical gender or may coexist (and sometimes co-occur) with the EECs. In the rest of this section, the diachronic development and the usages of such markers are discussed.

The marking of diminutives and augmentatives underwent massive change in Bafia, Eton, Northern Sotho, Shona, Swati, Tswana, Venda and Zulu. The innovations follow two paths: on the one hand, Northern Sotho, Shona, Swati, Tswana, Venda and Zulu – all spoken in the southeastern region of the Bantu-speaking area – have developed diminutive and augmentative suffixes; on the other hand, Bafia and Eton have developed strategies for the encoding of diminutives and augmentatives that differ both from class marking and from suffixation. I begin with an overview of the origin and uses of diminutive and augmentative suffixes in the southeastern Bantu languages and move on to Bafia and Eton thereafter.

Previous studies (Creissels 1999; Güldemann 1999) have shown that the diminutive suffix in southeastern Bantu originates from the Proto-Bantu word for 'child' **jana* (Creissels 1999: 34) or **yana* (Güldemann 1999). Table 6.3 provides an overview of the suffixes attested across the six languages and notable facts concerning their most typical uses. Information about the co-occurrence of diminutive suffixes and diminutive noun classes is also given.

Table 6.3: The diminutive suffix in some southeastern Bantu languages

Language	**Dim. suffix**	**Notable facts**	**Dim. noun classes**
Northern Sotho	*-ana*, *-nyana*	Lobedu dialect: noun class shift + suffix	Yes
Shona	*-ana*	Very incipient strategy, often co-occurring with the prefix	Yes
Swati	*-ana*, *-anyana*	*-ana*= small size ; *-anyana*= tiny size	No
Tswana	*-ana*,*-nyana*	With animate nouns: *-ana*= young age; *-nyana*= small size	No
Venda	*-ana*; *-nyana*	Noun class shift + suffix	Yes
Zulu	*-ana*	–	No

As shown in the table, languages vary in their use of the diminutive markers and in the extent to which the diminutive noun class prefixes are maintained. However, some general patterns can be observed throughout the data. Four languages (Northern Sotho, Swati, Tswana and Venda) have two diminutive suffixes, the regular form and a form that is described by grammars as "reduplicated." While the reduplicated form is used for expressing tiny size in Swati, in Tswana, both forms alternate in order to express age (*-ana*) or size (*-nyana*) with animate nouns. This is exemplified in (6.13). With inanimate nouns, the two suffixes are interchangeable.

(6.13) The diminutive suffix in Tswana (Bantu) (Cole 1955: 109)

(a) *nku*
'sheep'

(b) *nku-ana*
sheep-DIM
'young sheep'

(c) *nku-nyana*
sheep-DIM
'small sheep'

In the Lobedu dialect of Northern Sotho, Shona and Venda, the Diminutive Classes are still very productive and the two strategies (prefixal and suffixal) often co-occur. In Venda, different sizes are expressed by the prefix on its own (either Class 7 or 20) and by the combination of prefix and suffix.

(6.14) Diminutive marking in Venda (Bantu) (Poulos 1990: 88)

(a) *tshi-kali*
CL7-clay.pot
'small clay pot'

(b) *tshi-kal-ana*
CL7-clay.pot.DIM
'very small clay pot'

(c) *ku-kali*
CL20-clay.pot
'small clay pot'

(d) *ku-kali-ana*
CL20-clay.pot-DIM
'very small clay pot'

One of the few cases in which the Venda diminutive suffix occurs on its own is given in (6.14).

(6.15) Venda (Bantu) (Poulos 1990: 86)

(a) *kholomo*
'cow'

(b) *nam-ana*
cow-DIM
'calf'

In (6.15), the diminutive suffix is attached to a different noun stem: such kinds of stem alternations are cross-linguistically very common for differentiating age in domestic animals (Ljuba Veselinova, personal communication). Moreover, the word for 'calf' in Venda may also be interpreted as a case of lexicalization, whereby speakers of Venda do not actually analyse it as a diminutive noun any longer (Mark Van de Velde, personal communication).

In addition to the diminutive suffix, an augmentative suffix is also found among the southeastern Bantu languages. However, its use as an augmentative marker is restricted to Swati, Zulu, and, more rarely, Tswana. In Shona, Northern Sotho, and Venda the same suffix is attested but is only used to derive nouns denoting females. The augmentative affix originated from the proto-Bantu noun **kádi* 'woman'. Examples of the augmentative suffix in Swati and Zulu follow.

(6.16) Swati (Bantu) (Taljard et al. 1991: 141)

(a) *li-tje*
CL5-stone
'stone'

(b) *li-tje-kati*
CL5-stone-AUG
'rock'

(6.17) Zulu (Bantu) (Poulos & Bosh 1997: 9)

(a) *umu-ntu*
CL1-person
'person"

(b) *umu-ntu-kazi*
CL1-person-AUG
'huge person'

Bantu nominal morphosyntax is prominently prefixal. Thus, the development of nominal suffixes such as the diminutive, the feminine/augmentative, and the locative suffixes among the southeastern Bantu languages poses a relevant problem in the general understanding of the typology of Bantu nominal morphology. In Güldemann (1999), the process is explained as a result of areal contact with predominantly suffixal languages spoken in the same geographical area. According to his analysis, the best candidate contact languages are Khoe and non-Khoe languages spoken in southern Africa (such as Nama or Ju|'Hoan). In these languages, diminutive, augmentative and locative meanings are encoded by means of suffixes or postposed relational nouns. The southeastern Bantu languages discussed in this section acquired suffixal strategies to mark evaluative morphology but not the markers as such. The phenomenon came as an effect of intense cultural, social and economic contact (Güldemann 1999). The sociocultural context which fostered the grammaticalization of the diminutive and augmentative/feminine suffixes is also discussed in §6.5.

I now turn to a short overview of the evaluative markers of the Bantu languages Bafia and Eton. As mentioned at the beginning of this section, the evaluative markers of these two languages are neither prefixal nor suffixal. In Bafia, diminutives are encoded periphrastically: the noun to be diminutivized is preceded by the word *mán*, 'child', Gender 1/2 (plural: *ɓɔ́ń*). This behaves as a fully lexical noun and triggers Gender 1/2 indexation. This is illustrated in example (6.18).

(6.18) Bafia (Bantu) (Guarisma 2003: 318)

m-á	*zaʔ*	*á*	*ń-púp*
CL1-child	CL9chicken	CL1	CL1-white

'small white chicken'

In (6.18) the noun that undergoes diminution, *zaʔ*, is by default assigned to Gender 9/10[45] but the nominal modifier for 'white' indexes the gender of the diminutive marker (which, as mentioned before, is a noun belonging to Gender 1/2). Augmentatives are also periphrastic in Bafia: the noun for 'thing' (*cóm / ɓyóm*, Gender 7/8) is used in combination with the noun to be augmented and, as in the case of the diminutive marker, triggers indexation.

In Eton, diminutives and augmentatives are encoded by means of proclitic words. The lexical source of the diminutive proclitic, *mɔ=* (plural *bɔ*), is the noun for 'child'(*mɔŋɔ/ bɔŋɔ*), from which the diminutive marker differs in virtue of some phonological erosion and the floating high tone. The augmentative proclitic, *mòd* (plural *bǒd*), is formally identical to the word for 'person', and differs from it only for its final floating high tone (Van de Velde 2008: 208).

From a morphotypological point of view, the innovations in Bafia and Eton follow the expected pattern of Bantu nominal morphology: the nouns and the clitics that encode evaluative meanings in Bafia and Eton, respectively, precede the nouns that undergo diminution or augmentation, as in the case of the prefixal diminutives and augmentatives. Bafia and Eton are both spoken in Cameroon and belong to the same Bantu sub-area. It might be that their innovations in the marking of diminutives and augmentatives are geographically restricted similarly to those encountered in the southeastern languages.

6.3.4 Diminutive marking in Sɛlɛɛ

In this section, I focus on the diminutive constructions of Sɛlɛɛ, a Kwa language of the Ghana-Togo-Mountain subgroup that is spoken in the Volta region of Ghana. In contrast with the other two Kwa languages in the sample, Akan and Ewe, which do not have gender, Sɛlɛɛ has a non-sex-based gender system similar to the Bantu and North-Central Atlantic type. Besides, Sɛlɛɛ has a very interesting inventory of diminutive constructions, whereby diminutive markers internal to the gender system are always used in combination with diminutive suffixes marked on nouns. These suffixes are extraneous

[45] In Bafia, when nouns assigned to Gender 9/10 begin with a voiced consonant, the noun class prefixes are not overtly coded on nouns (Guarisma 2003).

to the gender system and have historical origins similar to the diminutive suffixes of the southeastern Bantu languages (see 6.3.3). A detailed description of the gender system of Sɛlɛɛ is found in Agbetsoamedo (2014), whereas Agbetsoamedo & Di Garbo (forthcoming) account for Sɛlɛɛ evaluative morphology.

According to Agbetsoamedo & Di Garbo (forthcoming), Sɛlɛɛ does not have morphological augmentatives and the attested diminutive constructions conform to two main constructional patterns:

(1) Gender shift + suffixation. Nouns are shifted either to Gender 5/8 or to gender 3/7, and are marked by one of the following suffixes: *-bi*, *-mii*, *ɛ*, *-nyi*.

(2) Suffixation. *-bi* is the only suffix that can occur without gender shift. This happens when the diminutive of inanimate nouns is encoded.

It is worth mentioning that, among the diminutive suffixes mentioned above, *-bi* is the only one for which a clear grammaticalization path has been established. The suffix originates from the noun *obi* 'child,' which is still a productive lexical item in the language (for more details, see Agbetsoamedo & Di Garbo forthcoming).

In the rest of this section, I focus on the diminutive constructions of the type described in (1), as they are the only ones where interaction between the marking of gender and evaluation occurs. Examples (6.19), (6.20) and (6.21) illustrate some instances of the pattern described in (1). Note that for each gender, only the singular forms are given.

(6.19) The Diminutive in Sɛlɛɛ (Kwa): shift to Gender 5/8 + suffix *-bi*
(Agbetsoamedo & Di Garbo forthcoming)

(a) *ɔ-kla*
CL1-cat
'cat'

(b) *lɛ-kla-bi*
CL5-cat-DIM
'kitten'

(6.20) The Diminutive in Sɛlɛɛ (Kwa): shift to Gender 3/7 + suffix *-mi*
(Agbetsoamedo & Di Garbo forthcoming)

(a) *kansiɛ*[46]
CL1.bird
'bird'

(b) *ka-kansiɛ-mii*
CL3-bird-DIM
'small bird'

[46] In Sɛlɛɛ, some nouns within Gender 1/2 may lack overt coding of gender. These are mostly borrowings. See Agbetsoamedo (2014) for details.

(6.21) The Diminutive in Sɛlɛɛ (Kwa): shift to gender 3/7 + suffix *-nyi* (Agbetsoamedo & Di Garbo forthcoming)

(a) *o-ti*
CL1.person
'person'

(b) *ka-ti-nyi*
CL3-person-DIM
'tiny person' (derogatory)

Agbetsoamedo & Di Garbo (forthcoming) show that not all the possible combinations between gender shift and suffixation are grammatical or equally productive in Sɛlɛɛ. For instance, suffix *-bi* tends to combine with shift to Gender 5/8 whereas the suffixes *-mi*, *ɛ* and *-nyi* always combine with shift to Gender 3/7. Such distributional preferences are also semantically relevant. With nouns referring to animal nouns, the diminutive construction "shift to Gender 5/8 + *-bi*" encodes young age or offspring (6.19), whereas the construction "shift to Gender 3/7 + *-mii* (or *-ɛ*)" expresses small size (6.20). Finally, the construction "shift to Gender 3/7 + *-nyi*" is attested only with two nouns, 'person' – as in (6.21) – and 'hand,' and is always derogatory.

Such a division of labour between the different possible diminutive constructions involving gender shift is similar to the situation described in §6.3.3 for the southeastern Bantu languages. The major difference between Sɛlɛɛ and the southeastern Bantu languages is that in Sɛlɛɛ, none of the noun class markers involved in diminutive marking can express diminution independently of suffixation. Agbetsoamedo & Di Garbo (forthcoming) suggest that in Sɛlɛɛ – similarly to the Bantu languages – "gender shifts might have been the original strategy for the encoding of diminutives" and that, later on, this started being combined (at the beginning, presumably only optionally) with a set of diminutive suffixes. Interestingly, the most recent and semantically transparent diminutive suffix in Sɛlɛɛ, *-bi*, is the only one that, at least in some cases, can occur without gender shift.

6.3.5 Type 1: summary

In this section, I discussed languages for which diminutives and augmentatives are part of the inventory of gender distinctions (Type 1 in the classification presented in §6.2). The distribution of Type 1 languages within my sample is highly genealogically skewed: they are only attested among Bantu, North-Central Atlantic and Kwa, that is, in languages with large and non-sex-based gender systems. Two types of evaluative noun classes can be found in Type 1 languages: the EECs, which are *not* only used to express diminutive and augmentative meanings, and the OECs, which are *only* used to express evaluative meanings. Languages differ as for whether or not they have both or only one type of evaluative class. In §6.3.2, I discussed instances of multiple class marking that are related to the use of the diminutive and augmentative classes in the languages of Type 1. The diachronic stability of Type 1 languages was addressed in §6.3.3, where I discussed the

renewal of evaluative classes in southeastern Bantu languages. Similar phenomena were discussed in §6.3.4 in connection with the Kwa language Sɛlɛɛ. In Sɛlɛɛ, the diminutive noun classes cannot be used on their own but always co-occur with the diminutive suffixes.

6.4 Type 2: Diminutives and augmentatives in languages with sex-based gender

Type 2 languages (28 out of 100; see table 6.1) have sex-based gender. Depending on the number of distinctions that they exhibit in their gender systems, Type 2 languages use shifts from masculine to feminine, from feminine to masculine, or from masculine/feminine to neuter as a strategy for expressing evaluative meanings. This is summarised in table 6.4.

Table 6.4 shows that Type 2 languages with two genders most commonly use gender shifts from masculine to feminine to encode diminutive meanings but lack a similar mechanism for the encoding of augmentative meanings. This pattern is found in 12 languages. Slightly less common are systems in which the possibility of gender shift is bidirectional. In these systems, when nouns are shifted from masculine to feminine, the resulting meaning is diminutive; on the other hand, when nouns are shifted from feminine to masculine, the resulting meaning is augmentative. This pattern is attested in 10 of the 28 Type 2 languages in my sample. Finally, in the isolate Hadza, gender shifts from the Masculine to the Feminine Gender are used for encoding augmentative meanings (Edenmyr 2004: 17). In two languages with three gender distinctions, gender shifts from the masculine and/or feminine to the neuter gender are used as a strategy for encoding diminutive meanings. These languages are Karamojong and Turkana, both belonging to the Eastern Nilotic group. An interesting pattern is attested in the Khoe-Kwadi languages of the sample. Even though these languages have three genders – Masculine, Feminine and Common – the Common Gender is never involved in gender shifts encoding size variation. In two languages, Kxoe and Naro, shifts from the Masculine to the Feminine encode diminutive meanings, whereas shifts from the Feminine to the Masculine encode augmentative meanings. In Nama, only shifts from the Masculine to the Feminine Gender are documented in my sources and they are used to express diminution. Examples from Nama are given in §6.4.1.

The general characteristics of Type 2 languages are discussed in §6.4.1. As opposed to Type 1 languages, which are only attested among the North-Central Atlantic, Bantu and Kwa languages of the sample, Type 2 languages have a wider distribution across genealogical groupings. This is examined in detail in §6.4.2.

Table 6.4: Number of gender distinctions and type of attested gender shifts in Type 2 languages (28/100)

	No. of genders	**Type of shift**	**No. of lngs.**	**Abs. %**	**Genealogical groups**
	2	M → F = Dim. F → M = Aug.	10	35.8%	Berber (6/6) Eastern Nilotic (1/3) Semitic (1/7) Ta-Ne-Omotic (2/4)
	2	M → F = Dim.	12	43%	Cushitic (3/13) Chadic (2/8) Dizoid (1/1) Sandawe (1/1) Semitic (2/7) South Omotic (1/1) Ta-Ne-Omotic (2/4)
	2	M → F = Aug.	1	3.5%	Hadza (1/1)
	3	M → F = Dim. F → M = Aug.	2	7.1%	Khoe-Kwadi (2/5)
	3	F → M =Dim	1	3.5%	Khoe-Kwadi (1/5)
	3	M/F → N = Dim.	2	7.1%	Eastern Nilotic (2/3)
Total			28	100%	

6.4.1 Examples of Type 2 languages

Afro-Asiatic languages may be defined as hotbeds for Type 2 phenomena (see §6.4.2 for more details). To give an example, in all the the Berber languages of my sample, shifts from the Masculine to the Feminine Gender and *vice versa* are used to encode variation in size with inanimate nouns. The evaluative value conveyed by gender shifts depends on the default gender of a noun and the construal of the NP referent with respect to the parameter of size. Inanimate masculine nouns are shifted to the Feminine Gender when a diminutive construal is intended. On the other hand, inanimate feminine nouns are shifted to the Masculine Gender when an augmentative construal is intended. If the default gender of a noun meaning, for example, 'basket' is Feminine, the Masculine Gender may be used to refer to a basket that is bigger than expected. Gender shifts are mostly used with objects that appear in different sizes, for instance, a small and a big pot, a small and a big jewel, etc. (Amina Mettouchi, personal communication). Examples (6.22) and (6.23) show how size-related gender shifts in Tachawit.

(6.22) Tachawit (Berber) (Penchoen 1973a: 12)

(a) *aq-nmuš*
[M-]SG-pot
'pot'

(b) *taq.nmuš-t*
F-SG-pot-F.SG
'small pot'

(6.23) Tachawit (Berber) (Penchoen 1973a: 12)

(a) *taɣ-nžak-t*
F-SG-spoon-F
'spoon'

(b) *aɣ-nž*
[M-]SG-spoon
'big spoon, ladle'

In example (6.22), a masculine noun is shifted to the Feminine Gender to encode reduced size, whereas in example (6.23), an inherently feminine noun is shifted to the Masculine Gender to express increased size. What happens then if speakers want to express notions like 'big pot' or 'small spoon'? In other words, how do speakers of Tachawit (and other Berber languages) construe size values for which the gender shift strategy cannot be used? In such cases, speakers would use adjectives for 'big' and 'small' in combination with the noun. The notion of 'pot' and 'small pot,' and 'spoon' and 'big spoon,' as expressed by the nouns in examples (6.22) and (6.23), are, to a certain extent, lexicalized insofar as they are built into gender assignment. Thus, speakers can combine the adjective for 'big' and 'small' with any of these nouns. For instance, the adjective for 'big' could co-occur with the noun *taqnmušt* 'small pot' (6.22b), when speakers want

to make a distinction between two pots of different sizes and point at the *big small pot* (Amina Mettouchi, personal communication). The same principle holds for the use of the adjective for 'small' with the noun *aɣnž* 'big spoon, laddle' (6.23).

In the Berber languages, nouns denoting humans and higher animals can also undergo gender shifts. However, in such cases, the gender shift is interpreted as encoding sex rather than size. As pointed out by Grandi (forthcoming-a) with nouns that "designate living beings without sexual dimorphism," such as reptiles, fish, and bugs, gender shifts can be interpreted both as sex or size-related, depending on the context. An example is given in (6.25), where the feminine form of the noun for 'lizard' can be interpreted as referring either to a female lizard or to a small one.

(6.24) Tachelhit (Berber) (as cited in Grandi forthcoming-a)

t-a-herdan-t
F/DIM.F-SG-lizard-F/DIM.F
'small lizard' or 'female lizard'

Lexical restrictions to this pattern are also found. For instance, again in Tachelhit, when the noun for 'bird' is treated as feminine, this can only be interpreted as encoding diminution. This is illustrated in (6.25).

(6.25) Tachelhit (Berber) (as cited in Grandi forthcoming-a)

ɸ-a-fruq-ɸ
DIM.F-SG-bird-DIM.F
'small bird'

In Masai (Eastern Nilotic), speakers can use gender shifts with human nouns to convey size differences (and derogation) and not only biological gender. This is shown in (6.26).

(6.26) Masai (Eastern Nilotic) (Payne 1998: 166)

(a) *ɛnk-anáshɛ̀*
F.SG-sister
'sister'

(b) *ɔnk-anáshɛ̀*
M.SG-sister
'very large sister' (*pejorative*)

(c) *ɔ-aláshɛ̀*
M.SG-brother
'brother'

(d) *ɛnk-aláshɛ̀*
F.SG-brother
'weak brother'

The pattern illustrated in (6.26) is crosslinguistically very rare: the use of gender shifts to encode diminutives and augmentatives is usually restricted to inanimate nouns, as in the case of the Berber languages (see also data in Aikhenvald 2003). In my sample, Masai is the only language where this pattern is attested with animate nouns. The gender system of Masai is very peculiar also in other respects. Masai has two productive genders: Masculine and Feminine.[47] In addition, as shown by Payne (1998) and Shirtz & Payne (2012), nouns in Masai can be classified according to their compatibility with the Masculine Gender and the Feminine Gender:

> [s]ome nouns stems seem equally compatible with both. Other seem to have a 'default' gender and reserve the other for pejorative or unusual size readings. Only a small subset of the lexicon is compatible with only one gender. Finally, there is a small set of noun stems which is completely incompatible with the category of gender. (Shirtz & Payne 2012)

Based on this distinction, for the majority of nouns in Masai, gender assignment is not *lexically* specified but is the result of the interplay between the denotation of a noun and the speaker's *construal* of the NP referent. According to Payne (1998) and Shirtz & Payne (2012), the Masai gender system challenges existing models of gender assignment whereby nouns are allotted to a gender on the basis of their meaning, their inflectional class or a combination of both. I shall come back to the relevance of their argument in the final discussion chapter (see chapter 8).

Finally, in nearly all Type 2 languages with two genders, size-related gender shifts are driven by the association between feminine gender and small size, and masculine gender and big size. The opposite (*feminine is large/big* and *masculine is small*) is attested in one language, the isolate Hadza. Notice that Masculine is the unmarked gender in Hadza.

(6.27) Hadza (isolate) (Edenmyr 2004: 16)

(a) *ʔato*
small.axe.M
'small axe'

(b) *ʔato-ko*
axe-F
'large axe'

In the languages with three gender distinctions (masculine, feminine, neuter/common), gender shifts that are relevant for the encoding of evaluative meanings can be restricted to the masculine and the feminine genders only (as in the Khoe-Kwadi languages) or can extend to the third gender. This is the case of the Eastern Nilotic languages Turkana and Karamojong. The third gender in Turkana and Karamojong – labelled as Neuter

[47] There is also a third gender, the Locative, which is, however, very rare and can be regarded as an instance of inquorate gender (see §4.2.2)

Gender in grammars – is the lexical gender of a very limited number of nouns as opposed to the more productive Masculine Gender and Feminine Gender. In Turkana and Karamojong, inherently neuter nouns are nouns that denote the offspring of animate entities or individual members/instances of groups. In Turkana, animate and inanimate feminine and masculine nouns can be shifted to the Neuter Gender to encode young age (6.28 a and b), small size (6.28 c and d), or small quantity (6.28 e and f), depending on the semantics of the noun and its countability properties.

(6.28) Turkana (Eastern Nilotic) (Dimmendaal 1983: 218)

(a) *a-kale*
F.SG-goat
'she-goat that has not yet produced young'

(b) *ɪ-kle*
N.SG-goat
'kid-goat'

(c) *e-dya*
M.SG-boy
'boy'

(d) *i-dya*
N.SG-boy
'small boy'

(e) *ŋa-kɔt*
F.PL-blood
'blood'

(f) *ŋi-kɔt*
N.PL-blood
'A little blood'

In Karamojong, shifts to the Neuter Gender are only used with animate nouns and indicate offspring (Novelli 1985).

As mentioned above (see also table 6.4), in the Khoe-Kwadi languages, size-related gender shifts do not involve the third gender, traditionally labelled as the Common Gender. Gender shifts from Masculine to Feminine and from Feminine to Masculine are attested in Kxoe and Naro. In Nama, the only examples of gender shifts mentioned in the grammar involve shifts from Feminine to Masculine, which are used to express diminutive meanings. These are presented in (6.29).

(6.29) Nama (Khoe-Kwadi) (Hagman 1977: 23)

(a) *'om-s*
house-F.SG
'house'

(b) *'om-i*
house-M.SG
'big house, apartment or office building'

(c) *xãa-s*
penis-F.SG
'penis'

(d) *xãa-p*
penis-M.SG
'small penis'

In Nama, size-related gender shifts are restricted to inanimate nouns and "express that there is something unusual about the referent of the noun" (Hagman 1977: 23). Hagman states that the nature of the semantic shift is hard to pin down since it depends on the semantics of the noun that undergoes gender shift and on the general discourse context in which the noun is used. He continues by saying that if the typical appearance of a NP referent is not large/big, gender shifts encode increased, large/big size; on the other hand, if it is usual for a noun referent to be large/big, gender shifts mark unexpected small size. Usually, if the largeness or smallness of size are undesirable for the object in question, the encoding of dimensional variation also entails derogation. Nama has also specialised diminutive and augmentative suffixes (see discussion in §6.4.2).

6.4.2 The distribution of Type 2 languages

As mentioned in §6.4, Type 2 languages are found virtually among all the genealogical groupings with sex-based gender: Berber, Chadic, Cushitic, Dizoid, Khoe-Kwadi, Eastern Nilotic, South Omotic, Semitic, Ta-Ne-Omotic and the isolates Hadza and Sandawe (table 6.1). The distribution of size-related gender shifts within the individual genealogical groups varies: they are attested in all the sampled Berber varieties, in all the sampled Omotic groups, and in all the sampled Khoe-Kwadi languages (with the exception of ǁAni and Kwadi). The existence of size-related gender shifts appears to be less pervasive in Chadic, Cushitic, and Semitic. The phenomenon was explicitly mentioned by my sources only in the case of two Chadic languages out of eight (Gidar and Hausa), and three Cushitic languages out of thirteen (Awngi, Beja and Daasanach). I cannot exclude that similar phenomena might also exist in other Cushitic and Chadic languages of the sample for which I could not retrieve any information in the sources that I consulted.

Noteworthy are the findings from Semitic. The only languages that can be clearly classified as Type 2 are Amharic, Tigre and Maltese. In his survey of diminutive constructions in Hebrew, Bolozky (1994: 50) mentions the existence of a set of lexicalized diminutives that differ from their correspondent nondiminutivized forms only in gender (the lexicalized diminutives are all feminine). However, he continues, gender shifts are no longer perceived as a productive strategy of diminutive formation by speakers of Israeli Hebrew (1994: 51). No relationship between gender shift and evaluation is

found in Standard Arabic[48] and Moroccan Arabic, whereas no information was at all recoverable on evaluative constructions in the Aramaic language Lishan Didan. Due to intense contact with Romance languages, the evaluative morphology system of Maltese displays an interesting interplay between indigenous (Semitic) and borrowed (Romance) evaluative markers. In this section, I mainly focus on those aspects of Maltese evaluative morphology that concern its interaction with gender marking (see Grandi 2002: 221-231, for a complete inventory of the evaluative constructions of Maltese). As is typical of other Semitic languages, diminutives in Maltese are formed by means of discontinuous infixal morphemes, that is, by a set of vowel templates that are inserted in the root in different ways, mostly depending on the number of consonants that a given root has. In some cases, vowel infixation is accompanied by a shift in gender, notably from Masculine to Feminine. This is exemplified in (6.30).

(6.30) Discontinuous infixal diminutive + gender shift in Maltese (Semitic) (adapted from Grandi 2002: 225)

(a) *xat*
beach.M
'beach'

(b) *xt**ajt**a*
beach:DIM.F
'small beach'

In other cases, the diminutive of a noun can be encoded only by means of gender shift, that is, without infixation of the diminutive vowel template (Grandi 2002: 226). The relation between feminine and diminutive reference seems to have spread to some nouns of Romance origin. This is shown in (6.31).

(6.31) Gender shift as a strategy for diminutive formation with nouns of Romance origin in Maltese (Semitic) (adapted from Grandi 2002: 230)

(a) *forno*
oven.M
'oven'

(b) *forn-a*
oven-DIM.F
'small oven'

In (6.31), the diminutive of the borrowed noun *forno* 'oven' is formed by changing the gender of the base. More commonly, nouns of Romance origin are diminutivized by means of the diminutive suffix *-in*, also borrowed from the Romance contact languages. This is shown in example (6.32).

[48] In his typological survey of evaluative constructions in the languages of the Mediterranean area, Grandi (2002: 218-248) notices that evaluative morphology is generally not very productive in Standard Arabic and Hebrew. Diminutives and augmentatives are more frequently used in nonstandard varieties of Arabic.

(6.32) Romance diminutives in Maltese (Semitic) (adapted from Grandi 2002: 229)

(a) *biskott*
biscuit.M
'biscuit'

(b) *biskutt-in*
biscuit.M-DIM
'small biscuit'

There are no morphological augmentatives of Semitic origin in Maltese. The augmentative suffix *-un* has been borrowed from the Romance contact languages and is productively used with Semitic nouns as well (Grandi 2002: 229-231). Gender shift does not play any role in the formation of augmentatives in Maltese.

To summarise, given that size-related gender shifts are attested throughout all the Afro-Asiatic genealogical subgroupings included in the sample, there is evidence to claim that they are an inherited feature of nominal morphology across the language family, rather than an independent innovation of the individual subgroupings.

Gender shifts may not be the only strategy available in a Type 2 language to encode evaluation. In a handful of languages of my sample, evaluative markers independent of gender coexist with size-related gender shifts. This is the case of Male (Ta-Ne-Omotic), Nama (Khoe-Kwadi) and, as described earlier in this section, Maltese. In Male, gender marking with inanimate nouns always implies size. In addition, the diminutive suffix *-ómma* is found (Amha 2001: 70-71). In Nama, as previously mentioned, there exist a diminutive and an augmentative suffix, *-ró* and *-kra*. Interestingly, the diminutive suffix works independently of gender shifts, whereas the augmentative is often used in combination with gender shifts to refer to extra-large size (Hagman 1977: 27).

Not much can be said about the diachrony and stability of the phenomenon of size-related gender shifts. What is known is that they tend to occur in languages in which gender assignment is not rigid. In such languages nouns can be assigned to different genders on the basis of the context in which they occur. The possibility of manipulating gender assignment is, however, conditioned by certain semantic properties of nouns. On a general basis, nouns that denote human beings have a tighter relation with their default gender.

6.4.3 Type 2: summary

In this section, I discussed languages with sex-based gender in which gender assignment is not rigid and gender shifts are used to encode diminutive and augmentative meanings (Type 2 in the classification proposed in §6.2). In §6.4.1, I surveyed the general characteristics of languages of Type 2 and discussed examples from various languages within Type 2. The distribution of Type 2 languages within my sample has been discussed in §6.4.2, where I showed that this type of interaction between gender and evaluative morphology has a wide diffusion across different genealogical groupings of Africa.

6.5 Evaluative morphology and biological gender in languages without (sex-based) gender

In this section, I describe a pattern that I found attested in only 8 of the 100 languages of my sample. This pattern is concerned with the interaction between the linguistic encoding of biological gender (i.e., strategies to characterise animate entities as females or males) and evaluative morphology. Although it is not entirely relevant for the general typological picture of the relationship between grammatical gender and evaluative morphology in the languages of Africa, this pattern deserves some discussion. The pattern was found in languages with non-sex-based gender and in languages without grammatical gender, that is, in languages where the encoding of biological gender is outside the gender system, or where there is no gender system at all. Five of these nine languages belong to the Bantu group and are the languages discussed in §6.3.3 with respect to evaluative morphology renewal; they have non-sex-based gender: Northern Sotho, Swati, Tswana, Venda and Zulu. The remaining three languages are Luwo, Acoli and Akan. Luwo and Acoli belong to the Western Nilotic genealogical unit. Akan is an Atlantic-Congo language of the Kwa subgroup. None of the three languages mentioned last in this list has grammatical gender.

6.5.1 Bantu

Bantu languages do not encode biological gender within the noun class system. The most common strategies for conveying information about biological gender are suppletion or constructions of the types *female cattle* for 'cow' (Güldemann 1999: 59). Nonetheless, in some of the southeastern Bantu languages, morphological strategies for the overt coding of female and male reference exist and are somewhat related to the domain of evaluation in two ways:

1. The diminutive suffix derived from **jana* 'child' can be used as a marker of female reference in very specific contexts, mostly with animal names or with colour adjectives modifying animal names.

2. The suffix resulting from the grammaticalization of **kádi* 'woman' becomes productive for the derivation of feminine nouns. As shown in 6.3.3, in fewer languages within the southeastern Bantu area, this suffix is also used as augmentative.

Selected examples from the above-mentioned six languages will now be discussed to illustrate how these usages developed diachronically.

In Sotho, the suffix *-hadi* is used to encode female reference when added to nouns denoting animals but functions as an augmentative when used with nouns denoting human beings (Creissels 1999). Table 6.5 illustrates how the meaning of the suffix varies depending on the meaning of the base noun.

Poulos & Louwrens (1994) present a different set of facts with respect to the Northern Sotho variety. Both *-ana* (the Diminutive) and *-gadi* (the Augmentative-Feminine) are said to be used as feminine suffixes. However, differently from what reported by Creissels

Table 6.5: The meanings and function of *-hadi* in Sotho (based on Creissels 1999: 32)

Base noun		**Suffix**	**Suffix meaning**	**Resulting form**
pere,'horse'	+	*-hadi*	F	*pere-hadi*, 'she-horse'
monna, 'man'	+	*-hadi*	AUG	*monna-hadi*, 'big man'

(1999), according to Poulos & Louwrens (1994: 68-69), the suffix *-gadi* is no longer used as an augmentative marker in Northern Sotho, and its use as a feminine marker is also very rare. The diminutive suffix, *-ana*, on the other hand, can express feminine meaning but only when used with animal nouns or adjectives denoting colour. The same pattern is found in Tswana, geographically very close to Sotho.

Grammatical descriptions of Tswana (see, for instance, Cole 1955: 141-144) report that, when colour adjectives modify nouns denoting domestic animals, the diminutive suffix is always used to mark female reference on the adjective. When the adjective is not marked by the diminutive suffix, the combination "Noun+ adjective" is interpreted as having male reference. Creissels (1999: p.34) provides an interesting example in which such a construction is attested being used to denote a human entity rather than an animal noun. In Creissels' example the colour adjective for 'yellow' occurs with the word for 'woman,' *mosadi*, and is marked by the diminutive suffix *-ana*.

(6.33) Tswana (Bantu) (Creissels 1999: 34)

(a) *monna* *yomo-setlha*
CL1.man CL1-yellow
'a man with clear skin'

(b) *mosadi* *yomo-setlh-ana*
CL1.woman CL1-yellow-F
'a woman with clear skin'

The use of *-ana* with adjectives modifying human nouns is not mentioned in grammatical descriptions of Tswana and is described as fairly atypical by Tore Janson (personal communication). What is relevant about the example in (6.33b) is that two different indexation patterns appear on the adjective. The adjective indexes the gender of the noun by means of Class 1 prefix. In addition, the occurrence of the suffix *-ana* on the adjective can be read as some sort of referential indexation pattern. The suffix signals that the NP referent is female. This usage can be interpreted as a step forward in the grammaticalization of the diminutive/feminine suffix. The phenomenon is also attested in Zulu, both with the diminutive and the feminine/augmentative suffix (Güldemann 1999: 74).

The suffixes *-gadi* and *adi* in Tswana have a meaning "associated with the feminine, the marriage institution or the idea of the opposite sex" (Cole 1955: p.110). The two suffixes are rarely used to mark biological gender with animal nouns (e.g., *tau* 'lion,'

taugadi 'she-lion'). The suffix is more productive in the neighbouring (and closely related) languages Southern Sotho and Zulu. Similar patterns are also attested in Swati and Venda.

To summarise, the overt coding of female reference in the southeastern Bantu languages is the result of two recent developments, which do not seem to have reached the same stage of grammaticalization in all the languages where they appear:

(1) *diminutive* > *feminine*

(2) *feminine* > *augmentative*

The rise of markers for the encoding of female reference follows the development of the new diminutive morpheme in the case of suffixes derived from **jana*, 'child' (1). On the other hand, in the case of the suffixes derived from **kádi*, 'woman', the most plausible pattern is: from marker of female reference to augmentative (2). The two patterns reveal a substantial interaction between evaluative morphology and the marking of biological gender.

As shown in the present section, the overt coding of female reference is often limited to a narrow range of occurrences, such as gender specification with animal nouns or, as in Tswana, colour adjectives used to differentiate different kinds of animals, mostly cows (Tore Janson, personal communication). Thus, the use of these suffixes as markers of biological gender is limited to very salient domains of experience for societies whose economic subsistence is very often based on hunting, livestock farming and selling. As already mentioned in §6.3.3, the development of diminutive and augmentative/feminine suffixes in the southeastern Bantu languages is explained by Güldemann (1999) as the result of contact with Khoe and non-Khoe languages spoken in southern Africa. Livestock breeding and hunting play a crucial role in the economy of many (Bantu and non-Bantu) societies of southern Africa:

> this conceptual orientation towards animals suggests that such closely associated features as sex and size are prone to be expressed linguistically (Güldemann 1999: 71).

Regular and transparent encodings of such notions is thus quite likely to be borrowed cross-culturally across languages.

6.5.2 Western Nilotic

In Western Nilotic languages, there exists an inventory of morphological strategies for the categorisation of nouns in semantically homogeneous groups (see discussion in §5.6).

In this section, I examine a few cases of interactions between the word-formation strategies that are used in these languages for the encoding of biological gender, on the one hand, and size, on the other. The languages considered are Acoli and Luwo.

In Luwo, the diminutive marker is the prefix *ɲì-*, which literally means 'daughter of.' Interestingly, the prefix is also found in ethnonyms where it expresses female reference. The two possible uses of the prefixes are illustrated in examples (6.34) and (6.35).

(6.34) Diminutive formation in Luwo (Western Nilotic) (Storch 2005: 295)

ɲì-bòò
DIM-banana
'small quantity of banana'

(6.35) Feminine ethnonyms in Luwo (Western Nilotic) (Storch 2005: 295)

ɲì-jáŋ
F-dinka
'Dinka woman'

When used as diminutive, the prefix tends to express only small quantities as in (6.34). Diminution in size is constructed by combining the use of the diminutive prefix with the adjective *mát̪tíi̪n̪* 'small,' as in example (6.36).

(6.36) Luwo (Western Nilotic) (Storch 2005: 295)

ɲì-t̪íi̪n̪ *mát̪tíi̪n̪*
DIM-child small
'small child'

Similarly to Luwo, in Acoli the diminutive prefix, *ɲ-*, is etymologically related to the noun for 'child, daughter of.' Augmentatives are constructed by juxtaposing *míin*, 'mother/female of,' to the noun to be augmented (Storch 2005: 358).

6.5.3 Akan

The Kwa language Akan does not have grammatical gender. The diminutive marker of Akan is the suffix *-ba/-wa* – depending on the dialect – which, as described by Appah & Amfo (2011), originates from the noun *ɔba*, 'child' (in the sense of offspring). Appah & Amfo (2011) discuss the connection between the use of the diminutive suffix and the encoding of female reference. In Akan, many common and proper names denoting women are marked by the suffix *-ba/-wa*. Three nouns of this type are listed in (6.37).

(6.37) Diminutive marking and female reference in Akan (Kwa) (Appah & Amfo 2011)

- *abaayewa* 'young woman'
- *aborɔwa* 'European woman'
- *Egyrba* female name (the correspondent male name is *Egyr*)

With the exception of the first noun in the list, the semantics of these nouns is not related to diminution in any evident way. The suffix *-ba/-wa* only encodes female reference. Interestingly, in the Asante and Akuapem dialects of Akan, female names can also be formed by using the suffixes *-waa* and *-bea*, which are cognates derived from the noun *ɔbaa/ɔbea* 'female' (Appah & Amfo 2011: 91). Appah & Amfo (2011: 91) propose that the diminutive and the feminine suffixes derive from two different but morphophonologically very similar nouns – *ɔba* 'child' and *ɔbaa/ɔbea* 'female,' respectively – and are used interchangeably by speakers of Akan to derive female names from the correspondent male nouns. No such connection between diminutive and female reference is found in the two other Kwa languages of the sample, Ewe and Sɛlɛɛ.

6.5.4 Summary

In this section, I surveyed interactions between the encoding of biological gender and evaluative markers in eight languages with non-sex-based gender or no gender. In these languages, the encoding of female reference shows associations with both diminutive markers originated from the word for 'child' and augmentative markers originated from the words 'woman' or 'daughter.' These interactions are very rare in my language sample, but correspond to a rather frequent pattern worldwide (Bauer 2002; Jurafsky 1996; Matisoff 1992).

6.6 Absence of interaction between gender and evaluative morphology

As shown in table 6.1, the marking of diminutives and augmentatives shows no interaction with the gender domain in 13 of the 100 languages of the sample. The evaluative constructions of these languages are not surveyed in detail here. However, a few examples are discussed as an illustration of the types of evaluative markers that are attested in such languages (see also §6.3.3 for a discussion of evaluative morphology in the Bantu languages Bafia and Eton).

In ǁAni (Khoe-Kwadi), Ju|'hoan (Kxa) and !Xóô (Tuu), only diminutives are found. These are suffixal and all etymologically related to the word for 'child.'

In Kambaata (Cushitic), the encoding of diminutive and augmentative meanings does not interact with gender marking but rather with number. In certain contexts, the singular and plural suffixes can be used as diminutives and augmentatives/pejoratives, respectively.

(6.38) Singular marking as diminutive in Kambaata (Cushitic) (Treis 2011: 6)

mat-íta	*ann-uhúu*	*ám-atii*	*yoo-ba'í*
one-F.ACC	father-M.NOM.CRD	mother-F.NOM.CRD	COP.3-NEG.REL
wotar-ch-úta	*aass-íi*	*iitt-an-tóo'u*	
foal-SG-F.ACC	give-M.DAT	decide-PASS-3FPFV	

'It was decided to give [them] a [tiny] donkey foal which had no father and mother'

(6.39) Plural marking as augmentative/pejorative in Kambaata (Cushitic) (Treis 2008: 144)

hor-i-ssá	*inq-áakk-ant*	*hínn*	*y-itóo'u*
all-M.GEN-3.PL.POSS	tooth-PL-F.NOM<N>	smirk	say-3.F.PFV

'[As they said, their upper lips were lifted up and] the teeth of all of them smirked'

The use of the plural suffix in (6.39) can be interpreted as either signalling the big size of the donkey's teeth or "as indicating that the teeth are an object of ridicule" (Treis 2008: 144).

6.7 Summary of the chapter

The purpose of this chapter has been to survey patterns of interaction between gender and evaluative morphology in the languages of the sample. In particular, the chapter has aimed at examining (1) the frequency of occurrence of *size* as an independent gender value and (2) how different types of gender systems (*sex-based* vs. *non-sex-based*) and strategies of gender assignment (*rigid* vs. *manipulable* gender assignment) condition the attested patterns of interactions between gender and evaluative morphology. The research questions addressed in the chapter are listed in (6.1).

In general, the data presented in this chapter show that grammatical gender and the morphological encoding of evaluation interact in relevant ways in the languages of Africa. This is in line with what has been pointed out by previous investigations such as Allan (1977) (see also §2.6).

The occurrence of *size* as an independent gender value was found in 23 of the 84 languages with gender in my sample. The distribution of the phenomenon is restricted to the following genealogical groups: North-Central Atlantic, Bantu and Kwa. These languages are characterised by non-sex-based gender systems and a large number of gender distinctions (see §6.3). The most frequent type of interaction between gender and evaluative morphology, which was found in 28 of the 84 languages with gender, does not entail the presence of dedicated diminutive and augmentative genders. In languages with sex-based gender systems, gender shifts from masculine/feminine to neuter or from masculine to feminine (and *vice versa*) can be used to encode diminutive and augmentative meanings (see §6.4).

The data discussed in the chapter thus show that the nature of the interactions between gender and evaluative morphology varies according to the type of gender system that a language has. The relevant variables are: type of gender system (sex-based vs. non-sex-based gender), number of distinctions within the system and degree of manipulability of gender assignment. The data also confirm that *size* can be a productive criterion for

the classification of nouns but is never centrally prominent, neither synchronically nor diachronically, in the two types of gender systems that are found among the languages of the sample.

In languages with large gender systems, evaluative meanings are (1) part of the semantic potential of very polysemous classes, and/or (2) associated with noun classes that are only used to form diminutives and augmentatives. The latter type of evaluative classes may exhibit peculiar morphosyntactic properties. For instance, in some Bantu languages, they can be added to the default class marker of a noun (see §6.3.2). From a diachronic point of view, the evaluative classes of the type described in (2) tend to be less stable (see §6.3.3).

As shown in §6.4, in languages with sex-based gender, size-related gender shifts are almost always restricted to inanimate entities: the polarity between masculine and feminine genders is variously exploited to encode polar notions within the domain of size ('big' vs. 'small') . The only exception is Masai, where shifts from the Masculine to the Feminine Gender also involve animate nouns. In Turkana and Karamojong, which have a tripartite gender system, the Neuter Gender is used to encode young age of animate entities and small size of inanimate.

Animacy-driven restrictions on the encoding of evaluative meanings by means of gender shifts are only attested in languages with sex-based gender, where sex and size-related gender shifts are realised by the same markers. In the case of animate nouns, speakers seems to conceive sexual dimorphism as the most salient semantic property and any size-related gender shift is excluded. With inanimate nouns, the relevance of sexual dimorphism is instead ruled out. One could say that sex and size-related gender shifts are competing types of semantically driven gender assignment, insofar as they are both determined by the semantic properties of nouns and/or the cultural representation of the NP referent.

It is noteworthy that diminutive and augmentative markers that are *not* related to sex distinctions show rather strong preferences for highly animate referents. For instance, previous studies on the diachronic development of diminutives (see, among others, Grandi 2002, 2011;Jurafsky 1996) show that the first function of diminutive constructions that originates from the word for 'child' is to encode young age with nouns denoting humans or higher animals. Only at the second stage of grammaticalization do these constructions start marking small size with inanimate entities. Similarly, a very common source for the grammaticalization of augmentative markers are the animate nouns 'mother' and 'woman' (Matisoff 1992), and a diachronic connection between augmentative markers and markers of animacy has also been noted (Grandi 2002, 2011).

7 Gender and grammatical complexity

7.1 Introduction

The aim of this chapter is to propose a complexity metric for gender that can account for the interactions between gender and number and gender and evaluative morphology investigated in this thesis. The research question addressed in the chapter is:

(7.1) Research question concerning grammatical complexity (same as (2.15))

Q 9: How do interactions between gender and number and gender and evaluative morphology affect the grammatical complexity of gender systems? Is it possible to measure the role that these interactions play in the absolute complexity of individual gender systems?

The chapter is structured as follows. In §7.2, I provide some background on the approach to grammatical complexity adopted in the dissertation and discuss the dimensions of gender complexity suggested by Audring (2014). In §7.3, I discuss how the interactions between gender and number and gender and evaluative morphology considered in this dissertation can be added to Audring's criteria. In §7.4, I propose a method for maximally local complexity measures. The complexity metric used in this thesis is presented in §7.5, whereas §7.6 describes the method followed to calculate the complexity of the gender systems attested in the language sample. Results and discussion of the results are found in §7.7, whereas a summary of the chapter is provided in §7.8.

7.2 Background

The notion of grammatical complexity has been introduced in §2.7 and discussed in §5.4.1 in connection with my results on patterns of exponence and syncretism of gender and number. As pointed out on those occasions, I interpret grammatical complexity in its absolute sense, that is, as an objective property of a grammatical domain. Absolute complexity can be measured either based on the number of parts that make up a system, or based on the length of the description of that system. In addition, measures of absolute complexity can only be *local*, in the sense that they should address individual grammatical domains rather than languages in their entirety. As also discussed in §2.7, Miestamo (2008) and Sinnemäki (2011) suggest two major principles as general guidelines for elaborating complexity metrics for any grammatical domain: the Principle of Fewer Distinctions and the Principle of One-Meaning–One-Form. The two principles are repeated below in a simplified form:

The Principle of Fewer Distinctions. The higher the number of grammaticalized distinctions, the more complex the domain.

The Principle of One-Meaning–One-Form. The least complex domain is the one where there is a one-to-one mapping between meaning and form.

In §7.3, I argue that the two principles do not suffice to account for the whole range of phenomena that can be relevant to the study of language complexity. By taking gender as a case in point, I suggest that a third principle can be added. In addition, in §7.4, I suggest how the three principles can be further specified in order to allow for maximally local complexity measures.

In her pioneering paper on gender as a complex feature, Audring (2014) proposes three criteria to assess the grammatical complexity of gender systems throughout the world's languages:

I The number of gender values

II The number, nature and scope of assignment rules

III The amount of formal marking (i.e., how much gender indexation there is in a language).

The three dimensions are discussed by Audring by using a set of examples from different types of gender systems attested in different areas of the world (see §2.7 for a detailed description of the methods and aims of Audring's work on gender as a complex feature).

Audring's dimensions of gender complexity provide us with a baseline to calculate the complexity of gender systems along the three most characteristic properties of this grammatical domain: classification, assignment and indexation. Interactions of gender and other domains of grammar are not in the scope of Audring's investigation. In §7.3, I suggest a way to include the patterns of interactions investigated in this dissertation into a metric for gender complexity.

7.3 How to integrate interactions of gender into a metric for gender complexity

In this section, I suggest that the absolute complexity of a grammatical domain is affected by the number of interactions that it has with other domains in the following way: the more interactive a grammatical domain, the higher its absolute complexity. In addition, I suggest that this aspect of the absolute complexity of grammatical domains cannot be handled by the Principle of Fewer Distinctions nor by the Principle of One-Meaning–One-Form and rather calls for the introduction of a third principle.

The *interactivity* of a grammatical domain cannot be accounted for by the Principle of Fewer Distinctions because it is not strictly concerned with the inventory of grammaticalized distinctions within an individual domain. Similarly, it cannot be accounted

for by the Principle of One-Meaning–One-Form because it is not strictly concerned with the mapping between meaning and form. The interactivity of a grammatical domain rather has to do with how *isolated* or *independent* a domain is from other domains. Its relevance to language complexity can be assessed as follows:

> The least complex grammatical domain is the one whose description is *independent* of semantic and functional properties of other domains. The more interactive the domain, the less independent. The less independent the domain, the higher its absolute complexity.

I propose to refer to this principle as the *Principle of Independence*.

A short overview of the types of morphosyntactic and semantic interactions of gender and number and gender and evaluative morphology explored so far, as well as their role in the complexity of gender systems, is in place.

Concerning morphosyntactic interactions, in chapter 5, I examined morphosyntactic interactions of gender and number by looking at exponence and syncretism. Even there, I argued that cumulative exponence features higher grammatical complexity because it makes the mapping between meaning and form less straightforward. Similarly, in chapter 6, I discussed morphosyntactic interactions of gender and evaluative morphology in the languages where diminutives and augmentatives are part of the inventory of gender distinctions. In principle, the presence of diminutive and augmentative genders can also be seen as increasing the absolute complexity of gender systems. This is discussed in detail in §7.5.

Concerning semantic interactions, the following phenomena were counted as instances of semantic interaction between gender and number (§5.7), and gender and evaluative morphology (chapter 6) throughout the thesis:

(1) Manipulations of gender assignment to express variation in the countability properties/number values of a noun (see §5.7).

(2) Manipulations of gender assignment to express variation in the size of the referent or the speaker's attitude towards the referent (see chapter 6).

The role that the two types of interactions play in the absolute complexity of gender systems has not been discussed so far. As argued throughout this dissertation already, gender assignment can be rigid (i.e., nonmanipulable) or manipulable. Manipulable gender assignment is found in those languages in which nouns can be shifted from one gender to another in order to encode a different construal of the NP referent according to patterns of the type described in (1) and (2). The possibility of manipulating gender assignment may be seen as piling on top of the default gender assignment rules that are productive in a language. In languages with manipulable gender assignment, gender markers have default and add-on meanings. These add-on meanings are dependent on semantic associations between gender and other grammatical domains, notably countability and evaluation. Thus, based on the *Principle of Independence* introduced above, their presence features an increase in the absolute complexity of gender. Gender assignment is not only given in the lexicon for each and every noun, but it is also subject to change *depending on* associations with other functional domains.

An alternative analysis would be to interpret gender systems with manipulable assignment as being semantically more predictable and thus less complex than gender systems with rigid assignment. Nouns are assigned to different genders depending on the way the NP referent is construed (e.g., *big* vs. *small* or *countable* vs. *uncountable*), that is, based on pragmatic grounds. An analysis of this type, however, would not account for the fact that default gender assignment in languages with manipulable gender assignment is often semantically arbitrary. For instance, as shown in chapter 6, in the Berber languages, inanimate nouns can be both masculine and feminine (see e.g., Tachawit where *aqnmuš* 'pot' is masculine and *taɣnžakt* 'spoon' is feminine). Only in case of gender shift is gender assignment manipulated according to semantically transparent principles of the type:

(1) Masculine Inanimate → Feminine = Diminutive

(2) Feminine Inanimate → Masculine = Augmentative

To summarise, manipulable gender assignment is an add-on device, which acts on the default mechanisms of gender assignment according to semantically predictable patterns that are dependent on the interactions of gender with other domains. The presence of manipulable assignment increases the absolute complexity of gender systems.

7.4 How to elaborate maximally local complexity measures

As observed in §§2.7 and 7.2, Miestamo (2008) and Sinnemäki (2011) suggest that language complexity can only be measured on a local scale, by focussing on individual areas of grammar. This locality constraint lies on two fundamental problems related to the assessment of language complexity: *representativity* and *comparability* (see discussion in Miestamo 2008). On the one hand, it is practically impossible to elaborate complexity measures that can be fully representative of the grammar of a language (representativity). On the other hand, it is extremely difficult to make sure that the individual grammatical phenomena within a complexity metric can be compared with each other in such a way that their contribution to the global complexity of a language can be assessed (comparability).

Following a suggestion by Dahl (2011: 156), in this section I propose that, in order to guarantee maximal locality, the formulation of the three principles of absolute complexity – the Principle of Fewer Distinctions, the Principle of One-Meaning–One-Form and the Principle of Independence – should be based on *ceteris paribus comparisons*. In other words, the maximally local way to measure absolute complexity is by means of comparisons of the type: "Everything else being equal, X is more complex than Y" or "Everything else being equal, the addition of Z increases the complexity of X" (Östen Dahl, personal communication). Only by introducing the "Everything else being equal" condition can we be sure that the individual features of a complexity metric are truly local in their scope.

When reinterpreted in terms of *ceteris paribus* comparisons, the Principle of Fewer Distinctions, the Principle of One-Meaning–One-Form and the Principle of Independence can be reformulated as described in (7.2).

(7.2) The principles of absolute complexity reformulated in terms of *ceteris paribus* comparisons

Principle of Fewer Distinctions

> Everything else being equal, a grammatical domain with n distinctions is less complex than one with *n+1* distinctions.

Principle of One-Meaning–One-Form

> Everything else being equal, a grammatical entity with n forms is less complex than one with *n+1* forms.
>
> Everything else being equal, a grammatical entity with n meanings is less complex than one with *n+1* meanings.

Principle of Independence

> Everything else being equal, a grammatical domain that is *in*dependent of semantic and functional properties of other domains, is less complex than a grammatical domain that is dependent on n or $n + 1$ semantic and functional properties of other grammatical domains.

In the rest of the chapter, the three principles are operationalised in order to elaborate a complexity metric for gender systems. This is spelled out in §7.5.

7.5 A metric for grammatical gender complexity

As mentioned in §7.2, Audring (2014) formulates three dimensions along which to assess the absolute complexity of gender: *classification* (number of distinctions/classes), *assignment* (number, nature and scope of assignment rules) and *indexation* (number of indexing targets). Audring's gender complexity dimensions are all included in my metric. Based on what was discussed in §7.3, in order to account for morphosyntactic and semantic interactions between gender and number and gender and evaluative morphology, two dimensions were added to my metric: *morphosyntactic interactions* and *semantic interactions* (see §7.2).
The following features are proposed for each of the two dimensions:

- Morphosyntactic interactions
  - Cumulative exponence of gender and number
- Semantic interactions
  - Manipulation of gender assignment triggered by number/countability

– Manipulation of gender assignment triggered by size.

My suggestion is that the presence of each of these phenomena increases the absolute complexity of a gender system.

One additional adjustment of Audring's (2014) complexity criteria concerns assignment rules. According to Audring, gender assignment can be evaluated by looking at (1) types (or nature) of assignment rules and (2) scope of assignment rules: (1) covers whether gender assignment is semantic or is both semantic and formal; (2) measures how general assignment rules are, that is, the number of nouns whose gender assignment each rule is able to predict. Since it is rather difficult to evaluate the scope of gender assignment rules based mostly on descriptive sources, I decided to restrict Audring's dimension II (assignment) to only the *the nature of assignment rules*.

In sum, the complexity metric for grammatical gender that I use in this dissertation is an expanded, and slightly revised, version of the model developed by Audring (2014). Table 7.1 provides a list of the features that are included in my metric and compares them with those used by Audring (2014). The rightmost column of the table describes how each feature is expected to affect the complexity of grammatical gender. The feature values and their numerical interpretation are discussed in §7.6. Since the metric aims at capturing how minimally and maximally complex gender systems can be, absence of gender is not featured by any of the features in the metric. It follows that the metric can only be applied to languages with gender.

It is worth mentioning that the existence of dedicated diminutive and augmentative genders may be considered a type of morphosyntactic interaction between gender and evaluative morphology. This aspect of the interaction between gender and evaluative morphology is not counted as an independent feature of the metric because it is somewhat redundant. As shown in chapter 6, languages with diminutive and augmentative genders are languages with large gender systems where gender assignment can be manipulated to encode size variation. Both properties are already accounted for by two other features of the metric: (1) number of gender values (GV) and (2) manipulation of gender assignment triggered by size (M2).

Table 7.1: Cues for assessing grammatical complexity of gender systems

Feature ID	My features	Audring's (2014) features	Description
GV	Number of gender values	Number of gender values	Everything else being equal, a gender system with two distinctions is less complex than a gender system with three or more distinctions (Principle of Fewer Distinctions).
AR	Nature of assignment rules	Number and nature of assignment rules	Everything else being equal, a gender system with one type of assignment rules – e.g., only semantic or only formal – is less complex than a gender system with two types of assignment rules – both semantic and formal. (Principle of Fewer Distinctions).[49]
IND	Number of indexing targets	Amount of formal marking	Everything else being equal, a gender system with one indexing target is less complex than a gender system with two or more indexing targets (Principle of Fewer Distinctions).
CUM	Cumulative exponence of gender and number	–	Everything else being equal, a marker that only signals gender is less complex than a marker which signals gender + number (Principle of One-Meaning–One-Form and Principle of Independence).
M1	Manipulation of gender assignment triggered by number/countability	–	Everything else being equal, a gender system where gender assignment is only lexically given is less complex than a gender system where gender assignment is given in the lexicon + can be manipulated depending on the countability properties of the noun (Principle of Independence) or the NP.
M2	Manipulation of gender assignment triggered by size	–	Everything else being equal, a gender system where gender assignment is only lexically given is less complex than a gender systems where gender assignment is given in the lexicon + can be manipulated depending on the size of the NP referent (Principle of Independence).

[49] As mentioned before, gender systems with only semantic assignment rules are quite common cross-linguistically, whereas gender systems with only formal assignment rules are almost never encountered.

7.6 Method

Having defined the features (or cues) for measuring the absolute complexity of grammatical gender (see table 7.1), the next step is to establish the values associated with each feature and to convert them into numbers. This allows us to compute the absolute complexity of the gender system of each of the languages with gender in the sample. To achieve this purpose, I follow Parkvall (2008) who designed a method for computing the language complexity of creoles and noncreole languages on the basis of a set of features taken from the WALS database (Dryer & Haspelmath 2013). The values of each feature are assigned a number between 0 and 1. Features with three values are converted into the numerical format 0, 0.5, 1. Similarly, features with five values are converted by Parkvall into the format 0, 0.25, 0.5, 0.75, 1. For all the features taken into account in Parkvall's paper, 0 stands for *minimally complex* and 1 for *maximally complex*. The total complexity score for each language is divided by the number of features included for that language. This is done in order to allow languages for which less information is available on a given feature to get average scores comparable to those of the best documented languages. The same procedure is followed in this thesis. In addition, values have been set for each of the features in the metric. The feature values and correspondent numbers are illustrated in table 7.2.

The composition of the metric is thus such that the least complex possible gender system is the one that scores zero with respect to all the features of the metric and exhibit the following properties: two gender values, semantic gender assignment, one indexing target, no cumulation with number, no manipulation of gender assignment triggered by number/countability and no manipulation of gender assignment triggered by size. On the other hand, the most complex possible gender system is the one that scores 1 with respect to all the parameters considered in the metric and exhibit the following properties: five or more genders, semantic and formal assignment, four or more indexing targets, cumulation with number, and manipulation of gender assignment triggered by both number/countability and size.

It is worth mentioning that the choice of the feature values assigned to IND relied on mere (and potentially problematic) convenience choices. First, the metric accounts for the number of indexing targets but does not allow us to measure whether different indexing targets affect the complexity of gender in different ways. By using the feature IND, a rough count of how pervasive gender indexation is in a language is obtained but it is not possible to verify whether, for instance, "one indexing target" means "only personal pronouns" or "only adjectives", and whether this difference has relevant consequences for the overall complexity of gender. Second, indexes are identified on the basis of distinguishable functions. Two functionally different indexes (e.g., determiners and demonstrative pronouns) can have the same formal realisation in one language. However, the metric cannot account for the implications of these patterns of identity of forms on the complexity of individual gender systems. These are important questions for the understanding of how gender systems work, but, unfortunately, they cannot be explicitly addressed by the metric used in this dissertation.

Table 7.2: Gender complexity metric

Feature	**Feature value**	**Score**
Number of gender values (GV)	Two genders	0
	Three	1/3
	Four	2/3
	Five or more	1
Nature of assignment rules (AR)	Semantic assignment	0
	Semantic and formal assignment	1
Number of indexing targets (IND)	One	0
	Two	1/3
	Three	2/3
	Four or more	1
Cumulative exponence of gender and number (CUM)	Noncumulative	0
	Partially cumulative	1/2
	Cumulative	1
Manipulation of gender assignment triggered by number/countability (M1)	Absent	0
	Present	1
Manipulation of gender assignment triggered by size (M2)	Absent	0
	Present	1

Finally, as mentioned in §7.4, a general issue to be addressed is what Miestamo (2008) refers to as *the problem of comparability*: to what extent are the individual features of a complexity metric comparable with each other? Do the individual features in a metric contribute in the same way to the final complexity score? The two questions are addressed in detail in §7.7.4. Based on the individual complexity scores of the languages of my sample, I compare the features of the metric by looking at the way they correlate with each other and by discussing whether any of these features can be considered as the best predictor of the final score.

Before presenting the results of my calculations, it is worth mentioning that, in case of missing features, the index values resulting from the calculations should be taken with caution. In fact, even though average scores (rather than total scores) are used as index values, the index values of languages with missing features cannot be regarded as entirely comparable to the index values of languages for which all features are equally documented.

7.7 Results and discussion

The grammatical complexity of the gender systems attested in the 84 gendered languages of the sample has been calculated following the method described in §7.6. The results are presented in table 7.3. The table is divided in two macro-columns and languages are arranged from the most to the least complex. The leftmost columns of each macro-column provide the rank: languages with the same average complexity score share the same rank. Next to the rank come the language names and their ISO code; the average complexity score assigned to the gender system of each language is given in the rightmost columns of the two macro-columns. I refer to the score as the *Gender Complexity Score* (henceforth GCS). The complexity scores for each of the feature values in the metric, as well the GCSs, are given in appendix F.

The following observations can be made on the basis of the results presented in table 7.3:

1. Languages from the same genealogical units, or spoken within the same areas, tend to have similar or even identical GCSs. In many cases, areal pressure seems to be a relevant factor in explaining the distribution of the outliers.

2. If the languages with the highest GCS (= 1) are excluded, languages may display the same index value but arrive to it on different paths. In other words, identical GCSs do not stand for *same type of gender system*. In fact, identical GCS may result from different combinations of values for each feature in the metric.

3. Implicational relations seem to exist between some of the features in the metric.

4. Some features in the metric correlate more with each other and seem to have a stronger impact on the GCS than others.

These four points will be discussed in §§7.7.1, 7.7.2, 7.7.3 and 7.7.4, respectively.

Table 7.3: GCSc of the languages of the sample

Rank	Language	Isocode	GCS	Rank	Language	Isocode	GCS
1	Bandial	bqj	1	9	Bench	bcq	0.61
1	Bemba	bem	1	9	Hadza	hts	0.61
1	Bidyogo	bjg	1	9	Hausa	hau	0.61
1	Chiga	cgg	1	9	Moroccan Arabic	ary	0.61
1	Kagulu	kki	1	9	Nama	naq	0.61
1	Kikuyu	kik	1	9	Naro	nhr	0.61
1	Lega	lea	1	9	Sandawe	sad	0.61
1	Maasina Fulfulde	ffm	1	9	Standard Arabic	arb	0.61
1	Mongo-Nkundu	lol	1	9	Tigre	tig	0.61
1	Makaa	mcp	1	10	Miya	mkf	0.6
1	Ndengereko	ndg	1	11	Awngi	awn	0.56
1	Nyanja	nya	1	11	Lingala, Kinshasa	lin	0.56
1	Shona	sna	1	11	Male	mdy	0.56
1	Serer	srr	1	11	Wolaytta	wal	0.56
1	Swahili	swa	1	12	Borana-Arsi-Guji Oromo	gax	0.53
1	Timne	tem	1	12	Koorete	kqy	0.53
1	Tonga	toi	1	12	Lishan Didan	trg	0.53
1	Venda	ven	1	12	Qimant	ahg	0.53
1	!Xóô	nmn	1	12	ǁAni	hnh	0.53
2	Tunen	baz	0.94	13	Beja	bej	0.5
3	Eton	eto	0.83	13	Gidar	gid	0.5
3	Maltese	mlt	0.83	13	Masai	mas	0.5
3	Northern Sotho	nso	0.83	14	Daasanach	dsh	0.47
3	Sɛlɛɛ	snw	0.83	14	Dirasha	gdl	0.47
3	Swati	ssw	0.83	14	Kxoe	xuu	0.47
3	Tswana	tsn	0.83	14	Lele	lln	0.47
3	Turkana	tuv	0.83	14	Rendille	rel	0.47
3	Wamey	cou	0.83	15	Dizin	mdx	0.44
3	Zulu	zul	0.83	15	Hebrew	heb	0.44
4	Bafia	ksf	0.78	15	Tsamai	tsb	0.44
4	Dibole	bvx	0.78	16	Iraqw	irk	0.43
4	Kisi	kss	0.78	16	Somali	som	0.43
4	Nuclear Wolof	wol	0.78	17	Dime	dim	0.39
5	Tachawit	shy	0.75	18	Baiso	bsw	0.36
6	Karamojong	kdj	0.72	18	Ju|'hoan	ktz	0.36
7	Kabyle	kab	0.69	18	Kambaata	ktb	0.36
7	Nafusi	jbn	0.69	19	Dahalo	dal	0.28
7	Tamasheq, Kidal	taq	0.69	20	Kwadi	kwz	0.25
7	Tamazight, Central	tzm	0.69	21	Bila	bip	0.22
7	Zenaga	zen	0.69	22	Pero	pip	0.12
8	Amharic	amh	0.67	23	Mwaghavul	sur	0.08
8	Gola	gol	0.67				
8	Noon	snf	0.67				

Table 7.3 shows that the highest GCS is 1 and the lowest 0.08. None of the languages of my sample thus gets the lowest possible score, 0 (see §7.6).

The results given in table 7.3 are also displayed in the graph in figure 7.1.

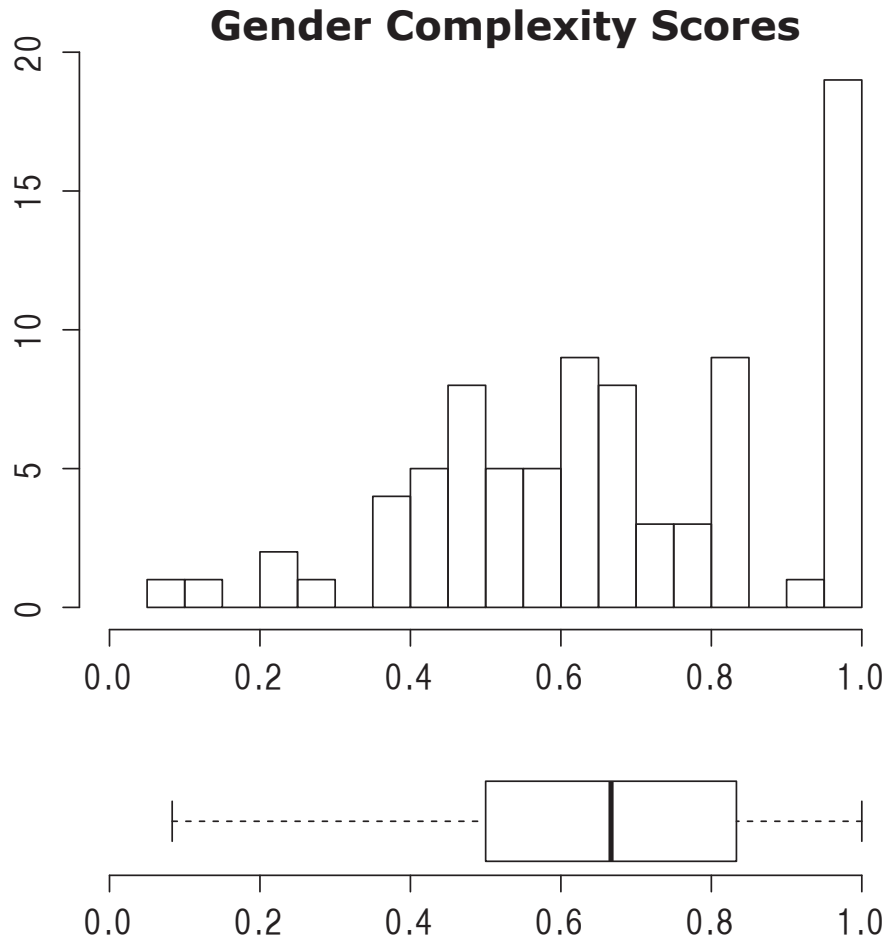


Figure 7.1: Distribution of the GCSs

The X-axis of the histogram displays the range of attested GCSs, whereas the Y-axis shows the distribution of the number of languages per GCS score. The box plot below the histogram provides the distribution of the GCSs per quartiles, and thus the boldface line represents the median. The figure shows that half of the languages of my sample have a GCS that ranges roughly from 0.5 to 0.8. In my data sample, high GCSs are substantially more frequent than low GCSs. The geographical distribution of the GCSs is represented in the map provided in figure 7.2.

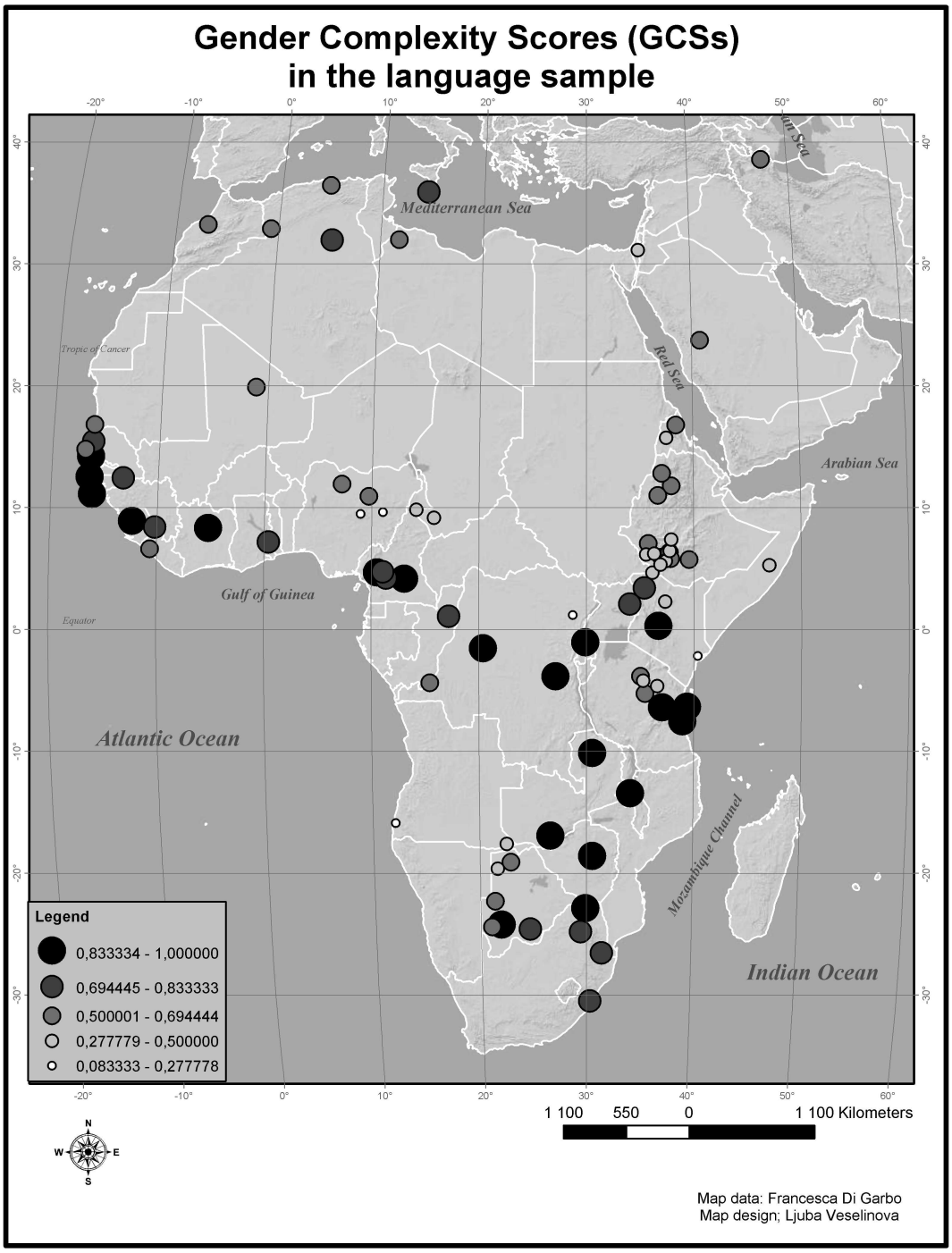


Figure 7.2: Geographical distribution of the GCSs

Before continuing with further discussion, an illustration of the procedure followed to calculate the GCSs of two languages of my sample is given. For the sake of clarity, I consider one language for which all features are documented, Turkana (Eastern Nilotic, rank 3 in table 7.3), and one for which two features are missing, Timne (Mel, rank 1 in table 7.3).

In order to classify the gender system of Turkana with respect to the features of my complexity metric, I followed the description provided by Dimmendaal (1983). Turkana has three gender values: Masculine, Feminine and Neuter. It thus gets 1/3 with respect to the feature GV. Gender assignment is both semantic and formal, and, as such, the value of AR is 1. According to Dimmendaal's description, there are three indexing targets for gender in Turkana: determiners, adjectives and pronouns (not the Personal Pronouns). Thus the language gets 2/3 with respect to the feature IND. In Turkana, gender distinctions are encoded cumulatively with number (CUM = 1). Finally, in Turkana gender shifts can be used to encode variation both in the countability properties of nouns (M1 = 1) or in the size of the NP referent (M2 = 1). In Turkana, when an uncountable masculine or feminine noun is shifted to the Neuter Gender, the resulting meaning is singulative. On the other hand, when countable masculine or feminine nouns are shifted to the Neuter Gender, the resulting meaning is diminutive (see §6.4 for more details). To summarise, for Turkana, the values assigned to each of the metric features are:

$$\text{GV} = 1/3;\ \text{AR} = 1;\ \text{IND} = 2/3;\ \text{CUM} = 1;\ \text{SYNC} = 1;\ \text{M1} = 1;\ \text{M2} = 1$$

Applying the formula illustrated in §7.6, $\frac{1/3+1+2/3+1+1+1}{6}$, the GCS of 0.83 is obtained.

The gender system of Timne is classified based on its description by Wilson (1961). Timne has more than five genders and thus gets 1 with respect to the feature GV. Gender assignment is both semantic and formal. Therefore, Timne gets a 1 with respect to the feature AR. According to Wilson's description, Timne has the following indexing targets for gender: adjectives, pronouns (of various type), verbs and the Indefinite Stabilizer, which is used with indefinite nouns in order to encode non-verbal predication (Wilson 1961: 11); the language thus gets a 1 with respect to IND. Gender and number are encoded cumulatively on the indexing targets (CUM = 1). The source does not provide any kind of information about gender shifts, which are, however, rather common phenomena in languages with similar gender systems. The features M1 and M2 cannot be documented for Timne. To summarise, for Timne, the values assigned to each of the metric features are:

$$\text{GV} = 1;\ \text{AR} = 1;\ \text{IND} = 1;\ \text{CUM} = 1;\ \text{M1} = \text{–};\ \text{M2} = \text{–}$$

Since two features are missing, the sum of the feature values is in this case divided by 4: $\frac{1+1+1+1}{4}$. The GCS of Timne is thus 1.

7.7.1 Genealogical and areal biases in the distribution of GCSs

In general, the results presented in table 7.3 suggest that closely related languages tend to have the same or very similar gender complexity scores. To give an example, all the

Berber languages in the sample have a gender complexity score of 0.69, with the only exception being Tachawit (shy) whose score is 0.75 (Tachawit scores 1 with respect to IND, whereas all the other Berber languages of the samples score 2/3; see appendix F). This tendency towards intragenealogical homogeneity in the absolute complexity of grammatical gender is perhaps unsurprising, given that grammatical gender tends to be a rather conservative feature of language families (see chapter 2). Outliers can be explained as the effect of more language-specific patterns, such as contact with other speech communities or language-internal historical developments. Let us consider a couple of examples.

Out of 84 languages, 19 scored 1, with all these being either Bantu, North-Central Atlantic or Mel. As observed throughout this thesis, typically, the gender systems of the Bantu and Atlantic type (i.e., North-Central Atlantic + Mel) exhibit features of high grammatical complexity: high number of gender distinctions, pervasive gender indexation, manipulability of gender assignment to express variation in the countability properties of nouns and/or in the size of the NP referents. Those Atlantic and Bantu languages which rank lower than 1 in table 7.3 have gender systems in which one or more of the above-mentioned features has/have been either weakened or lost. Many such cases of reduction and loss have already been presented in the previous chapters. For instance, as discussed in detail in §6.3.3, in 8 of the 23 Bantu languages of my sample – Bafia, Eton, Northern Sotho, Shona, Swati, Tswana, Venda, Zulu – diminutive and augmentative suffixes have grammaticalized from nouns. Of these eight languages, only Venda combines the use of the diminutive and augmentative suffixes with the uses of the diminutive and augmentative genders. In the remaining seven languages, the evaluative genders have been lost. As a result, the absolute complexity of the gender systems of these languages has decreased.

Two outliers with respect to the Bantu and Atlantic type of gender system are the Bantu languages Kinshasa Lingala and Bila. Kinshasa Lingala, the variety of Lingala spoken in the capital city of the Democratic Republic of Congo, has a gender complexity score of 0.56, that is, roughly 50% less than a Bantu language with a more prototypical gender system. Kinshasa Lingala preserves a variety of gender markers on nouns, but the only opposition that is marked on the indexing targets is *animate* vs. *inanimate*. In addition, the number of gender-indexing targets is massively reduced: gender distinctions are only marked on verbs and pronouns, and adjectives almost never index gender with nouns (Bokamba 1977: 70). Compared to Makanza Lingala, the standard variety, Kinshasa Lingala has undergone massive grammatical semplification. As Bokamba (1977: 75) puts it, "[t]he background linguistic heterogeneity of KL [Kinshasa Lingala] speakers certainly constitutes" one of the reasons behind this process of simplification and reduction.

Bila has a GCS of 0.22, the lowest score in comparison with the other languages of the Bantu group. Bila is spoken in the northeastern part of the Democratic Republic of Congo, which is also the northernmost corner of the Bantu-speaking area. The gender system of Bila is very simple with respect to nearly all the dimensions of the complexity metric used in this thesis (table 7.1). There are only two genders, the Animate Gender and the Inanimate Gender. Gender assignment is semantic, and there is no possibil-

ity of manipulating gender assignment. In addition, contrary to other Bantu languages where verbs are one of the most common gender-indexing targets, gender indexation in Bila is exclusively NP-internal (following Lojenga 2003: 462, the indexing targets of Bila are: adjectives, numerals and demonstratives). Such a simplified gender system is an innovation shared with other northern Bantu languages and explained by bantuists as a result of intense language contact with neighbouring non-Bantu languages (Lojenga 2003). The northern part of the Bantu-speaking area is often described as a true borderland between linguistically very diverse communities that have extensive contact with each other. In this area, Bantu speakers are surrounded by speakers of Nilo-Saharan and Ubangi (Niger-Congo) languages (Lojenga 2003: 451-452). Due to intense mutual contact, both the Bantu and non-Bantu languages spoken in this area are characterised by massive lexical borrowing as well as by grammatical innovations that are not shared with the respective cognate languages outside the area. The reduced gender system of Bila and other neighbouring Bantu languages is one of such innovative features.

The Semitic languages are an interesting case of closely related languages with non-homogeneous complexity scores. The highest ranking gender systems within the Semitic sample are found in Maltese (0.83) and Amharic (0.67). Moroccan Arabic, Standard Arabic and Tigre have the same complexity score, 0.61. The lowest ranking gender system is found in Hebrew (0.44), whereas Lishan Didan scored 0.53. Interestingly, the highest GCS, 0.83, is scored by Maltese, the Semitic language that is most prominent in its long-standing history of language contact with English and some Romance languages (Italian and Sicilian).

Two additional examples of outliers are Dahalo, with respect to the other Cushitic languages, and Kwadi, with respect to the Khoe-Kwadi group. Dahalo has a GCS of 0.28, and its gender system has been described by Tosco (1991: 20) as dying out as a result of contact with the neighbouring Bantu languages. Too little is known about Kwadi, a now extinct language of Angola. Güldemann (2004) describes its gender system as sex-based and pronominal, but nothing is said about the mechanisms of gender assignment nor about the use of gender shifts to encode diminutive and augmentative meanings (which is well documented in all the other Khoe-Kwadi languages of the sample).

Finally, the two lowest ranking languages in the complexity rank given in table 7.3 are the Chadic languages Mwaghavul (GCS = 0.08) and Pero (GCS = 0.12), both of which are spoken in Nigeria.

Mwaghavul scores 0 with respect to all the features of the complexity metric except for CUM, for which the score is 0.5. There are two genders in Mwaghavul (Masculine and Feminine), gender assignment is semantic and gender indexation is only pronominal. Finally, there seems to be no possibility of manipulating gender assignment in the language. With respect to the cumulation parameter, Mwaghavul shows at least some patterns of interaction with number on the indexing targets. The Third Person Human Anaphoric Subject and Object Pronouns encode gender and number cumulatively. On the other hand, the Third Person Non-human Pronoun, *nə̄*, encodes neither gender nor number distinctions (Frajzyngier & Johnston 2005). A similar type of system is found in Pero even though, from the description provided by Frajzyngier (1989), it is not entirely clear what type of assignment rules the language has and whether or not gender assign-

ment can be manipulated in any of the ways considered in this thesis. The remaining four Chadic languages with gender in the sample have higher GCSs (between 0.62 – Lele – and 0.5 – Gidar).

To summarise, genealogically related languages tend to have similar GCSs. On the other hand, multilingualism, language contact and second language learning seem to be very likely triggers of changes in the gender system of a language as opposed to its closest relatives (see, among others, Trudgill 1999; McWhorter 2001), and are also rather ordinary sociolinguistic settings in the African macro-area. A systematic account of the effects of language contact on the absolute complexity of gender in the languages of the sample lies, however, outside the scope of this dissertation.

7.7.2 Same complexity score does not mean same type of gender system

As mentioned at the beginning of the section, it is important to emphasise that, with the exception of the languages with the highest complexity score (= 1), identical complexity scores do not mean that languages have the same type of gender system. Different gender systems may arrive to the same complexity score via different paths. These differences become clear when considering how the individual languages score with respect to the individual features of the complexity metric (see appendix F). For instance, let us consider the case of Noon (North-Central Atlantic, rank = 8) and Amharic (Semitic, rank = 8). The two languages have the same GCS, 0.67, but have very different types of gender systems. Noon has a non-sex-based gender with more than five gender distinctions, whereas Amharic has a sex-based gender system with two gender distinctions, Masculine vs. Feminine. Thus, Noon scores 1 with respect to GV, whereas Amharic scores 0. The situation is reversed with respect to IND: gender indexation is more pervasive in Amharic (1) than in Noon (0). The two languages share the same scores with respect to the remaining features of the metric.

7.7.3 Implicational relationships between the features in the metric

On the basis of the results presented in table 7.3 an interesting relationship can be observed between the features GV and AR, and AR and IND.

Strictly semantic systems of gender assignment are only found in 6 of the 84 gendered languages within the sample: Bila (Bantu), Dahalo (Cushitic), Dime (South Omotic), Dizin (Dizoid), Masai (Eastern Nilotic), Mwaghavul (Chadic) (see also section 4.2). All these languages have two gender distinctions, and all but Bila have sex-based gender. Within my language sample then, strict semantic gender assignment is only found in languages with two or a maximum of three gender values. Moreover, there seems to be a preference for strictly semantic gender assignment in African languages to be based on cognitively basic oppositions such as *human* vs. *non-human*, *male* vs. *female*, *animate* vs. *inanimate*. It would be interesting to investigate what type of preferences exist, if they exist, in areas of the world where strictly semantic gender assignment is more common.

Finally, it is worth mentioning that the six languages of my sample with strictly

semantic gender assignment all score less than 1 with respect to IND: thus in none of these languages is gender indexation maximally pervasive. These results are in line with what was suggested by Audring (2009) with respect to the correlation between pervasiveness of indexation and type of assignment rules. Audring analyses the assignment rules of a number of pronominal gender systems from different areas of the world, and considers aspects of the diachrony of gender in English and Dutch. She shows that pronominal gender systems – where manifestations of gender throughout the discourse are rather poor – display a strong preference towards strictly semantic assignment rules. Within my language sample, only Mwaghavul (Chadic) has pronominal gender and semantic assignment. The remaining five languages with strict semantic assignment score either 1/3 or 2/3 with respect to IND. These results suggest that when strict semantic gender assignment is found in non-pronominal gender systems, gender indexation is still not maximally pervasive. In other words, semantic assignment seems to generally tolerate lower amount of formal marking.

7.7.4 Some features may be stronger predictors of gender complexity than others

In several occasions throughout this chapter, I mentioned that a major issue when investigating grammatical complexity is how to quantify the contribution that the individual features of a metric bring to the overall complexity score (what Miestamo 2008 refers to as *the problem of comparability*). Given that it is extremely difficult to measure the relative weight of the individual features of a complexity metric, overall complexity results can be extracted based on a language's average behaviour with respect to a set of features. These results describe tendencies in the absolute complexity of a grammatical domain with respect to an artificially designed set of cues. Therefore, they cannot be interpreted as uncontroversial measurements (for a similar discussion in their study of complexity in nominal plural allomorphy, see also Dammel & Kürschner 2008).

I would like to suggest here that one way of indirectly investigating the behaviour of the features of a complexity metric is to correlate the individual features with each other. In order to do so with my own complexity metric, I calculated the Squared Spaerman rank correlation coefficients between the individual features of the metric. The results are represented in the graph in figure 7.3.

Figure 7.3 is organised as follows. The individual features of the metric are displayed both horizontally and vertically. In this way, correlations coefficients between pairs of features can be read both row-wise and column-wise. Correlation coefficients are visualised according to a colour scale whereby white stands for *no correlation* and grey for *high correlation*. The grey diagonal area that cuts across the two halves of the figure represents correlation coefficients between pairs of the same features (that is, CUM with CUM, M2 with M2, etc.). These grey boxes correspond to a correlation coefficient that equals to 1 since each feature obviously has the highest correlation with its own copy. These results are thus not relevant to our analysis. With respect to correlations between pairs of different features, the figure shows that the highest correlation coefficients are found between IND and M1 (= 0.372), GV and M1 (= 0.324), and GV and IND (=

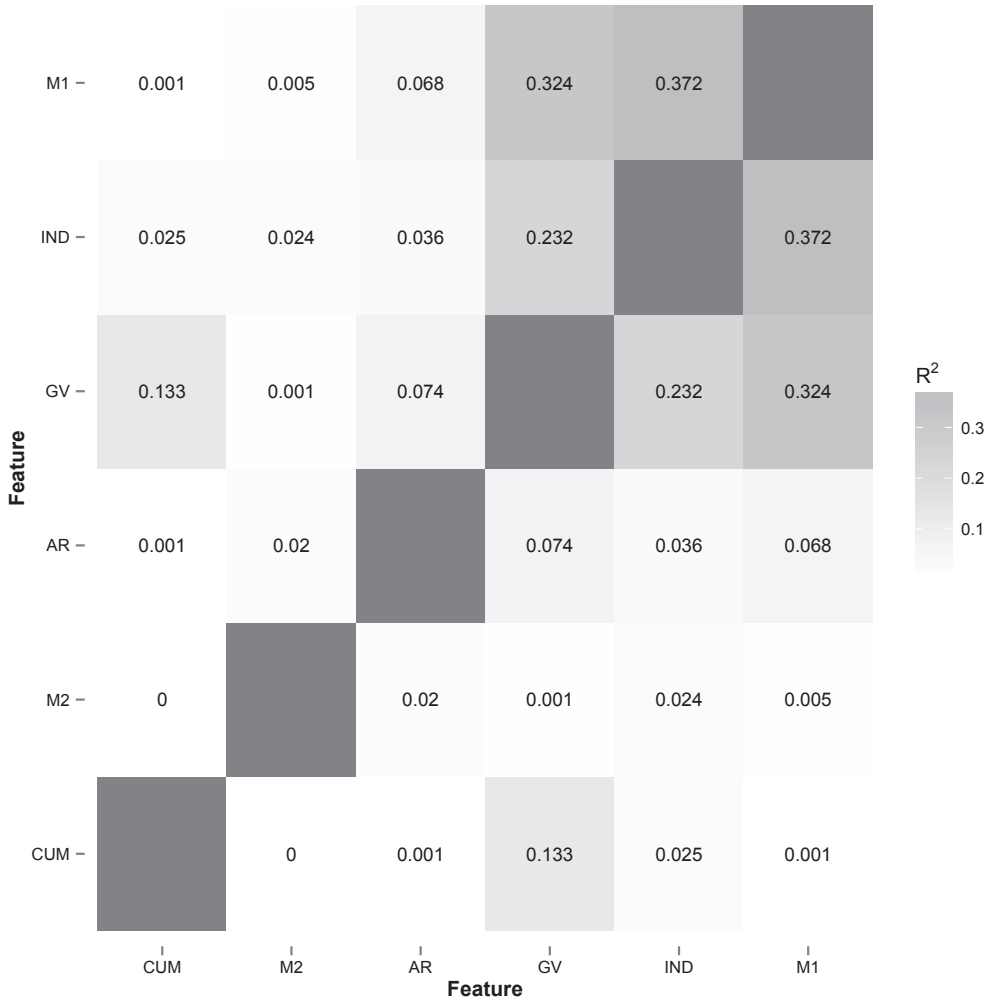


Figure 7.3: Correlation coefficients between the features of the metric

0.232).

The high correlation coefficients between IND and M1, on the one hand, and GV and M1, on the other, can be interpreted as follows. In the languages of my sample, the possibility of manipulating gender assignment to encode variation in the countability properties of nouns goes hand in hand with the presence of very pervasive gender indexation or, to a slightly lower degree, high number of gender values. As shown in section 5.7, M1 is not widely distributed across the gender systems of the language sample. It is only found in Bantu, North-Central Atlantic, Berber, a subset of the Semitic languages, and in the Eastern Nilotic language Turkana. In a way then, both the distribution of M1 and its correlation coefficients with IND and GV suggest that M1 is a very special property of gender systems, which can only be found in systems with a high amount of formal marking (IND) and/or a high number of gender distinctions (GV). On the contrary, the results show that M2, that is, manipulation of gender assignment to express diminutive and augmentative meanings, has extremely low correlation coefficients with both IND and GV as well as with all the other features of the metrics.

As mentioned above, the correlation coefficient between GV and IND is 0.232. The two features correlate with each other but not as strongly as one would expect based on Audring's (2014) argument, whereby pervasive indexation is likely to be found in languages with a high number of gender values (see §7.2).

Moreover, figure 7.3 shows that AR (Assignment Rules) has extremely low correlation coefficients with all the features of the metric. These results might depend on the fact that only 6 of the 84 languages with gender in my sample have semantic gender assignment. In other words, nearly all the languages of the sample behave similarly with respect to this parameter. It would be interesting to investigate the behaviour of this feature in areas of the world where semantic gender assignment is more frequent and compare it with my results from Africa. Finally, equally low correlations are found with the feature CUM.

One question that is worth asking is whether the correlation coefficients presented in figure 7.3 can tell us anything about which of these features is the best predictor of the GCS of each language. Since the GCS is the averaged sum of the values that a language takes for each feature in the metric, the features that show the highest correlations with each other (M1, IND and GV) can be expected to be those which also have a stronger impact on the final score. This can be verified by examining the associations between the independent variables (the features in the metric) and the dependent variable (the GCS) in a purely descriptive way, that is, by stratifying our dependent variable, the GCSs, according to the potential predictors, the individual features in the metric (Harrell 2001: 125). This is shown in figure 7.4.

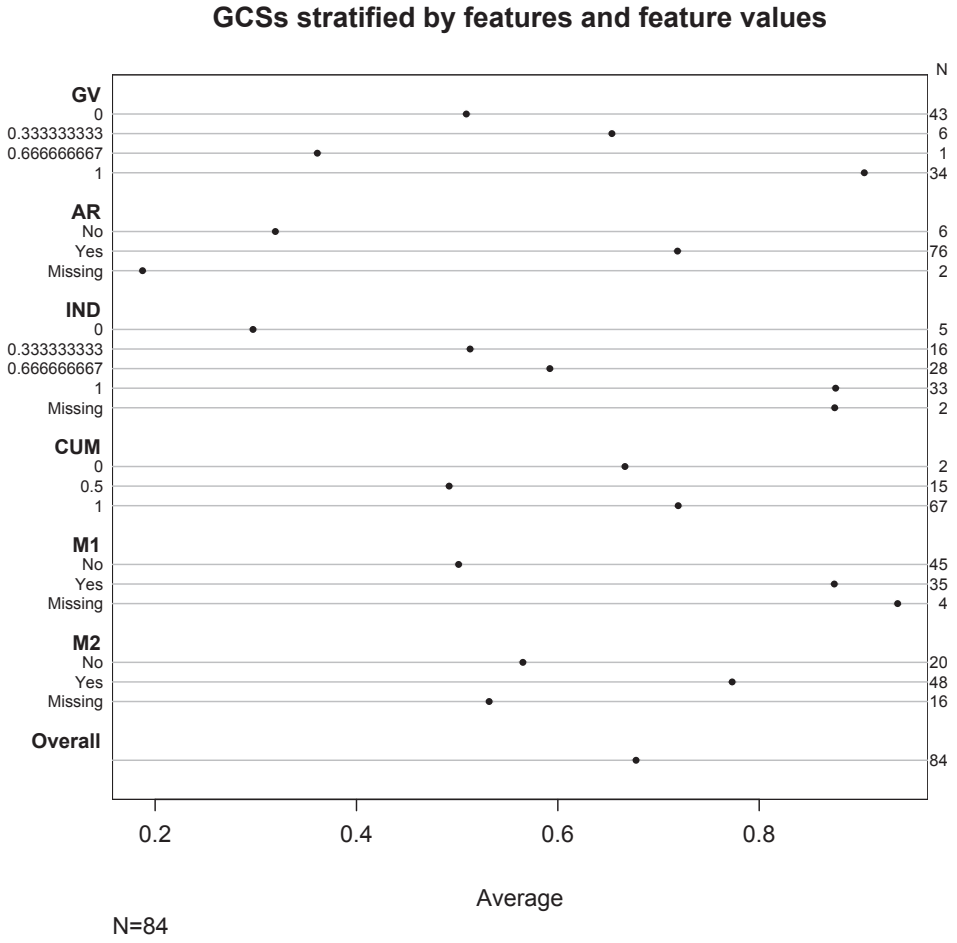


Figure 7.4: GCSs (Average) stratified according to feature values

Figure 7.4 is organised as follows. The GCSs are displayed on the X-axis. The left Y-axis represents the values assigned to each feature in the metric; the right Y-axis shows the number of languages in the sample where each of the feature values is found. The black dots represent the mean of the GCSs that languages displaying a certain feature value have. For instance, it shows that languages that score 1/3 (0.3333333333) with respect to GV have a GCS which, on average, ranges between 0.6 and 0.8. The black dots thus allow us to see which of the features and feature values can trigger the highest GCSs in the languages of the sample. As hypothesised based on the correlation coefficients shown in figure 7.3, in the languages of my sample, the highest scores in GV, IND and M1, trigger higher GCSs. With respect to GV, the figure shows that the impact of the different feature values on the GCSs grows from 0 to 1/3, drastically drops at 2/3 and grows again at 1. This is likely to be an effect of the fact that only one language within my sample has four gender distinctions, Ju‡'Hoan (Kxa). As discussed in §7.7, Ju‡'Hoan has a GCS of 0.36, which is one of the lowest scores in my language sample.

To summarise, even though the quantitative analysis applied to the complexity data does not provide a solution to the problem of comparability, it provides good tools for describing the behaviour of the complexity metric with respect to the dataset considered in this dissertation. Provided that my metric is a good measure for gender complexity, the results suggest that M1, IND and GV are the features which correlate more strongly with each other and those which seem to have the stronger impact on the final complexity scores of the languages of my sample.

7.8 Summary of the chapter

The purpose of this chapter has been to elaborate a metric for the grammatical complexity of gender systems that could also account for the interactions that gender has with the domains of nominal number and evaluative morphology.

The notion of grammatical complexity that I have worked with in this chapter is of the absolute type (see §7.2). I argued that the absolute complexity of a grammatical domain can be assessed by using three principles as major guidelines: the Principle of Fewer Distinctions, the Principle of One-Meaning–One-Form and the Principle of Independence. While the first two principles are based on Miestamo (2008) and Sinnemäki (2011), the Principle of Independence is my own and is used as a means of assessing interactions between grammatical domains and their contribution to the complexity of individual domains. Following Dahl (2011), I also argued that, in order to be maximally local, complexity metrics ought to be based on *ceteris paribus* comparisons.

The bulk of my complexity metric is an expanded version of the criteria for gender complexity suggested by Audring (2014). In order to translate the values of each features into numbers, I followed the method designed by Parkvall (2008). The feature values, their numerical interpretation and the methodology followed in order to compute the grammatical gender complexity of the languages of the sample have been presented in §7.6.

In §7.7.1, I showed that the genealogical biases within the sample are reflected by the results on gender complexity: languages belonging to the same genealogical unit tend to have similar complexity scores. The distribution of the outliers can be usually explained as the effect of language contact. In §7.7.2, I showed that languages with the same complexity score do not necessarily have the same type of gender systems. Possible implicational relationships between some of the features in the metric were discussed in §7.7.3. Finally, in §7.7.4, I showed that some of the features in the metric correlate more with each other than others and can be seen as better predictors of the final complexity score. As shown in figure 7.1, when they have gender, the languages of my sample tend to have rather complex gender systems, whereas gender systems with lower levels of complexity are very rare. If this sample is considered to be representative of the African macro-area, these results suggest that gender is a complex feature of African languages.

8 Summary of the main results, prospects for future research and concluding remarks

In this chapter, I summarise the results of my investigation and assess their contribution to the understanding of gender, viewed in its interactions with number and evaluative morphology. The methodology followed in the study, and its relevance to linguistic typology are also discussed. Shortcomings and limitations of the work are pointed out together with suggestions for future research on the topic.

8.1 Summary and assessment of results

Three research foci were established in chapter 2: (1) to investigate the interactions between gender and number; (2) to investigate the interaction between gender and evaluative morphology; and (3) to measure the impact of these interactions on the absolute complexity of gender systems. These three research foci were tackled individually in three independent chapters: chapters 5, 6 and 7, respectively. The research questions addressed in each of the three chapters are here repeated for convenience. The results and findings of each chapter are summarised and assessed thereafter.

(8.1) Research questions concerning gender and number (chapter 5)

- Q 1: How common is cumulative exponence of gender and number in the languages of the sample?
- Q 2: What are the formal and semantic factors that trigger gender syncretism in the context of number? Does gender syncretism in the context of number presuppose cumulative exponence?
- Q 3: What are the implications of cumulative exponence and syncretism on the absolute complexity of gender and number systems?
- Q 4: Can these types of interaction between gender and number be seen as a reflex of a nominal relevance hierarchy?
- Q 5: Can gender and number compete through indexation?
- Q 6: Is there any correlation between types of encoding of gender and types of encoding of number?
- Q 7: What types of semantic interactions can be found between gender and number?

(8.2) Research questions concerning gender and evaluative morphology (chapter 6)

Q 8: How frequently does *size* occur as an independent gender value? How stable and how widely distributed is this phenomenon within genenealogical units?

Q 9: Do the interactions between gender and evaluative morphology differ across types of gender systems and/or strategies of gender assignment (e.g., *sex-based* vs. *non-sex-based* gender systems, or *manipulable* vs. *rigid* gender assignment)?

(8.3) Research question concerning gender complexity (chapter 7)

Q 10: How do interactions between gender and number and gender and evaluative morphology affect the grammatical complexity of gender systems? Is it possible to measure the effect of these interactions on the absolute complexity of individual gender systems?

8.1.1 Gender and number

In chapter 5, morphosyntactic and semantic interactions between gender and number were considered. The chapter was divided in four parts.

In the first part of the chapter (§§5.2, 5.3 and 5.4), I examined patterns of exponence of gender and number as well as syncretism of gender in the context of number. Exponence and syncretism are often regarded as two crucial dimensions along which gender and number interact in language. However, to my knowledge, this dissertation offers the first systematic survey of the patterns of exponence and syncretism of gender and number that are attested in a large sample of languages. Cumulative exponence of gender and number on the indexing targets turned out to be extremely frequent in the languages of the sample as did syncretism. Since not all languages with cumulative exponence have syncretism, but nearly all languages with syncretism have cumulative exponence, I suggested that syncretism of gender in the context of number presupposes cumulative encoding of the values of the two grammatical domains. I suggested that cumulation should be viewed as a factor that increases the absolute complexity of gender systems insofar as it introduces noise in the mapping between meaning and form. Finally, I argued that my results on cumulation and syncretism may be read as possible indicators of the existence of a relevance hierarchy for nominal features. In agreement with Greenberg (1963b), Carstairs (1987) and Carstairs & Stemberger (1988), and Vafaeian (2013) – each of whom worked on different grammatical phenomena and different datasets – these results suggest that if a nominal relevance hierarchy existed, number would be the highest ranking feature.

In the second part of the chapter (§5.5), I focussed on the existence of competing patterns of gender and number indexation in a rather small subset of the language sample. These languages all belong to the Afro-Asiatic phylum and display a bipartite sex-based gender system. In addition, in these languages, there is a split in the indexation patterns associated with plural nouns. Some plural nouns trigger the same indexation as either masculine singular or feminine singular nouns, whereas other plural nouns trigger

an indexation pattern that differs both from the masculine and the feminine singular. Since the latter indexation pattern is only used to index plurality, I proposed to refer to it as Dedicated Plural Indexation (DPI). Furthermore, I suggested to label languages of this type as languages with split plural indexation systems. I showed that the use of DPI may be constrained either by the animacy (Standard Arabic, Miya) or by the lexical plurality of the head nouns (Cushitic). A case study of the Cushitic languages of the sample was carried out in order to examine how different languages within the group behave with respect to the use of DPI. In the literature on Cushitic, languages with a rather limited use of DPI are traditionally described as displaying a third indexation class, and thus a third gender, besides masculine and feminine: the plural (for an overview, see Mous 2008). This approach has been thoroughly criticised in more typologically oriented literature (see, for instance, Corbett & Hayward 1987; Corbett 2000). In these studies, however, the use of DPI has often been labelled as a mere exception without any explanation of its idiosyncrasies being attempted (a recent exception is Corbett 2012, who, for the first time, attempts to explain gender and number indexation in Baiso by shifting the attention to the role of lexical plurality). In my own study, I showed that an alternative overarching explanation for this long-debated issue becomes possible when looking at the distribution of DPI across different languages of the family and examining the factors that play a role in constraining its occurrence. In the Cushitic languages with split plural indexation, DPI is only legitimate with those indexation triggers that are inherently associated with plurality. This does not necessarily mean that DPI is the syntactic manifestation of a gender. It simply shows that two domains, one grammatically bound to indexation (gender) and the other potentially encodable through indexation (number) may compete with each other on (at least) some indexing targets. I showed that the division of labour between the two is done at different cut-off points in different Cushitic languages. In most languages nominal plurality (be it morphological or lexical) is always indexed syntactically, in other languages only lexical plurality is indexed.

In the third part of the chapter (§5.6), I showed that in the languages of the sample, the development of pervasive indexation systems always involves the presence of both gender and number and that the languages of the sample tend to exhibit the same number of gender- and number-indexing targets.

In the fourth and last part of the chapter (§5.7), I showed that in languages with manipulable gender assignment, that is, in languages where nouns can be assigned to multiple genders, gender shifts can be used to encode variation in the countability property of nouns. This phenomenon has a very skewed distribution in the languages of the sample.

8.1.2 Gender and evaluative morphology

In chapter 6, morphosyntactic and semantic interactions between gender and the morphological encoding of evaluation were considered. Through this chapter, this dissertation provides the first extensive typological account of how the encoding of *size* can be incorporated in the gender system of a language.

Interactions between gender and evaluative morphology were explored taking different types of gender systems as the independent variable. I found that, in the African macroarea, languages with large gender systems tend to have dedicated evaluative genders that are used to encode diminutive and augmentative meanings. On the other hand, in languages with smaller and sex-based gender systems, there are no dedicated evaluative genders but the masculine, feminine or neuter genders, which may be used to encode diminutive and augmentative meanings when a noun is shifted from its default gender to (one of) the other(s). The two types are widespread throughout my language sample and across genealogical groupings. In some cases, it was possible to assess their stability as well as aspects of their diachronic development.

Ultimately, what the two types of systems have in common is the possibility to manipulate gender assignment in order to modify the construal of the NP referent with respect to the parameter of *size*.

8.1.3 Interactions of gender and grammatical complexity

The aim of chapter 7 was to measure the impact that the interactions between gender and number and gender and evaluative morphology have on the absolute complexity of a gender system. The suggestion made in the chapter is that the more interactive a grammatical domain, the higher its absolute complexity.

Elaborating metrics to assess the absolute complexity of individual grammatical domains is a relatively novel field within language typology. In the case of gender, this has been discussed in pioneering work by Audring (2014). The paper proposes three dimensions along which gender complexity can be computed but does not provide a method to convert the variables associated to these dimensions into measurable values. In chapter 7, Audring's model of gender complexity was expanded with two new aims:

(1) Accounting for the morphosyntactic and semantic interactions between gender and number and gender and evaluative morphology explored throughout the thesis.

(2) Converting the values associated to each dimension of Audring's model into numbers.

The second aim was achieved by using the method designed by Parkvall (2008). The results of the calculations suggest that: (1) complexity scores are genealogically biased; (2) languages with the same complexity score may arrive to it through different paths; (3) some of the features in the metric interact with each other implicationally (e.g., strict semantic gender assignment is only found in languages with small gender systems); and (4) some of the features in the metric correlate with each other and seem to have a higher impact on the final complexity scores of the languages of the sample.

8.2 Assessment of the sampling methodology

The sampling methodology followed in the dissertation was introduced in chapter 3, where I discussed general matters concerning sampling in linguistic typology, my own

sampling design and the procedure of data collection and organisation.

Methodologically, this dissertation was designed as an investigation of one continental area, Africa, and as an attempt at practising crosslinguistic comparison both within and across genealogical units. The choice of combining intra- and intergenealogical typology in the study of one of the world's macro-areas proved to be beneficial to answer the research questions posited by this investigation.

For instance, having subsamples of individual genealogical units within Africa was particularly crucial for reaching a better understanding of the following phenomena:

(1) The indexation patterns associated with plural nouns in Cushitic languages. Had I not had a relatively extensive sample of the languages of the group, I could not have explored the distribution of DPI throughout different members of the family. This ultimately led me to propose an interpretation of the phenomenon that, hopefully, sheds new light on our understanding of these languages and, more generally, on the typology of indexation systems.

(2) The distribution of the evaluative genders in the Bantu languages. Had I not had a relatively extensive sample of the languages of the group, I could not have shown how common and stable these genders are throughout the family and what types of innovations are attested in different subdivisions of the group.

On the other hand, intergenealogical comparison was essential for achieving an understanding of the distribution of the detected patterns across genealogical units. For instance, I showed that cumulation of gender and number and syncretism of gender in the context of number are very common throughout the languages of the sample; the distribution of these phenomena cuts across genealogical groupings and typological differences among the gender and number systems of the individual languages of the sample. In addition, I showed that there is a correlation between type of gender system and expected type of interaction with evaluative morphology and that this correlation holds true across the different genealogical units represented in the language sample. Large-scale genealogical biases in the distribution of individual phenomena were also found. For instance, split plural indexation systems were found in all the genealogical units of the Afro-Asiatic phylum included in the sample but Berber (that is, in Chadic, Cushitic and Semitic).

Finally, in some cases, sub-areal phenomena could be detected. For instance, innovations in the gender systems of the Bantu languages were found to be distributed according to sub-areas. The renewal of evaluative morphology – from evaluative genders to evaluative suffixes – is concentrated in the southeastern Bantu languages (chapter 6), whereas substantial reduction of the noun class system is mostly found in the northern part of the Bantu-speaking area.

An intra- and intergenealogical survey of one macro-area, as the one conducted in this thesis, cannot lead to the formulation of large-scale typological generalisations. On the other hand, a continent-based sampling methodology provides an understanding of the geographical distribution and diachronic stability of grammatical phenomena within restricted areas of the world that cannot be easily reached by means of large-scale language

comparison. The generalisations that result from continent-based typologies can only hold at the areal level but constitute a set of assumptions that may be in turn tested on other macro-areas of the world.

8.3 Prospects for future research

In this section, I discuss four domains of research for which more work is needed: nominal relevance hierarchy, split plural indexation systems, gender assignment and its manipulation, and absolute complexity of gender systems. For each of the four domains, the limitations of my own results are discussed before I provide some suggestions for further research on the topic.

8.3.1 Nominal relevance hierarchy

On more than one occasion throughout this thesis, it was pointed out that number is the feature of nominal morphology most likely to serve as the context of syncretism, but which is less subject to undergo syncretism itself. In other words, certain distinctions which are grammatically salient for nouns tend to be reduced or lost under the pressure of number values. For instance, when nouns are marked as nonsingular, other grammatical distinctions (such as case or gender) can be reduced or neutralized. This suggests that nominal number has more relevance to nouns than other nominal features. However, as Vafaeian (2013: 122) puts it, "[a] relevance hierarchy for nouns is yet to be proposed and must be based on semantic arguments and confirmed by typological data in the same way as was done for verbs."

In my opinion, further research on the nominal relevance hierarchy should be carried out on the basis of a clear-cut definition of the domains under investigation. In particular, when collecting and analysing the typological data, it may be useful to distinguish between two major domains:

(1) Patterns of encoding of nominal features on the noun stem.

(2) Patterns of encoding of nominal features via indexation.

A systematic scrutiny of these two domains may provide different insights into the way in which the semantic relevance of nominal categories is reflected in morphosyntax. For instance, as shown in this thesis, since manifestations of gender are bound to indexation, gender syncretism in the context of number can only be investigated by looking at how the gender and number of a noun are encoded on the indexing targets. On the other hand, as shown by Vafaeian (2013), the analysis of suppletive noun stems provides insights on the hierarchical relationships that exist between other sets of nominal categories, such as case, possession and, again, number. Methodologically, distinguishing between the domains of investigation outlined in (1) and (2) could be a useful way of classifying the typological data that may account for the existence of the nominal relevance hierarchy.

8.3.2 Split plural indexation systems and Dedicated Plural Indexation

The notions of *split plural indexation* and Dedicated Plural Indexation have proved to be very useful in order to account for the indexation patterns associated with plural nouns in a small subset of Afro-Asiatic languages within my sample. In addition, I was able to show that within Cushitic languages, split plural indexation systems, where the use of DPI is restricted to some plural triggers only, coexist with systems in which the use of DPI is generalised to all plural nouns.

What my case study of Cushitic was not able to provide is a diachronic interpretation of the data. What is the diachronic relationship between systems with generalised use of DPI, systems with split plural indexation and systems where there is almost no use of DPI? Are systems with generalised use of DPI an innovation or a retention with respect to the two other types of systems? In order to answer these questions, a more extensive investigation of the Cushitic languages would be needed. The sample should be expanded, and, possibly, data from descriptive sources should be combined with data from spoken corpora. This would allow us to verify whether the parameters that, according to my findings, condition the use of DPI in the split plural indexation systems of my sample also account for Cushitic languages outside the sample. Once a more precise classification of the types of indexation systems synchronically attested within Cushitic is reached, it may be easier to analyse their distribution from a historical point of view and attempt at accounting for their diachronic development.

At a more general level, the validity of the two notions – split plural indexation and DPI – should be tested outside the African macro-area. In particular, it might be useful to verify whether the following two generalisations hold outside Africa:

(1) That split plural indexation systems are only attested in languages with sex-based gender.

(2) That the distribution of DPI within split plural indexation systems is only constrained by the animacy and the lexical plurality of the indexation triggers.

8.3.3 Gender assignment and its manipulation

In §5.7 and chapter 6, I introduced the notion of *manipulable gender assignment*, and I used it to refer to those languages where the gender of a noun can change according to the way the NP is construed. In such languages, there usually are default assignment rules, whereby nouns have lexically specified genders, and add-on assignment rules which allow speakers to modify the construal of the NP referent by changing its gender. This is mainly done in order to express variation in the size of the noun referent or in the speakers' attitude towards it. In §5.7, I showed that there are languages in which similar mechanisms of gender shift seem to be used to encode variation in the countability properties of nouns. This phenomenon is attested both in small and large, sex-based and non-sex-based gender systems but is generally not very frequent across the different genealogical groupings within the sample. Its distribution is rather concentrated within a few genealogical groupings.

If both my analyses are correct, the notion of manipulable gender assignment seem to challenge our current understanding of gender systems and, in particular, the assumption that gender is always rigidly specified in the lexicon as an inherent property of a noun. Further research in this field should be both empirical and theoretical in its aims. Empirically, the crosslinguistic distribution of manipulable gender assignment outside the African macro-area should be explored in a systematic fashion. The following research questions could be addressed:

(1) How common is manipulable gender assignment in the gender systems of the world languages?

(2) What meanings are expressed by gender shifts crosslinguistically?

(3) In languages with manipulable gender assignment, is it always the case that nouns have a default, lexical gender?

Theoretically, in my view, at least two questions are worth being asked:

(1) How do manipulable gender assignment systems challenge our model of gender assignment?

(2) How can manipulable gender assignment be distinguished from mere word-formation processes where derived nouns differ in gender from their base?

The theoretical question outlined in (1) has been rarely addressed in the literature and, mostly, in connection with the gender system of individual languages. For instance, Payne (1998) and, later, Shirtz & Payne (2012) point out that in the Eastern Nilotic language Masai most nouns do not have a default gender but are potentially compatible with any gender (see also the discussion in §6.4.1). Accordingly, they propose that gender assignment in Masai is based on the construal of the NP referent rather than on the semantics of the noun or its lexical properties. The majority of the languages with manipulable gender assignment in my sample display gender systems that are less atypical than the Masai system insofar as the majority of nouns do have default, albeit manipulable, gender. Yet, the results of the present investigation show that manipulable gender assignment is a frequent enough phenomenon to start considering a revision of our current understanding of gender assignment rules, their implications for the typology of gender systems, and, on a larger scale, their interaction with word-formation processes.

8.3.4 Absolute complexity of gender systems

Research on the grammatical complexity of gender systems, and how to compute it, is a relatively novel field within typology, and much still needs to be done.

As pointed out in §7.6, one of the shortcomings of my complexity metric is not to be able to provide a qualitative account of how different types of indexing targets can affect the absolute complexity of gender systems. In other words, the metric computes how many gender indexing targets there are in a language but does not account for whether or not it makes a difference that these targets might be, for instance, adjectives, pronouns,

verbs or determiners or any possible combination between them. In addition, the metric does not account for systems with *animate concord* and *split plural indexation* systems where, as shown throughout the dissertation, certain referential and lexical properties of the indexation triggers interfere with gender (and number) indexation patterns.

Finally, the metric presented in chapter 7 aims to measure the impact that interactions of gender with other nominal features have on the absolute complexity of gender systems. The general idea behind this attempt is that the more interactive a gender system, the higher its absolute complexity. The domains considered in this thesis are only two, number and evaluative morphology, but, as is known, grammatical gender tends to be deeply intertwined with other nominal features such as case and definiteness. An overview of these patterns of interactions is needed in order to incorporate them to the metric and assess their role on the complexity of gender systems.

A more general question that would be worth asking in future research pertains to the correlation between grammatical complexity of gender and *maturity* in Dahl's (2004) sense. However, since very little is known about young, nonmature gender systems, comparing the complexity of gender systems at different stages of maturation is a difficult task to undertake. As observed many times throughout the dissertation, gender is a special feature of grammar. It is stubbornly stable but does not tend to emerge often in the recent history of languages.

Finally, the results of the complexity calculations presented in chapter 7 suggest that typological distributions connected with gender systems in the African macro-area are somewhat bimodal. Languages either have complex gender systems – and this is the case for the majority of the languages in the sample – or they lack gender altogether. On the other hand, low levels of complexity are very rarely associated with gender. If these facts are interpreted in terms of stability, one can speculate that not only are noncomplex gender systems typologically infrequent but that they also represent diachronically unstable stages in the history of languages. In order to verify this hypothesis, the absolute complexity of the gender systems attested in other macro-areas of the world should be assessed following methods comparable to those adopted in this thesis.

8.4 Concluding remarks

This dissertation has focussed on an under-investigated aspect of the nature of gender systems: their interactions with other domains of grammar – number and evaluative morphology – and the relevance of these interactions to the absolute complexity of gender. This approach has turned out to be useful for analysing the gender systems of a large sample of African languages, challenge some pre-existing assumptions in the literature on gender, and spark further research along these lines.

Morphosyntactic and semantic interactions between individual features of grammar have been thoroughly investigated in the verbal domain where, as a result, grammatical features such as tense, aspect, mood and evidentiality are now seen as a congruent whole, both at a descriptive and theoretical level.

This investigation is certainly far from providing the ultimate methodology and theo-

retical justifications for an enterprise of this type to be applied to the nominal domain. The challenges are different because the factors at stake substantially differ. The results of this dissertation may however be seen as illustrating some of the advantages in looking at nominal features in their morphosyntactic, semantic and functional interaction, rather than as isolated domains of grammar.

A The language sample

A.1 Genealogical units and internal composition of the subsamples

Table A.1 lists the languages of the sample based on the genealogical units they are selected from (see also table 3.2). The language ISO codes are ordered alphabetically and are followed by the correspondent language names. Both are taken from Glottolog (Nordhoff et al. 2013) as of August 2014. Areal groupings, lower and higher levels of genealogical affiliation are indicated if relevant. In particular, lower subdivisions within a genealogical unit are spelled out only if they correspond to a genus in WALS. This is done in order to allow comparability with the WALS classification, which is genus-based. When the genealogical unit selected in my sample already corresponds to a genus in WALS (e.g., Berber or Semitic), no lower subdivision is given. Similarly, if the genealogical unit selected in my sample corresponds to a level below genus in WALS (e.g., Bantu, Eastern Nilotic or Western Nilotic), no lower subdivisions are indicated in the table. Finally, information about the sources that I consulted for each of the sampled languages is given in the table.

The following abbreviations are used in the table to refer to different kinds of language groupings:

AG = Areal Grouping

LGU = Lower Genealogical Unit

HGU = Higher Genealogical Unit(s)

An alphabetical index of the sampled languages is given in table A.2.

Table A.1: The language sample

Genealogical Unit	LGU	ISO code	Language	AG	HGU	Source
Bantu		baz	Tunen		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Mous (2003a)
		bem	Bemba		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	van Sanbeek (1955)
		bip	Bila		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Lojenga (2003)
		bvx	Dibole		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Leitch (2003)
		cgg	Chiga		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Taylor (1985)
		eto	Eton		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Van de Velde (2006, 2008)
		kik	Kikuyu		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Barlow (1951); Stump (1993)
		kki	Kagulu		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Petzell (2008)
		ksf	Bafia		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Guarisma (2003)
		lea	Lega		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Botne (2003)
		lin	Lingala		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Guthrie (1966)
		lol	Mongo		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Hulstaert (1965)
		mcp	Makaa		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Heath (2003)

Table A.1: (continued)

Genealogical Unit	LGU	ISO code	Language	AG	HGU	Source
Bantu		nya	Nyanja		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Watkins (1937)
		nso	Northern Sotho		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Poulos & Louwrens (1994)
		run	Rundi		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Meeussen (1959); Mel' čuck & Bakiza (1997)
		sna	Shona		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Fortune (1955)
		swa	Swahili		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Myachina (1981)
		ssw	Swati		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Taljard et al. (1991)
		toi	Tonga		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Carter (2002)
		tsn	Tswana		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Cole (1955)
		ven	Venda		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Poulos (1990)
		zul	Zulu		Bantoid, Benue-Congo, Volta-Congo, Atlantic-Congo	Poulos & Bosh (1997)
Berber		jbn	Nafusi		Afro-Asiatic	Beguinot (1942)
		kab	Kabyle		Afro-Asiatic	Mettouchi (2000)
		shy	Tachawit		Afro-Asiatic	Penchoen (1973a)
		taq	Tamasheq (Kidal)		Afro-Asiatic	Heath (2005)
		tzm	Tamazight, Central Atlas		Afro-Asiatic	Penchoen (1973b)
		zen	Zenaga		Afro-Asiatic	Nicolas (1953)

Table A.1: (continued)

Genealogical Unit	LGU	ISO code	Language	AG	HGU	Source
Chadic	**Biu-Mandara**	gid	Gidar		Afro-Asiatic	Frajzyngier (2008)
		hna	Mina		Afro-Asiatic	Frajzyngier & Johnston (2005)
		xed	Hdi		Afro-Asiatic	Frajzyngier & Shay (2002)
	East Chadic	lln	Lele		Afro-Asiatic	Frajzyngier (2001)
	West Chadic	hau	Hausa		Afro-Asiatic	Newman (2000)
		mkf	Miya		Afro-Asiatic	Schuh (1989)
		sur	Mwaghavul		Afro-Asiatic	Frajzyngier (1993)
		pip	Pero		Afro-Asiatic	Frajzyngier (1989)
Cushitic	**Beja**	bej	Beja		Afro-Asiatic	Roper (1928)
	Central Cushitic	ahg	Qimant		Afro-Asiatic	Appleyard (1975)
		awn	Awngi		Afro-Asiatic	Hetzron (1978)
	East Cushitic	bsw	Baiso		Afro-Asiatic	Corbett (2000); Corbett & Hayward (1987)
		dal	Dahalo		Afro-Asiatic	Tosco (1991)
		dsh	Daasanach		Afro-Asiatic	Tosco (2001)
		gax	Borana-Arsi-Guji Oromo		Afro-Asiatic	Andrzejewski (1960); Stroomer (1995)
		gdl	Dirasha		Afro-Asiatic	Sasse (1974)
		ktb	Kambaata		Afro-Asiatic	Treis (2008, 2011, 2014)
		rel	Rendille		Afro-Asiatic	Oomen (1981)
		som	Somali		Afro-Asiatic	Saeed (1987); Serzisko (1992)
		tsb	Tsamai		Afro-Asiatic	Savà (2005, 2007)
	South Cushitic	irk	Iraqw		Afro-Asiatic	Mous (2003a)
Defoid		yor	Yoruba		Atlantic-Congo	Bamgbose (1966); Ajíbóyè (2010)
Dizoid		mdx	Dizin	Omotic ?	Afro-Asiatic ?	Beachy (2005)
Eastern Nilotic		kdj	Karamojong		Nilotic	Novelli (1985)
		mas	Masai		Nilotic	Payne (1998)
		tuv	Turkana		Nilotic	Dimmendaal (1983)
Hadza	**Hadza**	hts	Hadza	Khoisan		Edenmyr (2004); personal communication
Igboid		ibo	Igbo		Atlantic-Congo	Green & Igwe (1963)
Kxa		ktz	Ju|'hoan	Khoisan		Güldemann (2000); Heine & König (2013)

Table A.1: (continued)

Genealogical Unit	LGU	ISO code	Language	AG	HGU	Source
Khoe-Kwadi		hnh	ǁAni	Khoisan		Heine (1999)
		kwz	Kwadi	Khoisan		Güldemann (2004)
		naq	Nama	Khoisan		Hagman (1977)
		nhr	Naro	Khoisan		Vossen (1986)
		xuu	Kxoe	Khoisan		Kilian-Hatz & Heine (2010)
Kwa		aka	Akan		Atlantic-Congo	Appah & Amfo (2011)
		ewe	Ewe		Atlantic-Congo	Agbedor & Agbetsoamedo (forthcoming); Westermann (1930); Yvonne Agbetoamedo, personal communication
		snw	Sɛlɛɛ		Atlantic-Congo	Agbetsoamedo & Di Garbo (forthcoming); Yvonne Agbetoamedo, personal communication
Mande	**Western Mande**	bam	Bambara			Travélé (1955)
		dyu	Dyula			Delafosse (1901)
		sus	Susu			Lacan (1942)
	Eastern Mande	mev	Mann			Maria Khachaturyan, personal communication
Mel		gol	Gola	Atlantic	Atlantic-Congo	Fachner (1990)
		kss	Kisi	Atlantic	Atlantic-Congo	Childs (1983, 1995)
		tem	Timne	Atlantic	Atlantic-Congo	Wilson (1961)
North-Central Atlantic	**Central Atlantic**	bjg	Bidyogo	Atlantic	Niger-Congo	Segerer (2002)
		bqj	Bandial	Atlantic	Niger-Congo	Sagna (2008, 2011, 2012)
	North Atlantic	cou	Wamey	Atlantic	Atlantic-Congo	Santos (1996); Pozdniakov (2010); personal communication
		ffm	Maasina Fulfulde	Atlantic	Atlantic-Congo	Breedveld (1995)
		snf	Noon	Atlantic	Atlantic-Congo	Soukka (2000)
		srr	Serer	Atlantic	Atlantic-Congo	McLaughlin (1992)
		wol	Nuclear Wolof	Atlantic	Atlantic-Congo	McLaughlin (1997)
Sandawe	**Sandawe**	sad	Sandawe	Khoisan		Eaton (2010)
Semitic		amh	Amharic		Afro-Asiatic	Leslau (1995); Desalegn Hagos Asfawwesen, personal communication

Table A.1: (continued)

Genealogical Unit	LGU	ISO code	Language	AG	HGU	Source
Semitic		arb	Standard Arabic		Afro-Asiatic	Belnap & Shabaneh (1994); Belnap (1999); Ryding (2005)
		ary	Moroccan Arabic		Afro-Asiatic	Harrell (1965)
		heb	Hebrew		Afro-Asiatic	Coffin & Bolozky (2005); Glinert (1994); Grandi (2002)
		mlt	Maltese		Afro-Asiatic	Borg & Azzopardi-Alexander (1997); Grandi (2002)
		tig	Tigre		Afro-Asiatic	Leslau (1945, 1948)
		trg	Lishan Didan		Afro-Asiatic	Garbell (1965); Khan (2008)
South Omotic		dim	Dime	Omotic ?	Afro-Asiatic ?	Seyoum (2008)
Ta-Ne-Omotic		bcq	Bench	Omotic ?	Afro-Asiatic ?	Rapold (2006)
		kqy	Koorete	Omotic ?	Afro-Asiatic ?	Teketal (2004)
		mdy	Male	Omotic ?	Afro-Asiatic ?	Amha (2001)
		wal	Wolaytta	Omotic ?	Afro-Asiatic ?	Lamberti & Sottile (1997)
Tuu		nmn	!Xoo	Khoisan		Güldemann (2000); Traill (1994)
Western Nilotic		ach	Acoli		Nilotic	Storch (2005, 2007)
		bxb	Belanda Bor		Nilotic	Storch (2005)
		dik	Dinka (Southwestern)		Nilotic	Storch (2005, 2007)
		lwo	Luwo		Nilotic	Storch (2005, 2007)
		mfz	Mabaan		Nilotic	Storch (2005)
		nus	Nuer		Nilo	Storch (2005, 2007)

A.2 The language sample: alphabetical index

In table A.2, the languages of the sample are listed alphabetically. The language names are followed by the ISO codes, and the names of the genealogical units that each language is assigned to (see also table A.1).

Table A.2: The language sample: alphabetical index

Language	ISO code	Genealogical unit
Acoli	ach	Western Nilotic
Akan	aka	Kwa
Amharic	amh	Semitic
Awngi	awn	Cushitic
Bandial	bqj	North-Central Atlantic
Bafia	ksf	Bantu
Baiso	bsw	Cushitic
Bambara	bam	Mande
Beja	bej	Cushitic
Belanda Bor	bxb	Western Nilotic
Bemba	bem	Bantu
Bench	bcq	Ta-Ne-Omotic
Bila	bip	Bantu
Borana-Arsi-Guji Oromo	gax	Cushitic
Bidyogo	bjg	North-Central Atlantic
Chiga	cgg	Bantu
Daasanach	dsh	Cushitic
Dahalo	dal	Cushitic
Dibole	bvx	Bantu
Dime	dim	South Omotic
Dinka (Southwestern)	dik	Western Nilotic
Dirasha	gdl	Cushitic
Dizin	mdx	Dizoid
Dyula	dyu	Mande
Eton	eto	Bantu
Ewe	ewe	Kwa
Gidar	gid	Chadic
Gola	gol	Mel
Hadza	hts	Isolate
Hausa	hau	Chadic
Hdi	xed	Chadic
Hebrew	heb	Semitic
Igbo	ibo	Igboid
Iraqw	irk	Cushitic
Ju|'hoan	ktz	Kxa
Kabyle	kab	Berber
Kagulu	kki	Bantu
Kambaata	ktb	Cushitic
Karamojong	kdj	Eastern Nilotic
Kikuyu	kik	Bantu

Table A.2: (continued)

Language	**ISO code**	**Genealogical unit**
Kisi	kss	Mel
Koorete	kqy	Ta-Ne-Omotic
Kwadi	kwz	Khoe-Kwadi
Kxoe	xuu	Khoe-Kwadi
Lega	lea	Bantu
Lingala (Kinshasa)	lin	Bantu
Lele	lln	Chadic
Lishan Didan	trg	Semitic
Luwo	lwo	Western Nilotic
Masai	mas	Eastern Nilotic
Maasina Fulfulde	ffm	North-Central Atlantic
Mabaan	mfz	Western Nilotic
Makaa	mcp	Bantu
Male	mdy	Ta-Ne-Omotic
Maltese	mlt	Semitic
Mann	mev	Mande
Mina	hna	Chadic
Miya	mkf	Chadic
Mongo-Nkundu	lol	Bantu
Moroccan Arabic	ary	Semitic
Mwaghavul	sur	Chadic
Nafusi	jbn	Berber
Nama	naq	Khoe-Kwadi
Naro	nhr	Khoe-Kwadi
Ndengereko	ndg	Bantu
Noon	snf	North-Central Atlantic
Northern Sotho	nso	Bantu
Nuer	nus	Western Nilotic
Nyanja	nya	Bantu
Pero	pip	Chadic
Qimant	ahg	Cushitic
Rendille	rel	Cushitic
Sandawe	sad	Isolate
Sɛlɛɛ	snw	Kwa
Serer	srr	North-Central Atlantic
Shona	sna	Bantu
Somali	som	Cushitic
Standard Arabic	arb	Semitic
Susu	sus	Mande
Swati	ssw	Bantu
Swahili	swa	Bantu
Tachawit	shy	Berber
Tamasheq (Kidal)	taq	Berber
Tamazight (Central Atlas)	tzm	Berber
Tigre	tig	Semitic
Timne	tem	Mel

Table A.2: (continued)

Language	**ISO code**	**Genealogical unit**
Tonga	toi	Bantu
Tsamai	tsb	Cushitic
Tswana	tsn	Bantu
Tunen	baz	Bantu
Turkana	tuv	Eastern Nilotic
Venda	ven	Bantu
Wamey	cou	North-Central Atlantic
Wolaytta	wal	Ta-Ne-Omotic
Wolof (Nuclear)	wol	North-Central Atlantic
Yoruba	yor	Defoid
Zenaga	zen	Berber
Zulu	zul	Bantu
ǁAni	hnh	Khoe-Kwadi
!Xóô	nmn	Tuu

B Database coding sheet

Information about the individual languages of the sample has been stored in the database based on the following coding sheet.

B.1 Language data

- Language name
- ISO code
- Number of speakers
- African micro-area:
  - Eastern Africa
  - Central Africa
  - Northern Africa
  - Southern Africa
  - Western Africa
- Genealogical group
- Source

B.2 Gender systems

- Language ID = ISO code
- Type of gender system
  - Sex-based
  - Non-sex-based
  - No gender
  - No data
- Number of genders
  - Two genders
  - Three genders
  - Four genders
  - Five or more
  - No gender
  - No data

- Gender assignment
  - Semantic assignment
  - Semantic and formal assignment
  - No gender
  - No data
- Number of gender-indexing targets
  - One indexing target
  - Two indexing targets
  - Three indexing targets
  - Four or more
  - No gender
  - No data
- Occurrence of gender marking on nouns
  - Gender marking on nouns: Yes
  - Gender marking on nouns: No
  - No gender
  - No data
- Cumulation with number on agreement targets
  - Cumulative with all indexing targets
  - Cumulative with some indexing targets
  - Noncumulative
  - No gender
  - No number indexation
  - No data
- Cumulation with number on nouns
  - Cumulative
  - Cumulative SG vs. noncumulative PL
  - Noncumulative with specific gender (G) and number (N) markers
  - Noncumulative
  - No gender marking on nouns
  - No gender
  - No data
- Syncretism of gender across number
  - Yes
  - No
  - No data

B.3 Number systems

- Language ID = ISO code
- Type of number system in terms of obligatoriness
  - Obligatory number
  - General number
  - No number
  - No data
- Number values
  - Singular vs. plural
  - Singular vs. plural vs. dual
  - Singular vs. plural vs. dual vs. trial
  - Singular vs. plural vs. dual vs. paucal
  - No number
  - No data
- Number of number-indexing targets
  - One indexing target
  - Two indexing targets
  - Three indexing targets
  - Four or more indexing targets
  - No number indexation
  - No data

B.4 Evaluative morphology

- Language ID = ISO code
- Evaluative distinctions
  - Diminutive and augmentatives
  - Only diminutives
  - Only augmentatives
  - None
  - No data
- Type of diminutive markers
  - Suffix
  - Prefix
  - Infix
  - Stem change
  - Clitic
  - Analytical
  - Other (reduplication, phonosymbolic patterns etc...)

  - Different diminutive markers on different nouns
  - Co-occurring diminutive markers
  - Different diminutives on different nouns + co-occurring diminutives
  - No diminutives
  - No data
- Type of augmentative markers
  - Suffix
  - Prefix
  - Infix
  - Stem change
  - Clitic
  - Analytical
  - Other (reduplication, phonosymbolic patterns etc...)
  - Different augmentative markers on different nouns
  - Co-occurring augmentative markers
  - Different augmentatives on different nouns + co-occurring augmentatives
  - No augmentatives
  - No data
- Relations among gender and evaluation
  - Present
  - Absent
  - No gender
  - No data
- Number of diminutive genders
  - One
  - More than one
  - None
  - No data
- Number of augmentative genders
  - One
  - More than one
  - None
  - No data
- Type of gender shift in sex-based gender systems
  - M $\rightarrow$ F = Dim. and F $\rightarrow$ M = Aug.
  - M $\rightarrow$ F = Dim.
  - F $\rightarrow$ M = Aug.
  - M/F $\rightarrow$ N = Dim.
  - None
  - No data

B.5 Gender complexity

- Language ID = Iso code
- Number of gender values (GV)
  - Two = 0
  - Three = 1/3
  - Four = 2/3
  - Five or more = 1
  - No data = NA
- Nature of assignment rules (AR)
  - Semantic assignment = 0
  - Semantic and formal assignment = 1
  - No data = NA
- Number of indexing targets (IND)
  - One = 0
  - Two = 1/3
  - Three = 2/3
  - Four or more = 1
  - No data = NA
- Cumulative exponence of gender and number (CUM)
  - Noncumulative = 0
  - Partially cumulative = 1/2
  - Cumulative = 1
  - No data = NA
- Manipulation of gender assignment triggered by number/countability (M1)
  - Absent = 0
  - Present = 1
  - No data = NA
- Manipulation of gender assignment triggered by size (M2)
  - Absent = 0
  - Present = 1
  - No data = NA

C Gender- and number-indexing targets in the languages of the sample

Table C.1: Number of gender- and number-indexing targets in the languages of the sample

G. indexes	N. indexes	No. of lngs.	%	Genealogical groups
One	One	4	4%	Chadic (1/8) Khoe-Kwadi (1/5) Kxa (1/1) North-Central Atlantic (1/7)
Two	Two	15	15%	Bantu (2/23) Chadic (1/8) Cushitic (3/13) Khoe-Kwadi (4/5) Mel (1/3) South Omotic (1/1) Ta-Ne-Omotic (3/4)
Three	Three	25	25%	Bantu (3/23) Berber (5/6) Chadic (1/8) Cushitic (7/13) Eastern Nilotic (1/3) North-Central Atlantic (1/7) Sandawe (1/1) Semitic (5/7) Ta-Ne-Omotic (1/4)
Four or more	Four or more	34	34%	Bantu (17/23) Berber (1/6) Chadic (2/8) Cushitic (1/13) Eastern Nilotic (2/3) Hadza (1/1) Kwa (1/3) Mel (1/3) North-Central Atlantic (5/7) Semitic (2/7) Tuu (1/1)

Table C.1: (continued)

	G. indexes	**N. indexes**	**No. of lngs.**	*%*	**Genealogical groups**
	Three	Two	1	1%	Dizoid (1/1)
	Three	One	1	1%	Cushitic (1/13)
	Two	Three	1	1%	Cushitic (1/13)
	One	Two	1	1%	Chadic (1/8)
	No gender	One	6	6%	Igboid (1/1) Kwa (1/3) Mande (4/4)
	No gender	Two	4	4%	Chadic (2/8) Defoid (1/1) Kwa (1/3)
	No gender	No number indexation	6	6%	Western Nilotic (6/6)
	No data	No data	2	2%	Bantu (1/23) Mel (1/3)
Total		100	100%		

D Examples of lexical plurals in four Cushitic languages

Table D.1: Lexical plurals in Iraqw (based on Mous 2008)

Semantic groupings	**Nouns**
Abstract nouns	*aldafiri* 'interest, something returned with what was borrowed'; *di-idaa* 'boasting pride'; *fayda* 'profit'; *huwaa* 'burden'; *iilo* 'weight, load'; *loeemaa* 'truth'; *qatsuwa* 'heroic success (in hunting or war)'; *slaaħareri* 'aroma'; *tsaxwa* 'danger'; *waaqooda* 'hypochrisy'
Activities involving multiple actions/participants	*aai* 'journey'; *da'ri* 'witchcraft'; *gila* 'quarrel, fight'; *ibyaa* 'pointless activity with the hands'; *waayaa* 'work of different kind, not heavy, routine'
Animate and inanimate collectives	*haywa* 'term to address children'; *kumbeeri* 'women accompanying the bride's mother during the wedding cerimony'; *wa'ree* 'boys and girls accompanying the bride'; *makay* 'animals'; *laqaya* 'thorns'; *yakawaa/hikwaa* 'cattle'; *hurwa'i* 'bad maize grains'; *baynu* 'pigs'; *maanda* 'Nyiramba, Bantu (land and people)'; *afi* 'scraping of stiff porridge at the sides of the pot'
Body parts	*afeetlo* 'waist, loin'; *duunga* 'nose'; *gitsee'a* 'forehead, face'; *gwe'edo* 'buttocks'; *hayso* 'tail, penis'; *xxatli* 'afterbirth, placenta of an animal, trees'; *xaxardu* 'palate'; *da'awa* 'chest'
Diseases	*kuuko* 'mumps'
Location	*alu* 'behind, reverse'; *baray* 'down (on a slope), low, inside'; *dimbé* 'side, far, separate, different'; *tsee'a* 'outside'
Miscellaneous	*ayla* 'song improvised for the occasion'
Masses	*da'ata* 'red of blood'; *duwa* 'milk from plant'; *ilwa* 'milk'; *hinqeereeri* 'saliva'; *tsnuqaa* 'saliva as blessing, gifts in the form of money to newly wed'; *dara'ma* 'roasted meat and intenstines for the skinners'; *fu'naay* 'meat (for eating)'; *kund'i* 'bundle to carry, bale'; *ma'ay* 'water' *slaħoo* 'mucus'
Internally complex objects	*dara'ma* 'roasted meat and intestines for the skinners'
Places	*maraay* 'houses'; *doori* 'sky, heaven'; *irqwá da'áw* 'montaneous area southeast of Mbulu'; *siħú* 'far land'; *yaamu* 'earth, world below'; *uwa* 'west'
Periods of time	*ameetleemu* 'midday'; *aymadu* 'lunch time'; *amsi* 'midnight, night'; *axweeso* 'evening, night'; *baloqa* 'day after tomorrow'; *buhaaree* 'rainy season'; *de'ma* 'time'; *ki'ima* 'turn, time, coming back'; *matlo* 'tomorrow'; *xweeraa* 'night'; *tsiindo* 'evening (before dark)'

Table D.2: Lexical plurals in Borana-Arsi-Guji Oromo (based on Andrzejewski 1960: 70)

Semantic groupings	**Nouns**
Body parts	*fúnnaan* 'nose(s)'; *áfaan* 'mouth, mouths'; *ílkaan* 'tooth, teeth'; *tʃ'íɗaan* 'penis, penes'
Mass nouns	*ímimmaan* 'tear, tears', *údaan* 'faeces'; *fíntʃ'aan* 'urine'; *fôon* 'meat'; *bísaan* 'water'
Animate and inanimate collectives	*lôon* 'cattle'; *indzaan* 'lice' (or louse); *míɖaan* 'a berry, a grain, a fruit, *or* berries, grains, fruit'; *mítʃiraan* 'stones, stone'; *séep'an* 'leather straps'; *búusan* 'Pleiades'
Periods of time	*halkán* 'night'

Table D.3: Lexical plurals in Rendille (based on Oomen 1981: 51)

Semantic groupings	**Nouns**
Body parts	*marát* 'brains'; *sombób* 'lungs'
Animate collective	*biná* 'wild animals'; *ħolá* 'domesticated animals'
Mass nouns	*ħaanú* 'milk'; *bicé* 'water'; *oncabá* 'maize'; *sonxór* 'sugar'; *basbás* 'different kinds of milk mixed'; *dúubát* 'fog'; *banáy* 'light'
Miscellaneous	*anxád* 'lightning'

Table D.4: Lexical plurals in Baiso (based on Corbett & Hayward 1987: 9)

Semantic groupings	**Nouns**
Body parts	*ilkoo* 'tooth, teeth'; *kalaljaa* 'kidneys'; *lukk͡aa* 'foot, feet, leg(s)'; *il͡oo* 'eye(s)'; *ogorroo* 'hair'; *moo* 'hips, lumber region'
Animate collective	*saé* 'cattle'
Mass nouns	*eenoo* 'milk'; *soo* 'meat'; *udú* 'faeces'
Objects coming in pairs	*keferoo* 'sandals'

E Singular and plural suffixes in Mabaan, Dinka and Luwo

Table E.1: Singular markers in Mabaan, Dinka and Luwo (based on Storch 2005: 382)

Mabaan	**Dinka**[50]	**Luwo**	**Semantics**
-(C)ɪ̀	*V3-*		long, one-dimensional, dominant objects
	V:		
	-w		
-gɔn			derogative concepts
-tɪ̀	*-V:-*	*-ɔ̤*	singulative concepts
	[ˈ]	*-ɔ̤̀*	
	HL		
-n	*-V:-*	*-a*	general singular
	-V3-	*-ɔ̤*	
ɪ̀	*-V3-*		abstract concepts
	-o		
-gɪ̀		*-k*	abstract concept
-à	*-V:-*	*-ɔ̤*	spherical, round, small, objects; mass items; specialized people
	[ˈ]		
	-w		
	-V2-		
-i			fast moving objects
ɔ̀	*-o*	*-a*	locatives, domestic objects
-ù / Nù	*-V2-*	*-ʊ̤*	shape + possession, body, spatial orientation
	-V3-		
	[ˈ]		
-Nɪ̀	[ˈ]	*-a*	soft, circular objects
	-w		
	-l		
àn̤ /ɲàn̤	*-w*		part of a larger unit
-(C)in̤	*-l*		animacy, mass

[50] The cardinal numbers that appear next to the vocalic symbol *V* refer to the degrees of vowel alternation in Dinka. For a more detailed discussion on the grammatical functions of vowel alternation in Dinka, see Storch (2005: 165-166).

Table E.2: Plural markers in Mabaan, Dinka and Luwo (based on Storch 2005: 385)

Mabaan	**Dinka**	**Luwo**	**Semantics**
-kʌ̀	*-k* *-V2-* *-V3-* *-Vː-* [ˈ]	*-kʌ̀*	general plural
-(C)lŋ	*-N* *-V2-* *-V3-* *-Vː-* [ˈ]	*-V$_{-F+X}$-ɛ* *-VNɛ́*	general plural
áŋ		*ɛ̀*	round, mass, small
-kʊ̀		*ì*	body, space
ʌ́			semantically unspecified
-t(àn	*-t*	*-t*	semantically unspecified

F Complexity scores for the individual features in the metric

Table F.1 shows how the languages with gender in the sample scored with respect to the features of the complexity metric presented in chapter 7. Unlike in table 7.3, where the GCSs are rounded up to numbers with two decimal places, unrounded figures are provided in table F.1. The data are ordered alphabetically based on the ISO codes of the sampled languages. See table A.2 for the correspondent language names. The following abbreviations are used in the headings of the table:

GV = Number of gender values

AR = Nature of assignment rules

IND = Number of gender-indexing targets

CUM = Cumulation

M1 = Manipulation of gender assignment triggered by number/countability

M2 = Manipulation of gender assignment triggered by size

GCS = Gender Complexity Score

Table F.1: Complexity scores

ISO	GV	AR	IND	CUM	M1	M2	GCS
ahg	0	1	2/3	1	0		0.533333333
amh	0	1	1	1	0	1	0.666666667
arb	0	1	2/3	1	1	0	0.611111111
ary	0	1	2/3	1	1	0	0.611111111
awn	0	1	1/3	1	0	1	0.555555556
baz	1	1	2/3	1	1	1	0.944444445
bcq	1/3	1	1/3	1	0	1	0.611111111
bej	0	1	1	0	0	1	0.5
bem	1	1		1	1	1	1
bjg	1	1	1	1	1	1	1
bip	0	0	1/3	1	0	0	0.222222222
bqj	1	1	1	1	1	1	1
bsw	0	1	2/3	1/2	0	0	0.361111111
bvx	1	1	2/3	1	1	0	0.777777778
cgg	1	1	1	1	1	1	1
cou	1	1	1	0	1	1	0.833333333
dal	0	0	2/3	1	0	0	0.277777778

Table F.1: (continued)

ISO	GV	AR	IND	CUM	M1	M2	GCS
dim	0	0	1/3	1	0	1	0.388888889
dsh	0	1	1/3	1/2	0	1	0.472222222
eto	1	1	1	1	1	0	0.833333333
ffm	1	1	1	1	1	1	1
gax	0	1	2/3	1	0		0.533333333
gid	0	1	1	1	0	0	0.5
gdl	0	1	1/3	1	0		0.466666667
gol	1	1	1/3	1	0		0.666666667
hau	0	1	2/3	1	0	1	0.611111111
heb	0	1	2/3	1	0	0	0.444444445
hnh	1/3	1	1/3	1	0		0.533333333
hts	0	1	2/3	1	0	1	0.611111111
irk	0	1	2/3	1/2	0		0.433333333
jbn	0	1	2/3	1/2	1	1	0.694444445
kab	0	1	2/3	1/2	1	1	0.694444445
kdj	1/3	1	1	1	0	1	0.722222222
kik	1	1	1	1	1	1	1
kki	1	1	1	1	1	1	1
kqy	0	1	2/3	1/2	0	1	0.527777778
ksf	1	1	2/3	1	1	0	0.777777778
kss	1	1		1		0	0.75
ktb	0	1	2/3	1/2	0	0	0.361111111
ktz	2/3	1	0	1/2	0	0	0.361111111
kwz	0		0	1	0		0.25
lea	1	1	1	1	1	1	1
lin	1	1	1/3	1	0	0	0.555555556
lln	0	1	1/3	1	0		0.466666667
lol	1	1	1	1	1	1	1
mas	0	0	1	1	0	1	0.5
mcp	1	1	1	1	1	1	1
mdx	0	0	2/3	1	0	1	0.444444445
mdy	0	1	1/3	1	0	1	0.555555556
mkf	0	1	1	1	0		0.6
mlt	0	1	1	1	1	1	0.833333333
naq	1/3	1	1/3	1	0	1	0.611111111
nhr	1/3	1	1/3	1	0	1	0.611111111
ndg	1	1	1	1	1	1	1
nmn	1	1	1	1			1
nso	1	1	1	1	1	0	0.833333333
nya	1	1	1	1	1	1	1
pip	0		0	1/2	0		0.125
rel	0	1	1/3	1	0		0.466666667
sad	0	1	2/3	1	0	1	0.611111111
shy	0	1	1	1/2	1	1	0.75
sna	1	1	1	1	1	1	1
snf	1	1	0	1	0	1	0.666666667
snw	1	1	1	1	0	1	0.833333333
som	0	1	2/3	1/2	0		0.433333333
srr	1	1	1	1		1	1
ssw	1	1	1	1	1	0	0.833333333

Table F.1: (continued)

ISO	GV	AR	IND	CUM	M1	M2	GCS
sur	0	0	0	1/2	0	0	0.083333333
swa	1	1	1	1	1	1	1
taq	0	1	2/3	1/2	1	1	0.694444445
tem	1	1	1	1			1
tig	0	1	2/3	1	0	1	0.611111111
toi	1	1	1	1	1	1	1
trg	0	1	2/3	1	0		0.533333333
tsb	0	1	2/3	1	0	0	0.444444445
tsn	1	1	1	1	1	0	0.833333333
tuv	1/3	1	2/3	1	1	1	0.833333333
tzm	0	1	2/3	1/2	1	1	0.694444445
ven	1	1	1	1	1	1	1
wal	0	1	1/3	1	0	1	0.555555556
wol	1	1	2/3	1	0	1	0.777777778
xuu	0	1	1/3	1	0		0.466666667
zen	0	1	2/3	1/2	1	1	0.694444445
zul	1	1	1	1	1	0	0.833333333

Sammanfattning på svenska

Francesca Di Garbo
Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa

(Översättning från engelska av Eva Lindström)

Denna avhandling undersöker morfosyntaktisk och semantisk interaktion i genussystem med numerus och evaluativ markering (diminutiva och augmentativa konstruktioner).[51] Detta görs genom en kombination av tvärspråklig forskning och djupstudier av besläktade språk inom en kontinent, nämligen Afrika. Diakroniska och synkroniska infallsvinklar på interaktionen tas i beaktande. Begreppet interaktion används ganska löst i detta arbete. Faktum är, särskilt i afrikanska språk, att genus och numerus å ena sidan, och genus och evaluativ morfologi å den andra, kan vara så tätt sammanflätade att de inte bara interagerar utan visar tecken på att ha smält samman.

Avhandlingen centreras kring tre huvudsakliga forskningsfrågor:

(1) Interaktion mellan genus och numerus
   - Har genus och numerus samma markering? Hur fördelar sig genusdistinktioner över numerusdistinktioner?
   - Konkurrerar indexeringsmönstren för genus och numerus med varandra distributionsmässigt?
   - Har genus och numerus samma relevans för substantiv?
   - Interagerar genus och numerus semantiskt?

(2) Interaktion mellan genus och evaluativ morfologi:
   - Kan *storlek* vara en faktor i genustilldelning?
   - Hur interagerar diminutiva och augmentativa genus med de övriga genusdistinktionerna i ett språk?

(3) Interaktion mellan genus och grammatisk komplexitet:
   - Är det möjligt att mäta den grammatiska komplexiteten hos genussystem?
   - Kan ett sådant komplexitetsmått redogöra för interaktion mellan genus och andra grammatiska domäner?

[51] Begreppet genus används i denna avhandling på samma sätt som det oftast gör i typologisk litteratur, där genus och nominalklasser ses som ett och samma fenomen, och genus används som överordnad term.

– Vilken roll spelar sådan interaktion för den sammantagna komplexiteten hos ett genussystem?

Avhandlingen består av åtta kapitel. **Kapitel 1** utgörs av en kort presentation av de huvudsakliga undersökningstemana och en översikt över avhandlingens struktur.

Den teoretiska bakgrunden diskuteras i **kapitel 2**. I kapitlets första del presenteras de tre undersökta domänerna (genus, numerus och evaluativ morfologi) var för sig, med fokus på deras funktionella och semantiska egenskaper, strukturella uttryck och typologiska fördelning. I kapitlets andra del ges en översikt över tidigare litteratur om interaktion mellan genus och numerus, och genus och evaluativ morfologi, liksom begreppet absolut komplexitet och dess relevans för genussystem. Följande nyckelbegrepp introduceras i kapitlet:

Indexering används i denna avhandling i stället för *kongruens*, såsom det definieras av Corbett (2006). Det står för de grammatiska strategier som ett språk använder sig av (1) för att uttrycka lexikala och grammatiska egenskaper hos substantiv, eller (2) för att uttrycka semantiska egenskaper hos NP-referenter, i båda fallen *oavsett* om en syntaktisk antecedent eller nominalt huvudord finns overt uttryckt i kontexten. Termen *indexeringsmottagare* (indexing target) används om konstituenter i kontexten vars böjningsmorfologi uttrycker genus och numerus (t.ex. adjektiv, determinerare, verb, pronomen osv.). *Indexeringsutdelare* (indexation trigger) används för de konstituenter som aktiverar användningen av ett visst indexeringsmönster (t.ex. substantiv och pronomen) inom en viss kontext.

Exponens definieras som betydelsemängden per morfologisk enhet. Exponens är kumulativ när betydelser härrörande till åtminstone två grammatiska domäner realiseras i ett och samma morfem. Genus och numerus har ofta kumulativ exponens.

Synkretism definieras som en typ av paradigmatisk asymmetri där vissa grammatiska distinktioner neutraliseras eller reduceras givet vissa betingande faktorer. Det är typologiskt vanligt att genusdistinktioner neutraliseras eller reduceras när ett substantiv inte står i singularis.

Manipulerbar vs. strikt genustilldelning: genustilldelning är manipulerbar när substantiv kan tilldelas mer än ett genus, beroende på hur NP-referenten betraktas. Genustilldelning är strikt när varje substantiv konsekvent tillhör ett enda genus.

Absolut komplexitet definieras som antalet distinktioner inom en grammatisk domän, eller längden på dess beskrivning.

Studiens urvalsmetodologi och hur data samlats in och behandlats presenteras i **kapitel 3**. Avhandlingen bygger på ett urval om 100 afrikanska språk, varav 84 med genus och 16 utan genus. Urvalet utformades med avsikten att kunna göra jämförelser både inom och mellan språkfamiljer, och består därför av flera delurval inom skilda genetiska grupper (t.ex. bantu, kushitiska och berber).

Urvalet är snedvridet i åtminstone två avseenden: släktskapsmässigt och bibliografiskt. Det kan ses som släktskapsmässigt snedvridet för att det bygger på relationer mellan undergrupper av inbördes besläktade språk inom en avgränsad del av världen. Det är bibliografiskt snedvridet då urvalet inom varje språkfamilj gjordes på basis av tillgänglig litteratur snarare än med matematiska metoder, även om alla undergrenar är representerade. Dessa två typer av snedvridningar inverkar på analysen såtillvida att urvalet inte kan användas för att göra statistiskt grundade förutsägelser om dominerande typologiska mönster, varken inom eller utanför den afrikanska makroregionen. Denna studie syftar heller inte till att göra statistiska förutsägelser. Målet är i stället att redogöra för förekomst, stabilitet och fördelning av grammatiska fenomen som rör interaktionen mellan genus och numerus, och mellan genus och evaluativ morfologi. Jag har utgått ifrån att dessa mål bäst kan uppnås genom att se på ett flertal grupper av inbrdes besläktade språk inom en enda världsdel. För att strukturera data från de olika språken i urvalet skapades en relationell databas. Vad gäller släktskapsklassificeringen använder sig denna studie av *Glottolog* (Nordhoff et al. 2013).

Kapitel 4 ger en översikt över de system för genus, numerus och evaluativ morfologi som påträffades i de undersökta språken. De parametrar som använts för var och en av de tre domänerna i klassificeringen av databasen presenteras och diskuteras. Frekvensfördelningen för de värden som hör ihop med de respektive parametrarna ges också i kapitlet.

Interaktion mellan genus och numerus diskuteras i **kapitel 5**. Kapitlet har fyra delar. I den första delen undersöker jag exponensmönster hos genus och numerus, samt numerusbetingad synkretism i genusmarkering. Exponens och synkretism ses ofta som två grundläggande dimensioner i interaktionen mellan genus och numerus. Såvitt jag känner till utgör dock denna avhandling den första systematiska genomgången av exponensmönster och synkretism för genus och numerus som grundar sig på ett så pass stort urval av språk. Kumulativ exponens, där genus och numerus uttrycks i en och samma form, är extremt vanlig i de undersökta språken, och det gäller även synkretism. Eftersom inte alla språk med kumulativ exponens har synkretism, men nästan alla språk med synkretism har kumulativ exponens, framlägger jag hypotesen att numerusbetingad genussynkretism förutsätter kumulativ markering av dessa två grammatiska domäner. Jag menar vidare att kumulativitet bör ses som en faktor som ökar den absoluta komplexiteten i genussystem, såtillvida att det inför brus överensstämmelsen mellan form och betydelse. Slutligen argumenterar jag för att mina resultat rörande kumulativitet och synkretism kan förstås som möjliga indikatorer på att det finns en relevanshierarki bland nominala grammatiska särdrag. I överensstämmelse med Greenberg (1963a), Carstairs (1987), Carstairs & Stemberger (1988), and (Vafaeian 2013) - som alla studerade olika grammatiska fenomen och använde sig av olika dataset - visar dessa resultat att om det finns en nominal relevanshierarki så skulle numerus vara den högst rankade kategorin.

I kapitlets andra del fokuserar jag på de konkurrerande mönster för genus- och numerusindexering som förekommer inom ett relativt litet antal språk inom det större urvalet. Dessa språk hör alla till det afroasiatiska fylumet och uppvisar ett tudelat,

könsbaserat genussystem. De har dessutom en delning av indexeringsmönstren på substantiv i pluralis. Vissa substantiv i pluralis aktiverar samma indexering som antingen maskulinum singularis eller femininum singularis, medan andra substantiv i pluralis ger upphov till indexering som skiljer sig både från maskulinum och femininum singularis. Eftersom det senare indexeringsmönstret bara används för att uttrycka egenskapen pluralis kallar jag det *Specialiserad pluralindexering* (Dedicated Plural Indexation, DPI). Jag föreslår vidare benämningen språk med *delat pluralindexeringssystem* (languages with split plural indexation) för språk av denna typ. Jag visar hur användningen av DPI kan begränsas antingen av animacitet (t.ex. i modern standardarabiska), eller av lexikal pluralitet hos huvudordet (kushitiska).

En fallstudie utfördes på de kushitiska språken inom det större urvalet för att kartlägga hur olika språk inom gruppen beter sig vad gäller användandet av DPI. I litteraturen om kushitiska språk beskrivs språk med relativt begränsad användning av DPI traditionellt som att de har en tredje indexeringsklass, alltså ett tredje genus, dvs. pluralis förutom maskulinum och femininum (se Mous 2008 för en översikt). Detta synsätt har mött stark kritik i mer typologiskt inriktad litteratur (se bl.a. Corbett 2000; Corbett & Hayward 1987). I dessa studier har dock DPI ofta setts bara som ett undantag, utan något försök till förklaring av dess säregenhet (ett undantag är Corbett 2012 som försöker förklara genus- och numerusindexering i Baiso genom att i stället fokusera på den roll som lexikal pluralitet spelar). Min studie visar att en övergripande alternativ förklaring till detta länge debatterade fenomen blir möjlig om man ser till hur DPI fördelar sig över olika språk inom familjen och undersöker de faktorer som påverkar dess förekomst. I kushitiska språk med delad pluralindexering är DPI bara möjlig om indexeringsutdelaren har en inneboende koppling till pluralitet. Detta innebär inte att DPI är ett morfosyntaktiskt uttryck för genus. Det visar bara att två domäner kan konkurrera med varandra: en obligatoriskt uttryckt genom indexering (genus), och den andra potentiellt uttryckt genom indexering (numerus). Jag visar att arbetsfördelningen dem emellan görs upp på olika sätt i olika kushitiska språk. I de flesta språken indexeras nominal pluralitet alltid, antingen den är morfologisk eller lexikal; i andra språk indexeras bara lexikal pluralitet.

I kapitlets tredje del diskuterar jag korrelationer mellan förekomst av genus och typ av numerusmarkering. Jag visar hur framväxten av genomgripande indexeringssystem i språken i urvalet alltid förutsätter att det finns både genus och numerus, och att språken i denna studie tenderar att uppvisa samma antal indexeringsmottagare för genus som för numerus.

I den fjärde och sista delen av kapitlet visar jag att genusbyte i språk med manipulerbar genustilldelning kan användas för att uttrycka räknebarhet. Detta fenomen har en väldigt sned fördelning i undersökningens språk, i och med att det bara påträffas inom familjerna bantu, berber, nord-centrala atlantiska och semitiska, samt i ett östnilotiskt språk.

I **kapitel 6** undersöker jag morfosyntaktiska och semantiska interaktioner mellan genus och den morfologiska kodningen av diminutiver och augmentativer. Kapitlet syftar särskilt till att undersöka (1) hur vanligt det är att *storlek* fungerar som ett genus för sig, och (2) hur olika typer av genussystem (könsbaserade kontra icke-könsbaserade) och

strategier för genustilldelning (strikt kontra manipulerbart) ger förutsättningarna för de mönster för interaktion mellan genus och evaluativ morfologi som kunnat observeras. Genom detta kapitel bidrar denna avhandling med den första större typologiska redogörelsen för hur markering av *storlek* ('stor' eller 'liten') kan inlemmas i ett språks genussystem. Resultaten tyder på att arten av interaktion mellan genus och evaluativ morfologi varierar enligt vilken typ av genussystem ett språk har. De relevanta variablerna är: (1) typ av genussystem (könsbaserat kontra icke-könsbaserat), (2) antalet distinktioner inom systemet, och (3) graden av manipulerbarhet i genustilldelning.

Två typer av interaktion mellan genus och evaluativ morfologi förekommer i undersökningsspråken. Språk med stora genussystem har typiskt specialiserade evaluativa genus som används för att uttrycka diminutiva och augmentativa betydelser. I språk med mindre och könsbaserade genussystem finns å andra sidan inga specialiserade evaluativa genus, utan maskulinum, femininum eller neutrum kan anvïdas för att uttrycka diminutiva och augmentativa betydelser när ett ord används med ett annat genus än det brukar ha. Båda typerna har bred spridning i mina undersökningsspråk, även över språkfamiljsgränser. I vissa fall är det möjligt att uppskatta deras stabilitet, såväl som delar av deras framväxt diakroniskt. Data visar också att även om storlek kan vara ett produktivt kriterium för substantivklassificering är det aldrig centralt framträdande, varken synkront eller diakront, i de två typer av genussystem som undersökningsspråken företer. Slutligen begränsas möjligheten att uttrycka evaluativa betydelser genom genusbyte i språk med könsbaserat genus av NP-referentens animacitet. För animata substantiv, där könstillhörighet är den mest framträdande egenskapen, används genusbyte bara för att uttrycka biologiska genusskillnader. När det gäller inanimata substantiv är könstillhörighet irrelevant och genusbyte används för att uttrycka skillnader i *storlek*.

Kapitel 7 syftar till att mäta den effekt som interaktioner mellan genus och numerus, och mellan genus och evaluativ morfologi, har på ett genussystems komplexitet. Resultaten tyder på att ju mer interaktiv en grammatisk domän är, desto högre är komplexiteten. Jag använder begreppet grammatisk komplexitet i dess absoluta bemärkelse, dvs. som en objektiv egenskap hos en grammatisk domän. Måttet på absolut komplexitet kan grunda sig antingen på antalet delar som ingår i ett system, eller längden på systemets beskrivning. Det bör noteras att mått på absolut komplexitet bara kan vara lokala, i meningen att de bör gälla specifika grammatiska domäner snarare än hela språk. Miestamo (2008) och Sinnemäki (2011) lägger fram två grundläggande principer som allmänna riktlinjer för utveckling av komplexitetsmått inom vilken grammatisk domän det än vara månde:

> *Principen om färre distinktioner.* Ju högre antalet grammatikaliserade distinktioner är, desto komplexare är domänen.
>
> *Principen om en betydelse-en form.* Den minst komplexa domänen är den där det råder ett-till-ett förhållande mellan betydelse och form.

I kapitel 7 argumenterar jag för att dessa två principer inte räcker för att redogöra för hela den vidd av fenomen som kan vara relevanta för studiet av språklig komplexitet.

Jag menar att en tredje princip bör läggas till för att hantera interaktioner mellan grammatiska domäner. Jag kallar denna princip *Principen om oberoende*.

> *Principen om oberoende.* Den minst komplexa domänen är den vars beskrivning är oberoende av semantiska och funktionella egenskaper hos andra domäner.

Att utveckla mått för att avgöra den absoluta komplexiteten hos grammatiska domäner är ett relativt nytt område inom språktypologi. Vad gäller genus har det behandlats i banbrytande arbete av Audring (2014). Författaren föreslår tre dimensioner längs vilka genuskomplexitet kan beräknas, men ger ingen metod för att överföra dimensionernas variabler till mätbara värden. I kapitel 7 utvecklar jag Audrings modell för genuskomplexitet, med två huvudsakliga syften:

(1) att redogöra för de morfosyntaktiska och semantiska interaktioner mellan genus och numerus och mellan genus och evaluativ morfologi som utforskas i avhandlingen.

(2) att omvandla de värden som hör ihop med de respektive dimensionerna till tal.

Det komplexitetsmått som utarbetades för att uppnå dessa syften består av sex komponenter: GV (antal genusdistinktioner), AR (antal regler för genustilldelning), IND (antal indexeringsmottagare för genus), CUM (förekomst av kumulation av genus och numerus), M1 (manipulering av genustilldelning som styrs av numerus/räknebarhet), M2 (manipulering av genustilldelning som styrs av storlek). För att omvandla värdena för var och en av parametrarna till siffror använder jag en metod utvecklad av Parkvall (2008).

Resultaten av uträkningarna pekar på: (1) att komplexitetsgraden är knuten till språkfamilj, (2) att språk med samma komplexitetsgrad kan uppnå den på olika sätt, (3) att vissa av parametrarna i beräkningsmetoden kan interagera implikationellt med varandra (som att strikt semantisk genustilldelning bara färekommer i språk med små genussystem), och (4) att vissa av måttets komponenter (i synnerhet GV, IND och M1) korrelerar starkt med varandra och förefaller ha en större inverkan på den sammantagna komplexitetsgraden i undersökningsspråken. Slutligen visar resultaten att språk i mitt urval som har genus oftast har relativt komplexa genussystem medan genussystem med låg komplexitetsgrad är sällsynta. Om detta urval kan anses representativt för den afrikanska makroregionen så visar resultaten att genus är ett komplext drag i afrikanska språk.

Kapitel 8 ger en sammanfattning av avhandlingen, utvärderar den metodologi som använts, och skisserar möjliga teman för framtida forskning. Behov av vidare forskning identifieras särskilt för fyra områden: nominala relevanshierarkier; system med delad pluralindexering; genustilldelning och dess manipulering; och absolut komplexitet i genussystem. Morfosyntaktiska och semantiska interaktioner mellan enskilda grammatiska drag har utforskats grundligt inom domänen verb, med resultatet att drag som tempus, aspekt, modus och evidentialitet numera betraktas som en sammanhängande helhet, både på deskriptiv och teoretisk nivå. De resultat som redovisas i denna avhandling visar på några av de fördelar som kan nås genom att utvidga detta angreppssätt

till den nominala domänen och se på nominala drag i termer av deras morfosyntaktiska, semantiska och funktionella interaktion, snarare än som isolerade grammatiska domäner.

Bibliography

Acquaviva, Paolo. 2008. *Lexical Plurals: A morphosemantic approach*. Oxford: Oxford University Press.

Agbedor, Paul & Yvonne Agbetsoamedo. forthcoming. Ewe. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh handbook of evaluative morphology*, Edinburgh University Press.

Agbetsoamedo, Yvonne. 2014. Noun classes in Sɛlɛɛ. *The Journal of West African Languages* 41. 95–124.

Agbetsoamedo, Yvonne & Francesca Di Garbo. forthcoming. Selee. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh handbook of evaluative morphology*, Edinburgh University Press.

Aikhenvald, Alexandra. 2003. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

Aikhenvald, Alexandra. 2012. Round women and long men: Shape, size, and the meanings of gender in New Guinea and beyond. *Anthropological Linguistics* 54. 33–86.

Ajíbóyè, Ọládípọ̀. 2010. Plural strategies in Yorùbá. *Studies in African Linguistics* 39. 141–181.

Allan, Keith. 1977. Classifiers. *Language* 53. 285–311.

Allan, Keith. 1980. Nouns and countability. *Language* 56. 541–567.

Amha, Azeb. 2001. *The Maale language*. Leiden: Research School of Asian, African and Amerindian Studies.

Amha, Azeb. 2012. Omotic. In Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*, 423–504. Cambridge: Cambridge University Press.

Andrzejewski, B. W. 1960. The categories of number in noun forms in the Borana dialect of Galla. *Africa* 30. 62–75.

Appah, Clement K. I. & Nana Aba Appiah Amfo. 2011. The morphopragmatics of the diminutive morpheme (-ba/-wa) in Akan. *Lexis* 6. 85–102. [Special Issue: Diminutives and augmentatives in the languages of the world].

Appleyard, D. L. 1975. A descriptive outline of Kemant. *Language* 38. 316–350.

Audring, Jenny. 2009. Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology* 18. 93–116.

Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17. [Special issue: Exploring grammatical gender].

Baerman, Matthew, Dunstan Brown & Greville Corbett. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge: Cambridge University Press.

Bakker, Dik. 2011. Language sampling. In Jae Jung Song (ed.), *Handbook of linguistic typology*, 100–127. Oxford: Oxford University Press.

Bamgbose, Ayo. 1966. *A grammar of Yoruba*. Cambridge: Cambridge University Press.

Barlow, Michael. 1992. *A situated theory of agreement*. New York: Routledge.

Barlow, Ruffell. 1951. *Studies in Kikuyu grammar and idiom*. Edimburgh: William Blackwood.

Bauer, Laurie. 1996. No phonetic iconicity in evaluative morphology. *Studia Linguistica* 50. 189–206.

Bauer, Laurie. 1997. Evaluative morphology: In search of universals. *Studies in Language* 21. 533–575.

Bauer, Laurie. 2002. What you can do with derivational morphology. In O. E. Pfeiffer & M. Voeikova (eds.), *Morphology 2000: Selected papers from the 9th Morphology Meeting, Vienna 24-28 February 2000*, 37–48. Amsterdam: John Benjamins.

Beachy, Marvin Dean. 2005. *An overview of Central Dizin phonology and morphology*. Arlington: The University of Texas at Arlington dissertation.

Beguinot, Francesco. 1942. *Il berbero di Nefûsi di Fassâṭo*. Roma: Istituto per l'Oriente.

Belnap, Kirk R. 1999. A new perspective on the history of Arabic variation in marking agreement with plural heads. *Folia Linguistica* 33. 169–185.

Belnap, Kirk R. & Osama Shabaneh. 1994. Variable agreement and non-human plurals in Classical and Modern Standard Arabic. In Ellen Broselow, Mushira Eid & John McCarthy (eds.), *Perspective on Arabic linguistics IV*, 245–262. Amsterdam/Philadelphia: John Benjamins.

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic despription*, vol. 3: Grammatical categories and the lexicon, 169–240. Cambridge University Press.

Bokamba, E. 2009. The spread of Lingala as a lingua franca in the Congo Basin. In Fiona McLaughlin (ed.), *The languages of urban Africa*, 50–70. London: Continuum.

Bokamba, Eyamba. 1977. The impact of multilingualism on language structures: the case of Central Africa. *Anthropological Linguistics* 19. 181–202.

Bolozky, Shmuel. 1994. On the formation of diminutives in Modern Hebrew morphology. *Hebrew Studies* 35. 47–62.

Borg, Albert & Marie Azzopardi-Alexander. 1997. *Maltese*. London and New York: Routledge.

Botne, Robert. 2003. Lega (Beya dialect) (D25). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 422–449. London: Routledge.

Breedveld, J. O. 1995. *Form and meaning in Fulfulde: A morphological study of Maasinankoore*. Leiden: Research School CNWS.

Bybee, Joan. 1985. *Morphology*. Amsterdam: John Benjamins.

Bybee, Joan. 1994. The grammaticization of zero: Asymmetries in tense and aspect systems. In William Pagliuca (ed.), *Perspectives on grammaticalization*, 235–254. Amsterdam: John Benjamins.

Carrell, Patricia. 1970. *A transformational grammar of Igbo*. Cambridge: Cambridge University Press.

Carstairs, Andrew. 1984. Outlines of a constraint on syncretism. *Folia Linguistica* 18. 73–85.

Carstairs, Andrew. 1987. *Allomorphy in inflection*. London: Croom Helm.

Carstairs, Andrew & Joseph Paul Stemberger. 1988. A processing constraint on inflectional homonymy. *Linguistics* 26. 601–617.

Carter, Hazel. 2002. *An outline of Chitonga grammar*. Lousaka: Bookwords Publishers.

Childs, T. 1995. *A grammar of Kisi*. Berlin: Mouton de Gruyter.

Childs, Tucker. 1983. Noun class affix renewal in Southern West Atlantic. In Jonathan Kaye, Hilda Koopman, Dominique Sportiche & Andr Dugas (eds.), *Current approaches to African linguistics II*, 17–29. Dodrecht: Mouton de Gruyter; Foris Publications.

Clements, G. N. & Annie Rialland. 2008. Africa as a phonological area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 36–85. Cambridge: Cambridge University Press.

Coffin, Edna A. & Shmuel Bolozky. 2005. *A reference grammar of Modern Hebrew*. Cambridge: Cambridge University Press.

Cole, Desmond T. 1955. *An introduction to Tswana grammar*. Cape Town: Longmans Green.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51. 406–18.

Contini-Morava, Ellen. 1997. Noun classification in Swahili. In Robert K. Herbert (ed.), *African linguistics at the crossroads*, 599–628. Cologne: Rüdiger Köppe Verlag.

Contini-Morava, Ellen. 2000. Noun class as number in Swahili. In Hellen Contini-Morava & Yishai Tobin (eds.), *Between grammar and lexicon*, 1–30. Amsterdam: John Benjamins.

Corbett, Greville. 1991. *Gender*. Cambridge: Cambridge University Press.

Corbett, Greville. 1998. Prototypical inflection: implications for typology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1998*, 1–22. Dordrecht: Kluwer.

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press.

Corbett, Greville. 2006. *Agreement*. Cambridge: Cambridge University Press.

Corbett, Greville. 2012. *Features*. Cambridge: Cambridge University Press.

Corbett, Greville. 2013a. Gender typology. In Greville Corbett (ed.), *The expression of gender*, 87–130. Berlin: Mouton de Gruyter.

Corbett, Greville. 2013b. Number of genders. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, chapter 30. Available online at: http://wals.info/chapter/30. Accessed on 2014-02-14.

Corbett, Greville. 2013c. Sex-based and non-sex-based gender systems. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, chapter 31a. Available online at: http://wals.info/chapter/31. Accessed on 2014-02-14.

Corbett, Greville. 2013d. Systems of gender assignment. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: http://wals.info/chapter/32. Accessed on 2014-02-14.

Corbett, Greville & Richard J. Hayward. 1987. Gender and number in Bayso. *Lingua* 72. 195–222.

Corbett, Greville & Alfred Mtenje. 1987. Gender agreement in Chichewa. *Studies in African Linguistics* 72. 195–222.

Creissels, Denis. 1999. Origin et evolution des diminutifs et augmentatifs dans quelques langues africaines. *Silexicales* 2. 29–35.

Creissels, Denis, Gerrit J. Dimmendaal, Zygmut Frajzyngier & Christa König. 2008. Africa as a morphosyntactic area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 86–150. Cambridge: Cambridge University Press.

Crisma, Paola, Lutz Marten & Rint Sybesma. 2011. The point of Bantu, Chinese and Romance nominal classification. *Italian Journal of Linguistics* 23. 251–299.

Cristofaro, Sonia. 2012. Typological universals in synchrony and diachrony: The evolution of number marking. Paper presented at the 45th meeting of the *Societas Linguistica Europea*, Stockholm, September 2012.

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45. 145–171.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Longman.

Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Croft, William. 2013. Agreement as anaphora, anaphora as coreference. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: studies in memory of Anna Siwierska*, 107–129. Berlin: Mouton de Gruyter.

Dahl, Östen. 1990. Standard Average European as an exotic language. In Johannes Bechert, Giuliano Bernini & Claude Buridant (eds.), *Towards a typology of European languages*, 3–8. Berlin: Mouton de Gruyter.

Dahl, Östen. 2000a. Animacy and the notion of semantic gender. In Barbara Unterbeck, Matti Rissanen, Tettu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*, 99–115. Berlin: Mouton de Gruyter.

Dahl, Östen. 2000b. Elementary gender distinctions. In Barbara Unterbeck, Matti Rissanen, Terttu Nevalaien & Mirja Saari (eds.), *Gender in grammar and cognition*, 577–593. Berlin: Mouton de Gruyter.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.

Dahl, Östen. 2006. Diminutives and augmentatives. In Keith Brown (ed.), *Encyclopedia of language and linguistics*, 594–595. Oxford: Elsevier.

Dahl, Östen. 2011. Grammaticalization and linguistic complexity. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 153–162. Oxford: Oxford University Press.

Dahl, Östen & Maria Koptjevskaja-Tamm. 2001. Kinship in grammar. In Irene Baron, Michael Herslund & Finn Sørensen (eds.), *Dimensions of possession*, 201–225. Amsterdam: John Benjamins.

Dammel, Antje & Sebastian Kürschner. 2008. Complexity in nominal plural allormorphy. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 243–262. Amsterdam: John Benjamins.

Daniel, Michael & Edith Moravcsik. 2013. The associative plural. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: http://wals.info/chapter/36. Accessed on 2014-02-14.

De Mauro, Tullio, Federico Mancini, Massimo Vedovelli & Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etatslibri.

Delafosse, Maurice. 1901. *Essai de manuel pratique de la langue mande ou mandingue*. Paris: Ernest Leroux.

Denny, Peter & A. Chet Creider. 1976. The semantics of noun classes in Proto-Bantu. *Studies in African Linguistics* 7. 1–30.

Desmuth, Katherine. 2000. Bantu noun class systems: loanword and acquisition evidence of semantic productivity. In Senft Gunther (ed.), *Systems of noun classification*, 270–92. Cambridge: Cambridge University Press.

Dickens, Patrick. 1992. *Ju|'hoan grammar. ms.* Nyae Development Foundation of Namibia.

Diessel, Holger. 1999a. *Demonstratives: Form, function and grammaticalization*. Amsterdam: John Benjamins.

Diessel, Holger. 1999b. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3. 1–49.

Dimmendaal, Gerrit. 1983. *The Turkana language*. Dordrecht: Foris Publications.

Dimmendaal, Gerrit. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42. 214–261.

Dimmendaal, Gerrit. 2008. Language ecology and linguistic diversity on the African continent. *Language and Linguistics Compass* 2. 840–858.

Dixon, Robert M. W. 1982. *Where have all the adjectives gone? And other essays in semantics and syntax*. Berlin: Mouton de Gruyter.

Dixon, Robert M. W. 2004. Adjective classes in typological perspective. In *Adjective classes: A cross-linguistic typology*, 1–45. Oxford: Oxford University Press.

Dressler, Wolfgang & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German, and other languages*. Berlin: Mouton de Gruyter.

Dryer, Matthew. 1989a. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292.

Dryer, Matthew. 1989b. Plural words. *Linguistics* 27. 865–895.

Dryer, Matthew & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at http://wals.info, Accessed on 2014-02-14.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.

Dryer, Matthew S. 2013. Coding of nominal plurality. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: http://wals.info/chapter/33. Accessed on 2014-02-14.

Eaton, Helen. 2010. *A Sandawe grammar.* Dallas: SIL International.

Edenmyr, Niklas. 2004. The semantics of Hadza gender assignment: a few notes from the field. *Africa & Asia* 4. 3–19.

Fachner, Regine. 1990. The noun phrase in Gola. *Afrikanistische Arbeitspapiere* 23. 5–52.

Ferry, Marie-Paule & Konstantin Pozdniakov. 2001. Dialectique du régulier et de l'irrégulier: le système de classes nominales dans le groupe tenda des langues atlantiques. In Robert Nicolaï (ed.), *Leçons d'Afrique: Filiations, ruptures et reconstitution de langues. Un hommage à Gabriel Manessy*, 153–167. Paris: Peeters.

Fleming, Harold. 1969. The classification of West Cushitic within Hamito-Semitic. In Daniel McCall, Norman Bennet & Jeffrey Butler (eds.), *Eastern African history*, 3 – 27. New York: Praeger.

Foley, W. A. & R. Van Valin. 1984. *Functional syntax and universal grammar.* Cambridge: Cambridge University Press.

Fortune, G. 1955. *An analytic grammar of Shona.* London: Longmans Green.

Frajzyngier, Zygmunt. 1989. *A grammar of Pero.* Berlin: Dietrich Reimer Verlag.

Frajzyngier, Zygmunt. 1993. *A grammar of Mupun.* Berlin: Dietrich Reimer Verlag.

Frajzyngier, Zygmunt. 1997. Grammaticalization of number: From demonstratives to nominal and verbal plural. *Linguistic Typology* 1. 193–242.

Frajzyngier, Zygmunt. 2001. *A grammar of Lele.* Standford, CA: CSLI Publications: Center for the Study of Language and Information.

Frajzyngier, Zygmunt. 2008. *A grammar of Gidar.* Frankfurt am Main: Peter Lang.

Frajzyngier, Zygmunt & Eric Johnston. 2005. *A grammar of Mina.* Berlin: Mouton de Gruyter.

Frajzyngier, Zygmunt & Erin Shay. 2002. *A grammar of Hdi*. Berlin: Mouton de Gruyter.

Garbell, Irene. 1965. *The Jewish Neo-Aramaic dialect of Persian Azerbaijan*. Berlin: Mouton de Gruyter.

Gil, David. 1996. Maltese 'collective nouns': A typological perspective. *Rivista di Linguistica* 8. 53–87.

Gil, David. 2013. Numeral classifiers. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, chapter 55. Available online at: http://wals.info/chapter/55. Accessed on 2014-02-14.

Givón, Talmy. 1971. Some historical changes in the noun class system of Bantu. In C. W. Kim & S. Herbert (eds.), *Papers in African linguistics*, 33–54. Alberta: Linguistic Research Inc.

Glinert, Lewis. 1994. *Modern Hebrew: An essential grammar*. London and New York: Routledge.

Grandi, Nicola. 2001. Su alcune presunte anomalie della morfologia valutativa: Il rapporto con il genere ed il numero. *Archivio Glottologico Italiano* 1. 25–56.

Grandi, Nicola. 2002. *Morfologie in contatto: Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: Franco Angeli.

Grandi, Nicola. 2011. Renewal and innovation in the emergence of Indo-European evaluative morphology. *Lexis* 6. 5–25. [Special Issue: Diminutives and augmentatives in the languages of the world].

Grandi, Nicola. forthcoming-a. Berber languages. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh handbook of evaluative morphology*, Edinburgh University Press.

Grandi, Nicola. forthcoming-b. Evaluative morphology and number/gender. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Handbook of evaluative morphology*, Edinburgh University Press.

de la Grasserie, Raoul. 1898. La catégorie psychologique de la classification: révélée par la langage. *Revue de Philolosophique de la France et de l'Étranger* 45. 594–624.

Green, M. M. & Rev. G. E. Igwe. 1963. *A descriptive grammar of Igbo*. Berlin: Akademie-Verlag. KB.

Greenberg, Joseph. 1959. Africa as a linguistic area. In William Bascom & Melville Herskovits (eds.), *Continuity and change in African cultures*, 15–27. Chicago: Chicago University Press.

Greenberg, Joseph. 1963a. *The languages of Africa*. Bloomington: Indiana University.

Greenberg, Joseph. 1963b. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Greenberg, Joseph. 1978. How does a language acquire gender markers? In Joseph Greenberg, Charles Ferguson & Edith Moravcisk (eds.), *Universals of human language*, vol. 3: Word structure, 47–92. Stanford: Standford University Press.

Greenberg, Joseph. 1980. Circumfixes and typological change. In Elisabeth Traugott (ed.), *Papers from the 4th International Conference on Historical Linguistics*, 233–241. Amsterdam and Philadelphia: John Benjamins.

Greenberg, Joseph. 1983. Some areal characteristics of African languages. In Ivan Dihoff (ed.), *Current approaches to African linguistics 1*, 3–22. Dordecht/Cinnaminson: Foris.

Gregová, Renátá, Lívia Körtvélyessy & Július Zimmermann. 2010. Phonetic iconicity in the evaluative morphology of a sample of Indo-European, Niger-Congo and Austronesian languages. *Word Structure* 3. 156–180.

Guarisma, Gladys. 2003. Kpāʔ. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 307–334. London: Routledge.

Gudmundson, Anna. 2012. *L'accordo nell'italiano parlato da apprendenti universitari svedesi: Uno studio sull'acquisizione del numero e del genere in una prospettiva funzionalista*. Stockholm: Stockholm University dissertation.

Güldemann, Tom. 1998. The Kalahari Basin as an object of areal typology: A first approach. *Khoisan Phorum, Working Papers* 3. Köln: Institut für Afrikanistik, Universität zu Köln.

Güldemann, Tom. 1999. Head-initial meets head-final: nominal suffixes in eastern and southern Bantu from a historical perspective. *Studies in African Linguistics* 29. 49–91.

Güldemann, Tom. 2000. Noun categorization systems in Non-Khoe lineages of Khoisan. *Afrikanistische Arbeitspapiere* 63. 5–33.

Güldemann, Tom. 2004. Reconstruction through 'de-construction': the marking of person, gender and number in the Khoe family and Kwadi. *Diachronica* 21. 251–306.

Güldemann, Tom. 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern sub-Saharan Africa. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 151–185. Cambridge: Cambridge University Press.

Güldemann, Tom. 2010. Sprachraum and geography: Linguistic macro-areas in Africa. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, vol. 2: Language mapping, 561–585, Maps 2901–2914. Berlin: Mouton de Gruyter.

Güldemann, Tom & Rainer Vossen. 2000. Khoisan. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*, 99–122. Cambridge: Cambridge University Press.

Guthrie, Malcom. 1966. *Grammaire et dictionnaire de lingala*. Westmead: Greg Press Limited.

Hagman, Roy S. 1977. *Nama Hottentot grammar*. Bloomington: Indiana University.

Hämeen-Anttila, Jaakko. 2000. Grammatical gender and its development in Classical Arabic. In Barbara Unterbeck, Matti Rissanen, Tettu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*, 595–608. Berlin: Mouton de Gruyter.

Harrell, Frank E. 2001. *Regression modeling strategies*. New York: Springer.

Harrell, Richard S. 1965. *A short reference grammar of Moroccan Arabic*. Washington D.C.: Georgetown University Press.

Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook. (Handbücher zur Sprach- und kommonikationwissenshaft)*, 1492–1510. Berlin: Mouton de Gruyter.

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. 25–70.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11. 119–132.

Haspelmath, Martin. 2009. An empirical test of the agglutination hypothesis. In Sergio Scalise, Francesca Masini, Elisabetta Magni & Antonietta Bisetto (eds.), *Universals of language today*, 13–29. Dordrecht: Springer.

Haspelmath, Martin. 2012. Continent-wide typology: a recent trend within diversity linguistics. In *Diversity linguistics comment*, Available online at http://dlc.hypotheses.org/340.

Haspelmath, Martin. 2013. Occurrence of nominal plurality. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Max Planck Digital Library, Chapter 34. Available online at: http://wals.info/chapter/34. Accessed on 2014-02-14.

Heath, Jeffrey. 2005. *A grammar of Tamashek (Tuareg of Mali)*. Mouton de Gruyter.

Heath, Teresa. 2003. Makaa (A 83). In Derek Nurse & Gerard Philippson (eds.), *The Bantu languages*, 335–348. London: Routledge.

Heine, Bernd. 1970. *Status and use of African lingua francas*. München: Welt Forum Verlag.

Heine, Bernd. 1975. Language typology and corvengence areas in Africa. *Linguistics* 144. 27–47.

Heine, Bernd. 1976. *A typology of African languages based on the order of meaningful elements* Kölner Beiträge zur Afrikanistik 4. Berlin: Dietrich Reimer Verlag.

Heine, Bernd. 1982. African noun class systems. In H. Seiler & C. Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen*, 189–216. Tübingen: Gunter Narr Verlag.

Heine, Bernd. 1997. Grammaticalization chains across languages: an example from Khoisan. Paper presented at the Symposium on Grammaticalization, Houston, March 1997.

Heine, Bernd. 1999. The ǁAni: Grammatical Notes and Texts. *Khoisan Phorum* 11. Universität zu Köln, Institut für Afrikanistik.

Heine, Bernd & Christa König. 2011. *The !Xun language: a dialect grammar of Northern Khoisan.* Cologne: Rüdiger Köppe.

Heine, Bernd & Christa König. 2013. Northern Khoesan: !Xun. In Rainer Vossen (ed.), *The Khoesan languages*, 293–325. London and New York: Routledge.

Heine, Bernd & Tania Kuteva. 2006. *The changing languages of Europe.* Oxford: Oxford University Press.

Heine, Bernd & Derek Nurse. 2000. Introduction. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*, 1–10. Cambridge: Cambridge University Press.

Heine, Bernd & Derek Nurse. 2008. Introduction. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 1–14. Cambridge: Cambridge University Press.

Heine, Bernd & Rainer Vossen. 1983. On the origin of gender in Eastern Nilotic. In Rainer Vossen & Marianne Becchaus-Gerst (eds.), *Nilotic studies: Proceedings of the international symposium on languages and history of the Nilotic peoples, Cologne, January 4-6, 1982*, 245–268. Köln: Dietrich Reimer Verlag.

Hetzron, Robert. 1978. The nominal system of Awngi (Southern Agaw). *Bulletin of the School of Oriental and African Studies, University of London* 41. 121–141.

Hockett, Charles F. 1958. *A course in modern linguistics.* New York: Macmillan.

Hulstaert, Gustaaf. 1965. *Grammaire du lomongo. deuxième partie: Morphologie.* Tervuren: Musée Royale de l'Afrique Centrale.

Iemmolo, Giorgio. 2011. *Towards a typological study of Differential Object Marking and Differential Object Indexation.* Pavia: Università degli studi di Pavia dissertation.

Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41. 9–45.

Jespersen, Otto. 1922. *Language: Its nature, development and origin.* London: Allen and Unwin.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72. 533–578.

Katamba, Francis. 2003. Bantu nominal morphology. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 103–120. London: Routledge.

Kavari, Jekura U. & Lutz Marten. 2009. Multiple noun class prefixes in Otjiherero. In Peter Austin, Oliver Bond, Monik Charette, David Nathan & Peter Sells (eds.), *Proceeding of Conference on Language Documentation & Linguistic Theory 2*, London: SOAS.

Khan, Geoffrey. 2008. *Jewish Neo-Aramaic dialect of Urmi.* Piscataway, NJ: Gorgias Press.

Kibrik, Aleksander. 1992. Defective paradigms: number in Daghestanian. In *EUROTYP working papers, Theme 7*, vol. 16, Strassbourg: ESF.

Kilian-Hatz, Christa & Bernd Heine. 2010. On nominal gender marking in Khwe. *Khoisan Phorum* Working Paper 1. 2–26. Revised version of 1997.

Koptjevskaja-Tamm, M. 2004. Mass and collection. In Geert E. Booij, Christian Lehmann, Joachim Mugdan, Stavros Skopeteas & E. Booij (eds.), *Morphology: A handbook on inflection and word formation*, 1016–1031. Berlin: Mouton de Gruyter.

Koptjevskaja-Tamm, Maria. 2002. Number (book review). *Journal of pragmatics* 32. 1911–1919.

Koptjevskaja-Tamm, Maria. 2011. Language contact. In Jae-Jung Song (ed.), *The Oxford handbook of linguistic typology*, 568–590. Oxford: Oxford University Press.

Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: An areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages*, vol. 2: Grammar and typology, 615–750. Amsterdam: John Benjamins.

Körtvélyessy, Lívia. 2011. A cross-linguistic research into phonetic iconicity. *Lexis* 6. 27–39. [Special Issue: Diminutives and augmentatives in the languages of the world].

Körtvélyessy, Lívia. 2012. Evaluative morphology from cross-linguistic perspective. In *Paper presented at the conference on "Universals and Typology in Word-Formation II", Košice*, .

Körtvélyessy, Lívia & Pavel Stekauer. 2011. Introduction. *Lexis* 6. 3. [Special Issue: Diminutives and augmentatives in the languages of the world].

Kossman. 2014. Derivational gender in Moroccan Berber: examples from Ayt Seghrushen. *STUF – Language Typology and Universals* 67. 21–33.

Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflections*. Utrecht: LOT: University of Leiden dissertation.

Kusters, Wouter. 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 3–22. Amsterdam: John Benjamins.

Lacan, Ph. 1942. *Français-soussou soussou-français*. Bordeaux: Librairie d'Amérique et d'Orient.

Lakoff, George. 1987. *Women, fire and dangerous things*. Chicago: University of Chicago Press.

Lamberti, Marcello & Roberto Sottile. 1997. *The Wolaytta language*. Köln: Köppe. MI.

Langacker, Ronald. 1987a. Nouns and verbs. *Language* 63. 53–94.

Langacker, Ronald W. 1987b. *Foundations of cognitive grammar*, vol. 1: Theoretical prerequisites. Standford: Standford University Press.

Lass, Roger. 1990. How to do things with junk: Exaptation in language evolution. *Journal of Linguistics* 26. 79–102.

Leitch, Myles. 2003. Babole (C 101). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 392–421. London: Routledge.

Leslau, Wolf. 1945. Grammatical sketches in Tigré (North Ethiopic): Dialect of Mensa. *Journal of the American Oriental Society* 65. 164–203.

Leslau, Wolf. 1948. Supplementary observations on Tigré grammar. *Journal of the American Oriental Society* 68. 127–139.

Leslau, Wolf. 1995. *Reference grammar of Amharic*. Wiesbaden: Harassowitz.

Lewis, M. P., Gary F. Simons & Charles D. Fenning (eds.). 2013. *Ethnologue: Languages of the world, seventeenth edition*. Dallas: SIL International. Online version: http://www.ethnologue.com.

Lojenga, Constance Kutsch. 2003. Bila (D 32). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 450–474. London: Routledge.

Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45. 435–464.

Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Maho, Jouni. 1999. *A comparative study of Bantu noun classes*. Göteborg: Orientalia et Africana Gothoburgensia dissertation. Acta universitatis gothoburgensia.

Maho, Jouni. 2003. Remarks on a few "polyplural" classes in Bantu. *Africa & Asia* 3. 161–184.

Marten, Lutz. 2010. The great siSwati locative shift. In Anne Breitbarth, Christopher Lucas, Sheila Watts & David Willis (eds.), *Continuity and change in grammar*, 249–267. Amsterdam: Benjamins.

Matasovič, Ranko. 2004. *Gender in Indo-European*. Heidelberg: Winter.

Matisoff, James A. 1992. The mother of all morphemes: Augmentatives and diminutives in areal and universal perspective. In Martha Ratliff & Eric Shiller (eds.), *Papers from the first annual meeting of the Southeast Asian linguistic society*, 293–249. Arizona State University, Program for Southeast Asian Studies.

McGregor, William. 2003. The nothing that is, the zero that isn't. *Studia Linguistica* 57. 75–119.

McLaughlin, Fiona. 1992. *Noun classification in Seereer-Siin*. Austin: The University of Texas dissertation.

McLaughlin, Fiona. 1997. Noun classification in Wolof: When affixes are not renewed. *Studies in African Linguistics* 26. 1–28.

McLaughlin, Fiona (ed.). 2009. *The languages of urban Africa*. London: Continuum.

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.

Meeussen, A. E. 1959. *Essai de grammaire rundi*. Tervuren: Musée Royal du Congo Belge.

Mel' čuck, Igor & Élie Bakiza. 1997. Les classes nominales en kirundi. *Bulletine de la société de linguistique de Paris* 82. 283–341.

Mettouchi, Amina. 2000. La "t" n'est-il qu'une marque de féminin en berbère (kabyle)? *Faits de Langue* 14. 217–225.

Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 23–41. Amsterdam: John Benjamins.

Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, changes*. Amesterdam: John Benjamins.

Mithun, Marianne. 1988. Lexical categories and the evolution of verbal number. In Noonan Michael & Michael Hammond (eds.), *Theoretical morphology*, 211 – 233. New York: Academic Press.

Moravcsik, Edith. 2003. A semantic analysis of associative plurals. *Studies in Language* 27. 469–503.

Moscati, S. 1964. *An introduction to the comparative grammar of the Semitic languages.* Wiesbaden: Otto Harrasowitz.

Mous, Maarten. 1993. *A grammar of Iraqw.* Hamburg: Buske.

Mous, Maarten. 2003a. Nen (A 44). In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 283–306. London: Routledge.

Mous, Maarten. 2008. Number as an exponent of gender in Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *Interaction of morphology and syntax: Case studies in Afroasiatic*, 137–160. Amsterdam: John Benjamins.

Mous, Marteen. 2012. Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *The Afroasiatic languages*, 342–422. Cambridge: Cambridge University Press.

Mous, Marten. 2003b. Loss of linguistic diversity in Africa. In Mark Janse & Sijmen Tol (eds.), *Language death and language maintenance: Theoretical, practical and descriptive approaches*, 157–170. Amsterdam: John Benjamins.

Mufwene, Salikoko. 2003. Contact languages in the Bantu area. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 195–208. London: Routledge.

Myachina, E. N. 1981. *The Swahili language.* London: Routledge & Kegan Paul.

Newman, Paul. 1990. *Nominal and verbal plurality in Chadic.* Dordrecht: Foris Publications.

Newman, Paul. 2000. *The Hausa language: an encyclopedic reference grammar.* New Haven: Yale University Press.

Nichols, Johanna. 1971. Diminutive consonant symbolism in Western North America. *Language* 47. 826–848.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: University of Chicago Press.

Nicolas, Francis. 1953. *La langue berbère de Mauritanie.* Dakar: IFAN.

Nordhoff, Sebastian, Harald Hammarström, Robert Forkel & Martin Haspelmath. 2013. *Glottolog 2.2.* Max Planck Institute for Evolutionary Anthropology. Available online at http://glottolog.org, Accessed on 2014-08-19.

Novelli, Bruno. 1985. *A grammar of the Karimojong language.* Berlin: Drietrich Reimer Verlag.

Nurse, Derek & Gérard Philippson. 2003. Towards a historical classification of the Bantu languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 164–181. London: Routledge.

Oomen, Antoinette. 1981. Gender and plurality in Rendille. *Afroasiatic linguistics* 8. 85–75.

Osam, Emmanuel Kweku. 1993. The loss of the noun class system in Akan. *Acta Linguistica Hafniensia: International Journal of Linguistics* 26. 81–106.

Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 265–285. Amsterdam: John Benjamins.

Payne, Doris L. 1998. Maasai gender in typological perspective. *Studies in African Linguistics* 27. 159–175.

Payne, Thomas. 1997. *Describing morpho-syntax: A guide for field linguists*. Cambridge: Cambridge University Press.

Penchoen, Thomas. 1973a. *Etude sintaxique d'un parler berbère (Ait-Frah de l'Aurès)*. Napoli: Centro di studi magrebini.

Penchoen, Thomas G. 1973b. *Tamazight of the Ayt Ndhir*. Los Angeles: Undena Publications.

Petzell, Malin. 2008. *The Kagulu language of Tanzania: Grammar, texts and vocabulary*. Köln: Rüdiger Köppe Verlag.

Plank, Frans. 1999. Split morphology: How agglutination and flexion mix. *Linguistic Typology* 3. 279–340.

Plank, Frans & Elena Filimonova. 2009. *The universals archive*. Konstanz: Universität Konstanz. Available online at http://typo.uni-konstanz.de/archive/intro/index.

Plank, Frans & Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45. *Linguistic Typology* 1. 53–101.

Poulos, George. 1990. *A linguistic analysis of Venda*. Pretoria: Via Afrika Limited.

Poulos, George & Sonja Bosh. 1997. *Zulu*. München: Lincom Europa.

Poulos, George & Louis J. Louwrens. 1994. *A linguistic analysis of Northern Sotho*. Pretoria: Via Afrika Limited.

Pozdniakov, Konstantin. 2010. La classification nominale: à la croisée des paradigmes. In Franck Floricic (ed.), *Essais de typologie et de linguistique générale: Mélanges offerts à Denis Creissels*, 87–105. Paris: ENS Éditions.

Pozdniakov, Konstantin. 2013. Protolanguage and prototype: a "proto-letter" and a "proto-spirit" in noun classes in Niger-Congo. *Histoire, épistémologie, langage* 35. 65–82.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London and New York: Longman.

Rapold, Christian J. 2006. *Towards a grammar of Benchnon*. Leiden: University of Leiden dissertation.

Reid, Nicholas. 1990. *Ngan'gityemerri: A language of the Daly River region, northern territory of Australia*. Canberra: Australian National University dissertation.

Reid, Nicholas. 1997. Class and classifier in Ngan'gityemerri. In Mark Harvey & Nicholas Reid (eds.), *Nominal classification in aboriginal Australia*, 165–225. Amsterdam: John Benjamins.

Roper, E. M. 1928. *Tu Bedawiɛ: An elementary handbook for the use of Sudan governament officials*. Hertford: Stephen Austin and Sons, LTD.

Ryding, Karin. 2005. *A reference grammar of Modern Standard Arabic*. New York: Cambridge University Press.

Saeed, John I. 1987. *Somali reference grammar*. Wheaton: Dunwoody Press.

Sagna, Serge. 2008. *Formal and semantic proprties of the Gujjolay Eegima (a.k.a Banjal) nominal classification system*. London: University of London – School of Oriental and African Studies dissertation.

Sagna, Serge. 2011. Semantic categorizations in the Gújjolaay Eegima collectives and distributives. In Oliver Bond, Peter K. Austin, David Nathan & Marten Lutz (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, 237–246. London: Department of Linguistics, SOAS.

Sagna, Serge. 2012. Physical properties and culture-specific factors as principles of semantic categorization of the Gújjolaay Eegima noun class systems. *Cognitive Linguistics* 23. 129–163.

van Sanbeek, Jan Cornelius. 1955. *A Bemba grammar*. Cape Town: Longmans, Green.

Santos, Rosine Sachot. 1996. *Le mey*. Paris: Universite de la Sorbonne Nouvelle dissertation.

Sapir, Edward. 1951. A study in phonetic symbolism. In David G. Mandelbaum (ed.), *Selected writings of Edward Sapir in language, culture and personality*, 61–72. Berkeley and Los Angeles: University of California Press.

Sasse, Hans-Jrgen. 1974. Notes on the structure of Galab. *Bulletin of the School of Oriental and African Studies, University of London* 37. 407–438.

Savà, Graziano. 2005. *A grammar of Ts'amakko*. Köln: Rüdiger Köppe Verlag.

Savà, Graziano. 2007. Interaction between gender and number in Cushitic: The case of Ts'amakko. In Marco Moriggi (ed.), *XII incontro italiano di linguistica camito-semitica (afroasiatica)*, 203–212. Rubbettino: Soveria Mannelli.

Schuh, Russel G. 1989. Gender and number in Miya. In Zygmunt Frajzyngier (ed.), *Current progress in Chadic linguistics*, 171–181. Amsterdam: John Benjamins.

Segerer, Guillaume. 2002. *La langue bijogo de bubaque (guiné bissau)*. Paris, Louvain: Peeters.

Segerer, Guillaume. 2010. 'Isolates' in 'Atlantic'. Paper presented at the workshop on Language Isolates in Africa (Lyon, December 3-4, 2010).

Serzisko, Fritz. 1992. Collective and transnumeral nouns in Somali. In Adam M. Hussein & Charles L. Geshekter (eds.), *Proceedings of the first international congress of Somali studies*, 513–525. Atlanta: Scholar Press.

Seyoum, Mulugeta. 2008. *A grammar of Dime*. Leiden: Netherlands Graduate School of Linguistics dissertation.

Shadeberg, Tilo. 2003. Derivation. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 71–89. London: Routledge.

Shirtz, Shahar & Doris Payne. 2012. Gendered prefixes in Maa. Paper presented at the 45th meeting of the *Societas Linguistica Europea*, Stockholm, September 2012.

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.

Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity*. University of Helsinki: General Linguistics, Department of Modern Languages dissertation.

Smith-Stark, Cedric. 1974. The plurality split. *Chicago Linguistic Society* 10. 657–671.

Soukka, Maria. 2000. *A descriptive grammar of Noon*. München: Lincom Europa.

Storch, Anne. 2005. *The noun morphology of Western Nilotic*. Köln: Rüdiger Köppe Verlag.

Storch, Anne. 2007. Grammatical change and emblematic features in Western Nilotic. In Doris L. Payne & Jaime Peña (eds.), *Selected proceedings of the 37th annual conference on African linguistics*, 1–15. Somerville, MA: Cascadilla Proceedings Project.

Ström, Eva-Marie. 2011. The increasing importance of animacy in the agreement system of Ndengeleko and other Southern Coastal Bantu languages. In Jan-Olof Svantesson, Niclas Burenhult, Arthur Holmer, Anastasia Karlsson & Håkan Lundström (eds.), *Language Documentation and Description, Vol. 10*, 265–287. SOAS.

Stroomer, Harry. 1995. *A grammar of Boraama Oromo*. Köln: Rüdiger Köppe Verlag.

Stump, Gregory. 1993. How peculiar is evaluative morphology? *Journal of Linguistics* 29. 1–36.

Taljard, P.C., J.N. Khumalo & E.S Bosh S. 1991. *Handbook of Siswati*. Pretoria: J. L. Schaik.

Talmy, Leonard. 2000a. *Towards a cognitive semantics*, vol. 1: Typology and process in concept structuring. Cambridge, Massachussets: The MIT Press.

Talmy, Leonard. 2000b. *Towards a cognitive semantics*, vol. 2: Concept structuring systems. Cambridge, Massachussets: The MIT Press.

Taylor, Charles. 1985. *Nkore-Kiga*. Sidney: Croom Helm.

Teketal, Hirut Woldemariam. 2004. *The grammar of Haro with comparative notes on the Ometo linguistic group*. Adis Abab: Addis Ababab University dissertation.

Tendeng, Odile. 2000. *Le gusiilay: un essai de systématisation. une contribution à l'étude du jóola*. Freiburg: Albert-Ludwigs-Universität zu Freiburg dissertation.

Thompson, Kathrina D. & Antonia F. Schleicher. 2001. *Swahili's learner's reference grammar*. Madison, WI: NALCR Press.

Tosco, Mauro. 1991. *A grammatical sketch of Dahalo*. Hamburg: Helmut Buske.

Tosco, Mauro. 2001. *The Dhaasanac language*. Köln: Rüdiger Köppe Verlag.

Traill, Anthony. 1994. *A !Xóõ dictionary*. Köln: Köppe.

Travélé, Moussa. 1955. *Petite manuel franç-bambara*. Librairie Orientaliste Paul Geuthner.

Treis, Yvonne. 2008. *A grammar of Kambaata*. Cologne: Rüdiger Köppe.

Treis, Yvonne. 2011. Number in Kambaata. Paper presented at the Workshop on "Number in Africa and beyond", Cologne, 27-30 Sep.

Treis, Yvonne. 2014. Number in Kambaata. In Anne Storch & Gerrit Dimmendaal (eds.), *Number: Constructions and semantics. Case studies from Africa, India, Amazonia and Oceania.*, 111–133. Amsterdam: John Benjamins.

Trudgill, Peter. 1999. Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics* 35. 133–152.

Ultan, Russell. 1978. Size-sound symbolism. In Joseph Greenberg (ed.), *Universals of human language*, vol. 2: Phonology, 525–568. Stanford: Stanford University Press.

Vafaeian, Ghazaleh. 2013. Typology of nominal and adjectival suppletion. *STUF – Language Typology and Universals* 66. 112–140.

Van de Velde, Mark. 2006. Multifunctional agreement patterns in Bantu and the possibility of genderless nouns. *Linguistic Typology* 10. 183–221.

Van de Velde, Mark. 2008. *A grammar of Eton*. Berlin: Mouton de Gruyter.

Veselinova, Ljuba. 2006. *Suppletion in verb paradigms*. Amsterdam: John Benjamins.

Veselinova, Ljuba. 2013. Verbal number and suppletion. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: http://wals.info/chapter/80. Accessed on 2014-02-14.

Veselinova, Ljuba. forthcoming. Sampling procedures. In *WSK dictionary on theories and methods in linguistics*, Online publication under the auspices of Mouton de Gruyter.

Vossen, Rainer. 1986. Some observations on nominal gender in Naro. In F. Rottland (ed.), *Festschrift zum 60. Geburtstag von Carl F. Hoffmann*, 373–391. Hamburg: Buske.

Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

Wälchli, Bernhard. 2012. Grammaticalization clines in space: Zooming in on synchronic traces of diffusion processes. In Björn Wlemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical replication and borrowability in language contact*, 233–272. Berlin: Mouton de Gruyter.

Wald, Benji V. 1975. Animate concord in northeast coastal Bantu: Its linguistic and social implications as a case of grammatical convergence. *Studies in African Linguistics* 6. 267–314.

Watkins, Mark Hanna. 1937. A grammar of Chichewa, a Bantu language of British Central Africa. *Language* 13. 9–158.

Welmers, William. 1973. *African language structures*. Berkeley: University of California Press.

Westermann, Diedrich. 1930. *A study of the Ewe language*. London: Oxford University Press.

Whorf, Benjamin. 1941. The relation of habitual thought and behavior to languages. In Leslie Spier, Irving A. Hallowell & Stanley S. Newman (eds.), *Language, culture, and personality: Essays in memory of Edward Sapir*, 75–93. Menasha, Wisconsin: Stanley S. Newman.

Wierzbicka, Anna. 1984. Diminutives and dispreciatives: semantic representation for derivational categories. *Quaderni di semantica* 5. 123–130.

Wilson, W.A. A. 1989. Atlantic. In John Bendor-Samuel & Rhonda L. Hartell (eds.), *The Niger-Congo languages: A classification and decsription of Africa's largest language family*, 81–104. University Press of America.

Wilson, William A. A. 1961. *An outline of the Temne language*. London: School of Oriental and African Studies.

US-AB universitetsservice
Printed in Sweden 2014, www.us-ab.com

NORDIC ECOLABEL
341900
Printed matter

ISBN 978-91-7447-952-2

Department of Linguistics

Stockholm University